bioRχiv

Challenges in replicating layer-specificity of working memory processes in human dlPFC

Denis Chaimow[a,b,*], Jonas Karolis Degutis[a], Daniel Haenelt[b], Robert Trampel[b], Nikolaus Weiskopf[b,c,d] and Romy Lorenz[a,b]

[a]Cognitive Neuroscience & Neurotechnology Group, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[b]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[c]Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth System Sciences, Leipzig University, Leipzig, Germany
[d]Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3AR, UK
*Correspondence should be addressed to D.C. (denis.chaimow@tuebingen.mpg.de)

Although working memory reliably activates the dorsolateral prefrontal cortex (dlPFC), the functional significance of its distinct cytoarchitectonic layers is not well understood in humans. A recent functional magnetic resonance (fMRI) study at 7T demonstrated for the first time layer-specific responses in the human dlPFC during working memory. Superficial layers were more active during the delay period when working memory items needed to be manipulated compared to mere maintenance. In contrast, deeper layers were more active during the motor response to a probe compared to non-action. Like many current layer fMRI studies, this study relied on several manual and semi-manual processing steps, including the selection of regions of interest. To test the replicability of these findings, we conducted a pre-registered replication of this study in 21 subjects using a fully automated and reproducible analysis pipeline. Our results do not show the same layer-specific effects. Although we observed higher activity in the superficial layers in response to working memory manipulation during the delay period, we did not find any evidence for stronger deep layer involvement during motor response in the probe period. We argue that our results are biologically plausible in light of previous research as well as methodological considerations inherent in layer fMRI acquisition and analysis. Consequently, we conclude that the evidence regarding the functional role of different layers within the human dlPFC during working memory remains inconclusive. A focus on replicability, reproducibility, and a better understanding of the influence of methodological choices will help layer fMRI become a more routine tool in cognitive neuroscience.

Keywords: working memory, prefrontal cortex, high-resolution fMRI, cortical layers, VASO

Introduction

The dorsolateral prefrontal cortex (dlPFC) plays a central role in human high-level cognition, consistently activating during working memory (WM) tasks, which necessitate the temporary retention and manipulation of information for cognitive processes (Curtis and D'Esposito, 2003; D'Esposito and Postle, 2015). Despite its significance, the functional role of the dlPFC's laminar organization in human WM remains poorly understood. Research utilizing invasive electrophysiology in non-human primates suggests that distinct layers of the dlPFC contribute uniquely to WM processing. During delayed-response tasks, delay period activity in dlPFC has been mainly localized to superficial layers 2 and 3 (Bastos et al., 2018; Sawaguchi et al., 1990, 1989). Supported by neuroanatomical findings about local connectivity (Kritzer and Goldman-Rakic, 1995), this has led to a model explaining WM delay activity as recurrent activity in superficial layers of the prefrontal cortex across similarly tuned neighboring cortical columns (Compte et al., 2000; Goldman-Rakic, 1995). In contrast, neurons in deep layer 5 have been shown to be more active during motor responses following the delay (Opris et al., 2011). Additionally, Bastos et al. (2018) identified phase-amplitude coupling between beta activity in deeper layers and delay-specific gamma band activity in superficial layers, leading to a model where deep layers gate the WM processes carried out by superficial layers.

The development of such mechanistic models of cortical function and their experimental evaluation depends on the ability to measure neuronal activity at the resolution of cortical columns and layers. Until recently this has been the exclusive domain of invasive electrophysiological studies and was therefore only applicable in non-human animals. This poses significant limits on understanding human cognition as it is unclear whether and how these findings can be translated to humans. Non-human primates used in such studies are often overtrained, spending several years learning a single cognitive task, which can limit the generalizability of the findings (Miller et al., 2022). Multitask studies are particularly challenging due to the extensive time required for training. Furthermore, electrophysiological recordings in animals are typically restricted to a few preselected cortical sites. Ideally one would want to measure such activity directly in humans, not least because many human cognitive processes may not be suited to being studied in animals.

Fortunately, technological advances in high-resolution fMRI have now paved the way to study the human brain at the resolution of columns and layers. Most layer fMRI studies so far have targeted sensory (Kok et al., 2016; Koopmans et al., 2010; Olman et al., 2012) and motor

systems (Huber et al., 2017). However, recently, these methods have also been used to study higher-level cortices (Degutis et al., 2024; Finn et al., 2019; Sharoh et al., 2019). Finn et al. (2019) demonstrated layer-specific responses non-invasively in the human dlPFC for the first time during a WM task using VASO fMRI at 7T. This imaging sequence differs from the more commonly used gradient-echo (GE) blood oxygenation level dependent (BOLD) contrast in that it has higher spatial specificity making it more suitable for studying layer-specific activations. Importantly, this study showed the differential involvement of superficial and deep layers of the same region in different phases of the task. Superficial layers were more active during the delay period when WM items needed to be manipulated compared to their mere maintenance (in the following: "superficial layer delay effect"). In contrast, deeper layers were more active when participants had to respond to a subsequent WM probe by pressing a button (in the following: "deep layer action effect").

Finn et al. (2019) used an analysis approach that entailed the careful manual localization of active regions of interest (ROIs) in dlPFC and a manual delineation of layers to sample from. Such manual and semi-manual processing steps are common in many layer fMRI studies to date (Dresbach et al., 2023; Faes et al., 2023; Huber et al., 2023; Zhang et al., 2023), aiming to achieve high precision. This is understandable as until recently the focus has been on feasibility and demonstration studies. In addition to hand-crafted and custom developed analysis methods that were not pre-registered, sample sizes were often small and many layer fMRI results are awaiting replication.

To make the leap from methodological demonstration to routine use in cognitive neuroscience research it is important to confirm the replicability of such findings, ideally using more automated analysis approaches.

Here, we repeated the experiment conducted by Finn et al. (2019) in 21 participants and analyzed it using a fully automated and reproducible pre-registered analysis pipeline. To preview, we found higher superficial layer delay period activity that is specific to WM manipulation. However, we did not find any evidence for a deep layer action effect during the response period. We discuss our results in the context of previous neurophysiological findings and methodological considerations of layer fMRI acquisition and analysis and implications for layer fMRI in cognitive neuroscience.

Results

Pre-registered replication analysis

We determined the target sample size of 21 participants using conservative power analysis based on effect sizes of the original study (see Methods). Participants performed a WM task identical to Finn et al. (2019), while we simultaneously acquired VASO and BOLD fMRI data at 7T using an SS-SI-VASO sequence (Huber et al., 2014; Huber et al., 2021; Stirnberg and Stöcker, 2021).

Briefly, in each trial of two experimental runs a random set of five letters was presented visually, followed by a cue indicating whether the letters had to be mentally rearranged in alphabetical order or merely remembered in their original order (alphabetization vs. remembering). After a delay of 10 s one of the letters appeared as a probe and participants had to press one of five buttons indicating its alphabetical or

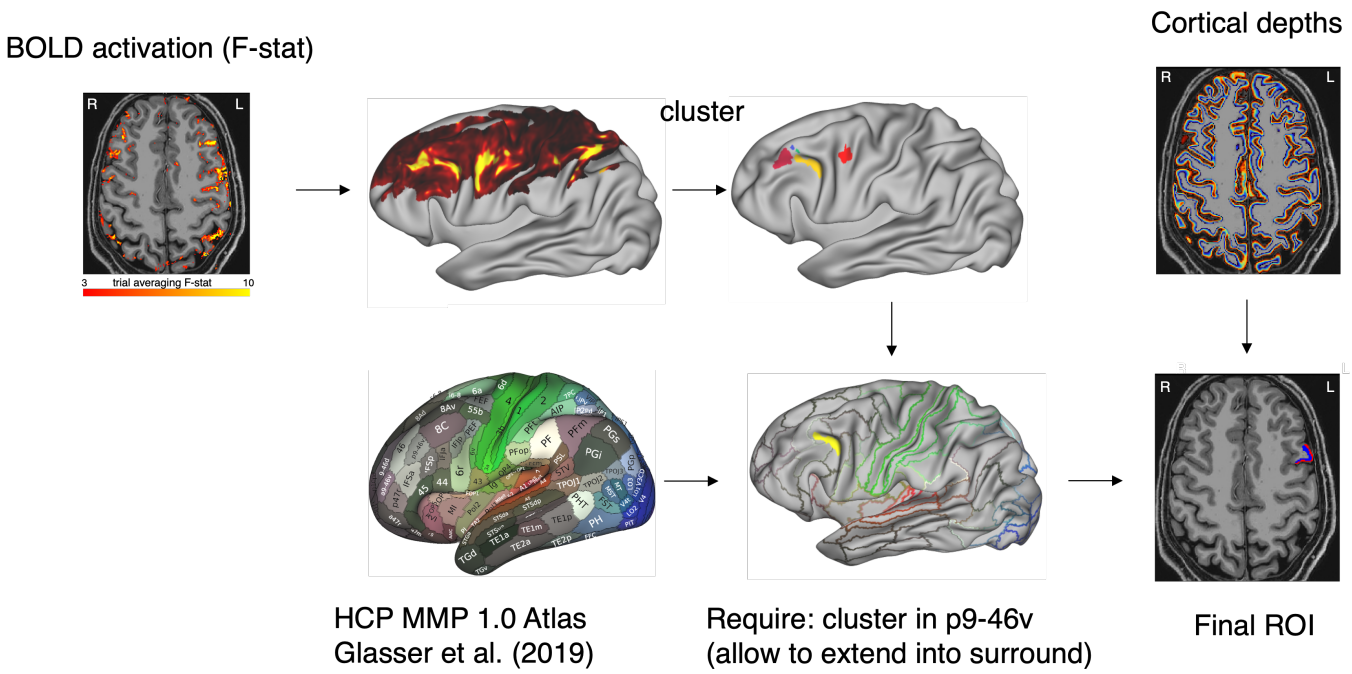

Figure 1: Automatic definition of ROIs. ROIs were determined by combining anatomical information obtained from the HCP MMP 1.0 atlas (Glasser et al., 2016) (bottom left) and on whether a location was activated by the experimental paradigm using the BOLD contrast (top row from left). We chose the largest activation cluster overlapping p9-46v and selected all contiguous voxels that belonged either to p9-46v or any of the surrounding parcels (46, 8C, IFSp, IFSa) accounting for uncertainties in single subject parcellation. This ROI was then split into a superficial and a deep layer based on cortical depth estimates.

ordinal position. In two additional runs only alphabetization trials were presented but were varied with respect to whether a response was required or not (action vs. non-action).

We attempted to replicate the findings of Finn et al. (2019) following an analysis pipeline that we had pre-registered on Open Science Framework Registries (https://doi.org/10.17605/osf.io/txh3s). We designed the analysis pipeline to closely follow the original study. However, consistent with our overall goal, we formalized all manual steps. Most notably, ROIs were automatically determined by combining precise anatomical information about dlPFC localization obtained from a fine-grained surface-based cortical parcellation (Glasser et al., 2016) with subject-specific clusters of task-evoked functional activation in the BOLD data (Fig. 1 and Methods). Following Finn et al. (2019), we leveraged BOLD fMRI for the ROI selection procedure due to its higher sensitivity compared to VASO fMRI. We then split each automatically defined subject-specific ROI into superficial and deep layer voxels based on cortical depth estimates (Fig. 1) and estimated average trial time courses from the VASO data as a function of layer and experimental condition.

We present the resulting laminar time courses averaged over subjects (Fig. 2A) in a format similar to Finn et al. (2019). The alphabetization condition resulted in elevated VASO signals following the delay period as compared to mere remembering (compare blue to green curve). Furthermore, action trials resulted in elevated VASO signals following the response period (compare red to orange curve).

To quantify these differences, we first averaged the time points associated with delay and response periods (Figure 2B) while accounting for hemodynamic delay. We then analyzed the dependence on cortical layer, experimental condition, and trial period using a series of two-way ANOVAs, identical to Finn et al. (2019). First, we analyzed each layer's response in each type of run separately for effects of experimental condition and trial period.

Manipulation dependent signal increases in both layers

Activity in both layers was higher in alphabetization trials (superficial layers: $p=1.7\times10^{-5}$, $F(1,20)=31.51$; $\eta_p^2=0.612$; deep layers: $p=3.8\times10^{-5}$, $F(1,20)=27.7$; $\eta_p^2=0.580$; effect of condition in a two-way ANOVA with factors condition x trial period). We found no significant interactions between trial type and period (superficial layers: $p=0.21$; $F(1,20)=1.71$; $\eta_p^2=7.89\times10^{-2}$; deep layers: $p=0.32$; $F(1,20)=1.05$; $\eta_p^2=4.97\times10^{-2}$).

Action dependent signal increase in superficial but not deep layers

For the action vs. non-action comparison, we found an interaction between trial type and trial period in superficial layers ($p=1.6\times10^{-3}$; $F(1,20)=13.2$; $\eta_p^2=0.398$), such that activity in action trials was higher than in non-action trials during the response period. Also in deep layers, activity was somewhat higher in action trials than in non-action trials but this difference was not statistically significant ($p=0.059$; $F(1,20)=4.01$; $\eta_p^2=0.167$; effect of condition in a two-way ANOVA with factors condition x trial period) and there was no significant interaction ($p=0.068$; $F(1,20)=3.71$; $\eta_p^2=0.156$).

Replicating superficial layer delay effect but not deep layer action effect

To determine whether response differences between conditions were specific to a particular layer and the type of experimental manipulation, we conducted a second type of two-way ANOVA, following methodology outlined by Finn et al. (2019). For each period, we tested for an interaction between layer and type of condition contrast. This is the main analysis which our pre-registered power calculation addressed. We found a significant interaction for the delay period ($p=5.6\times10^{-3}$; $F(1,20)=9.62$; $\eta_p^2=0.325$) but not the response period ($p=0.23$; $F(1,20)=1.50$; $\eta_p^2=7.00\times10^{-2}$). The pattern of interaction (Fig. 2B) for the delay period was qualitatively similar to Finn et al. (2019), such that superficial layer responded stronger to the alphabetization vs. remembering contrast. However, contrary to Finn et al. (2019), during the response periods superficial and not deep layers showed a trend of responding stronger to the action vs. non-action contrast.

Though not part of the preregistered replication analysis, we carried out the same statistical analyses on the BOLD part of the data (Supplementary Fig. 1). Here, we found significant interaction effects, with both alphabetization and action, both showing stronger responses in superficial layers (delay period: $p=8.6\times10^{-7}$, $F(1,20)=50.0$; $\eta_p^2=0.710$; response period: $p=8.9\times10^{-4}$, $F(1,20)=15.2$; $\eta_p^2=0.431$; two-way ANOVA interaction with factors layer and condition contrast). However, these results pointing to relative higher activity in superficial layers need to be interpreted with caution as BOLD is known to suffer from a draining vein bias resulting in higher responses in superficial layers (Havlicek and Uludağ, 2020; Heinzle et al., 2016; Markuerkiaga et al., 2016).

Failure to replicate deep layer action effect not due to limits in sensitivity

A lack of a statistical interaction effect for the response period in the VASO analysis might suggest that any true effect present in our data was smaller than previously reported, or that our study lacked the statistical power to detect effects comparable to those reported by Finn et al. (2019). To further compare our results to Finn et al. (2019) we used bootstrapping over subjects to estimate confidence intervals on the layer by condition contrast interaction effects, taking the direction of layer differences into account. For the superficial layer delay effect, 95% of bootstrap samples yielded η_p^2 values below 0.544 or failed to demonstrate an increased superficial layer delay contrast (blue dots below the zero line in Fig. 2C). These findings stand in contrast to the value of 0.86 reported by Finn et al. (2019). For the deep layer action effect more than 95% of

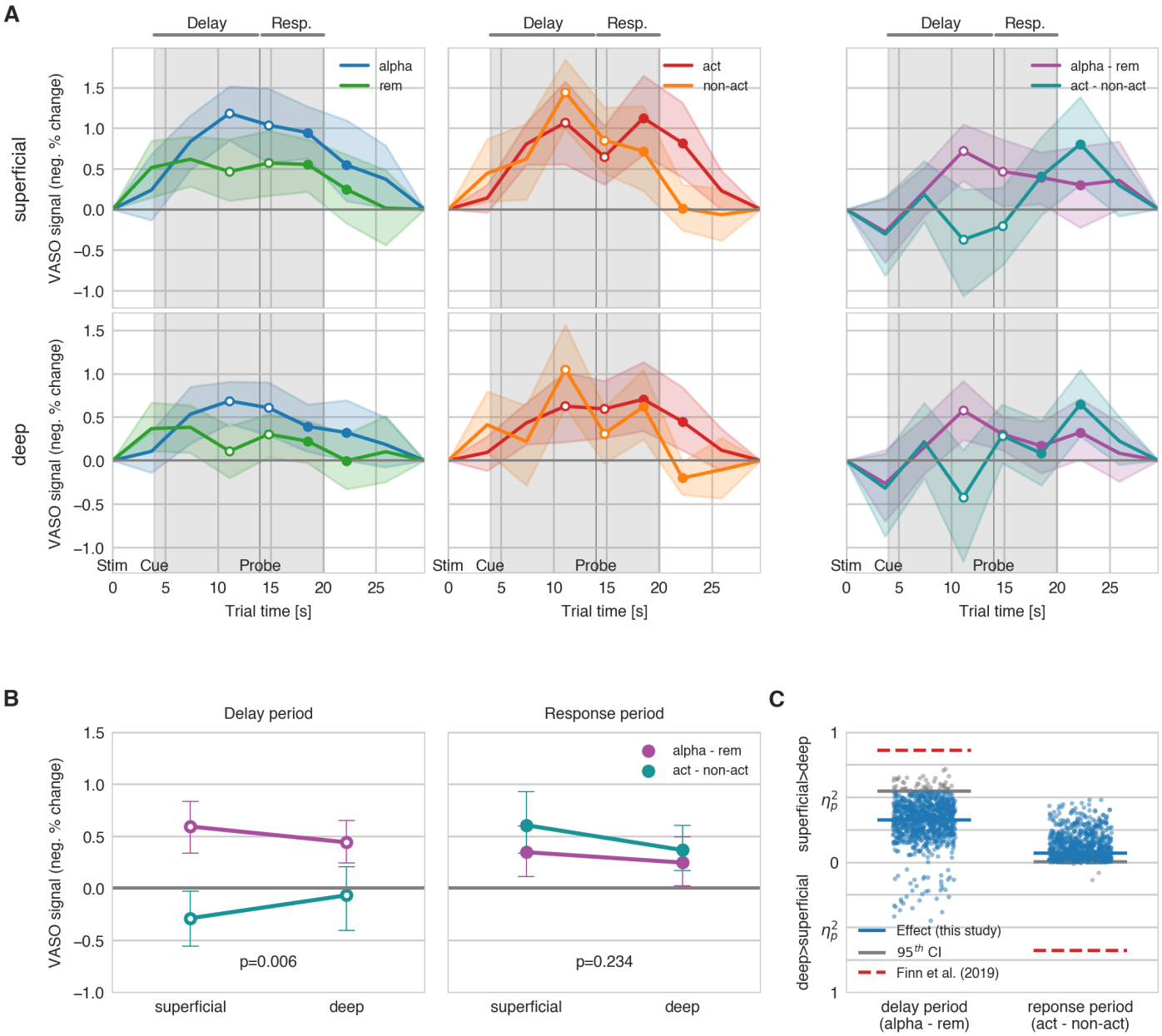

Figure 2: Results of preregistered replication analysis. (A) Trial averaged VASO time courses for superficial (top row) and deep layers (bottom row) for alphabetizations vs. remembering (left), action vs. non-action (middle) and differences between conditions from both types of runs. Each panel shows the negative percent VASO signal response, i.e. a proxy of positive cerebral blood volume change and activation (see Methods), as a function of trial time. A sequence of 5 random letters were presented at t = 0 s. At t = 4 s a cue indicated whether participants had to alphabetize the sequence or remember it in its original order. At t = 14 s one letter was probed, and participants had to respond using a button press. The shaded band indicates the 95% confidence interval over subjects. Open circles and filled circles represent delay and response period data points, respectively, that were used for further analysis. **(B)** VASO response contrasts obtained from delay (left) and response periods (right) are shown for alphabetization vs. remembering (purple) and action vs. non-action (teal) for superficial and deep layers. The interaction between type of contrast and layer was significant for the delay but not for the response period. **(C)** Estimated interaction effect sizes (partial eta squared; blue lines) for delay and response periods are shown together with their sampling distributions estimated using bootstrapping (blue circles). Effect size values belonging to a higher superficial layer contrast are plotted upwards, those belonging to a higher deep layer contrast downwards. For both periods the interaction effects' 95th percentiles (gray lines) are clearly distant from the results of Finn et al. (2019), shown as dashed red lines.

bootstrap samples did not result in any higher deep layer action contrast, which Finn et al. (2019) reported with an associated effect size of η_p^2=0.68 (Fig. 2C). This discrepancy suggests that any potential deep layer action effect is considerably smaller than reported by Finn et al. (2019) and that their superficial layer delay effect is also stronger than what our data suggests.

Control analyses demonstrate adequacy of methods and robustness of results

We conducted a series of control analyses to evaluate how data quality and methodological choices may have influenced our results.

First, we assessed our data quality by calculating tSNR maps on individual BOLD-corrected VASO runs (Supplementary Fig. 2) and calculating the mean within-ROI tSNR for each subject. The median tSNR across subjects

was 14.1, which can be considered moderate to excellent (Huber, 2020).

We inspected gray matter segmentation and structural to functional image registration focusing on each participant's individual ROI (Supplementary Fig. 3). The fully automatic segmentation and registration appears to have performed well within and around each ROI. However, in subjects 2, 3, 4 and 7 the automatically generated ROI was located near the edge of the slab, resulting in lower tSNR (Supplementary Fig. 2) and reduced ROI volume (Supplementary Fig. 6A). We therefore reran our main analysis without these subjects. Results for this subgroup were similar to those from the complete group (Supplementary Fig. 4) in that there was a significant interaction for type of contrast and layer during the delay period ($p=0.013$; $F(1,16)=7.80$; $\eta_p^2=0.328$) with the superficial layer showing higher responses specific to the alphabetization vs. remembering contrast. As in the main analysis there was no significant interaction during the response period ($p=0.79$; $F(1,16)=7.60\times10^{-2}$; $\eta_p^2=4.73\times10^{-3}$).

We have thus qualitatively demonstrated that the automatic segmentation and ROI definition performed reasonably well. To further quantitatively assess potential loss of accuracy compared to manual segmentations we also created manually drawn laminar ROIs. We hand-segmented superficial and deep layers directly on T1-weighted VASO images, a procedure comparable to Finn et al. (2019), but we used the automatic ROIs to determine location and extent. This procedure required about 50 hours of manual labor. We reran our main analysis using these carefully manually delineated ROIs. Resulting time courses (Supplementary Fig. 5) were qualitatively very similar. Importantly, we did not find a deep layer action effect ($p=0.55$, $F(1,20)=0.370$, $\eta_p^2=1.81\times10$). However, the superficial layer delay effect size decreased compared to the automatic ROI analysis ($p=0.127$; $F(1,20)=2.53$; $\eta_p^2=0.112$). Although the decrease of the superficial layer effect was somewhat unexpected, the confidence intervals of the manual and automatic analyses highly overlap (Supplementary Fig. 5C). Together these results point against limited sensitivity of our automatic ROI analysis approach.

In our automated ROI estimation we had to define a target ROI surface area. It is possible that smaller ROIs would have been more sensitive to a more localized effect. Alternatively, assuming a more distributed response, averaging over a larger ROI could have boosted statistical power. Supplementary Fig. 6 shows the effect of varying ROI size. We find some variations in the strength of the superficial delay layer effect during the delay period in that moderately small surface areas between 60 mm^2 and 120 mm^2 yield statistically significant results while very small surface areas (< 60 mm^2) as well as surface areas between 140 mm^2 and 360 mm^2 are not significant; for very large areas from 380 mm^2 on, results would again yield statistical significance. Importantly though, no ROI size resulted in a deep layer action effect.

In our analysis, we followed the ROI description provided by Finn et al. (2019), where the focus was on activations in and around the left dlPFC (approximately Brodmann area 9/46) which we pre-registered as the left p9-46v parcel from the HCP MMP 1.0 atlas (Glasser et al., 2016). However, it is important to note that their subject-specific ROIs were drawn manually and without the help of atlasing tools, and these ROIs are not publicly available. Given that these ROIs were manually defined, there is a possibility that the original analysis inadvertently concentrated on alternative dlPFC ROIs. In light of this, we should reconsider the focus on p9-46v and explore other candidate regions of the dlPFC to ensure a more comprehensive and accurate assessment of task-specific activations.

To that extent, we estimated clusters in the group averaged BOLD F-stat map, backprojected them to individual subjects and used these as constraining regions for automatic ROI calculation. Supplementary Fig. 7 shows that these clusters resemble the multiple demand (MD) system (Assem et al., 2020; Duncan, 2010), a frontoparietal network that typically activates during a wide variety of cognitively demanding tasks. Cluster 3 resembles our main ROI region in and around left p9-46v. In addition, clusters 2 and 4 situated in close posterior and anterior proximity, respectively, are candidates for alternative dlPFC ROIs. However, the ROIs based on both of these group clusters resulted in qualitatively similar layer and condition contrasts as our main ROI (Supplementary Fig. 7D). In particular, none of them showed a deep layer action effect.

Finally, we assessed whether a larger separation of layers within the ROI would affect results by increasing laminar specificity. We reran the main analysis on modified laminar ROIs by leaving a gap of ⅓ relative cortical depth between the superficial and deep layer ROI. Results (Supplementary Fig. 8) were qualitatively similar to the main analysis. Again superficial layers showed a higher alphabetization - remembering contrast during the delay period (contrast x layer interaction: $p=3.0\times10^{-3}$; $F(1,20)=11.4$; $\eta_p^2=0.363$) while there was no significant interaction for the response period ($p=0.22$; $F(1,20)=1.63$; $\eta_p^2=7.56\times10^{-2}$).

Discussion

In this pre-registered study we aimed to replicate a seminal study that for the first time demonstrated layer-specific responses in the human dlPFC during WM (Degutis et al., 2024; Finn et al., 2019; Sharoh et al., 2019). We were not able to fully replicate the results of Finn et al. (2019) using our pre-registered analysis. While we observed higher activity in the superficial layers compared to the deep layers for manipulation versus maintenance during the delay period, which is consistent with the original study, we found a different pattern during the response period. In contrast to Finn et al. (2019) we observed higher superficial layer activity for action than for non-action during the response period. Using confidence interval estimation and a series of control analyses, we demonstrated that the failure to replicate was likely not an issue of low sensitivity nor a

problem of potentially reduced accuracy of automatic layer ROI segmentation methods of our replication study.

Finn et al. (2019) suggested that their findings regarding superficial layer activity at least partially reflect recurrent excitatory connections. Our findings, showing a less distinct higher superficial layer response during manipulation trials, do not contradict their conclusion. However, it questions the degree to which manipulation-dependent increases in delay period activity are exclusively localized to superficial layers. Evidence from non-human primates indicates that some delay period activity is also present in deep layers, albeit to a lesser extent. Sawaguchi et al. (1990) using single cell recordings during an oculomotor delayed response task found that, while most delay-type neurons that did not differentiate between stimuli were found in superficial layers, most stimulus differentiating delay period neurons were found in deep layers. Similarly, deep layer delay period activity has been found in LFP and multi-unit activity (MUA) in multiple prefrontal areas (Bastos et al., 2018). These findings also align with our previous work in humans using BOLD fMRI in WM tasks. Specifically, we observed activation in the deep layers of the left dlPFC during the delay period for heightened WM load demands, albeit to a significantly lesser extent than in the superficial layers (Degutis et al., 2024). Importantly, in our previous work, we demonstrated that this pattern is generalizable to increases in WM load more broadly and does not appear to be specific to WM manipulation, further supporting the robustness of our findings.

In contrast to what was observed during the delay period in the original study, our findings did not reveal stronger deep layer responses for action compared to non-action trials during the probe period. Instead, we observed increased activity in both layers but especially superficial. These results do not rule out a deep layer involvement, but also do not support conclusions about an exclusive deep layer response. We speculate that such distinct deep layer responses are also not very likely. Even though there is ample evidence from electrophysiological studies in non-human primates that dlPFC deep layers are active or involved in generating a response in delayed-response tasks (Opris et al., 2011; Sawaguchi et al., 1990, 1989), these studies stem from recordings of spiking activity, which primarily reflect action potentials originating from the somata of the neurons. In contrast, fMRI signals are more closely related to local field potentials (Logothetis et al., 2001), which primarily reflect synaptic activity and dendritic integration, potentially extending into more superficial layers as is the case for apical dendrites of layer 5 pyramidal neurons. If the deep layers of the dlPFC are indeed responsible for motor response generation, the fMRI signal in the deep layers themselves might be weaker than expected as it primarily reflects synaptic input and dendritic activity rather than axonal output or somatic spiking, with stronger signals potentially arising in downstream targets where the motor commands are processed. However, this might not be the case for feedback connections that target deep layer synapses and could explain why layer fMRI studies have found deep layer-specific results in sensory cortices in the past (Carricarte et al., 2024; Kok et al., 2016) .

Moreover, in the paradigm that we and Finn et al. (2019) employed action trials differed from non-action trials not only in the requirement of a motor response but also in whether the probed letter was shown (action) or not (non-action). This points to a potential drawback of the experimental paradigm as it complicates the clear dissociation between motor response and retrieval, making it difficult to isolate the specific contributions of each cognitive process. Previous findings have shown ramping gamma-band activity at the end of memory delays, which is thought to result from its disinhibition following a drop in alpha/beta power, allowing for a successful read-out of information from WM (Buschman and Miller, 2022; Hussar and Pasternak, 2010). Since gamma-band activity has been linked to superficial layers (Bastos et al., 2018; Buffalo et al., 2011; Maier et al., 2010; Smith et al., 2013), this could explain why we did not observe exclusive deep layer activation in our study. Specifically, the readout from WM during the retrieval period in the action conditions likely contributed more prominently to superficial layer activity, compared to the non-action conditions where no such readout was required.

Interestingly, delayed response tasks in non-human primates also showed superficial layer activity well into the probe period (Bastos et al., 2018; supplementary Fig. 8). This was true for a search task in which images had to be compared to working memory content but also for a simpler delayed saccade task in which cognitive demands after the delay were minimal. Similarly, Markowitz et al. (2015) isolated WM storage-specific population activity, that could be localized preferentially to superficial layers, and that showed heightened activity not only during the delay period but also during the probe period when a memory guided saccade had to be made.

Even if laminar neuronal activity were perfectly localized to individual layers, a certain degree of laminar unspecificity in fMRI measurements is expected. Reasons are residual effects of subject motion and the sampling of cortical depths across approximately 2 mm cortical thickness into 0.8 mm or 0.9 mm wide voxels. The latter is expected to introduce some amount of cross-layer blurring (e.g. see simulations in Koopmans et al., 2011; Shmuel et al., 2007) due to varying alignment between voxels and local cortical geometry as well as biological and measurement uncertainty in exact layer boundaries. For deep layer activity there is the added effect of draining veins, whereby spatially specific vascular changes in deep layers are being propagated downstream by ascending cortical veins. This effect is most significant for deoxyhemoglobin changes measured by GE BOLD fMRI. However, some minimal sensitivity to blood volume changes in larger vessels, foremost intra-cortical arteries is expected when using VASO fMRI (Akbari et al., 2022). While varying degrees of task-dependent coactivation of deep and superficial layer activity has been demonstrated (Huber et

al., 2017), to date we are not aware of any VASO results showing completely isolated deep layer responses compared to baseline, except for Finn et al. (2019).

We attempted to develop a fully automated analysis while still focusing on precision. Nevertheless, it may well be the case that results like those presented in Finn et al. (2019) are currently only achievable with careful manual intervention. However, it should be noted that previous attempts of automated segmentation were able to yield layer-specific results (Degutis et al., 2024; Kok et al., 2016; Lawrence et al., 2019). Moreover, our manual delineated ROIs did not reveal qualitatively different results.

We believe that manual approaches benefit the advancement of current methodological possibilities, and are therefore well justified as a first step. However, the next step must be to give such methods and results a solid foundation that ensures replicability with adequate methodological effort. Moreover, defining what constitutes 'adequate' requires a deeper exploration of how various methodological choices in data acquisition and analysis impact the outcomes. General challenges in layer fMRI have been discussed elsewhere (Merriam and Kay, 2022; Viessmann and Polimeni, 2021). In our opinion, one important issue in the context of replicability and reproducibility is ROI selection. Especially in association cortices it is very difficult to localize a ROI due to less correspondence between macroscopic anatomy and function as well as unknown structure-function relationships at the mesoscale (Frost and Goebel, 2012). Moreover, given these uncertainties, the exact locations of specific functional regions is much more variable between subjects than those in primary sensory and motor cortices (Mueller et al., 2013). In this context, identifying ROIs functionally within each subject individually using adequate localizer tasks (Fedorenko et al., 2024) or dense-sampling resting-state approaches, where resting-state networks can be precisely mapped on an individual-level basis (Gordon et al., 2017) can help address these challenges by reducing inter-subject variability. Lastly, variability in local cortical geometry and local vasculature may influence resulting laminar response profiles adding a further ROI selection dependence to the results. A better understanding of all these factors and objective yet precise methods of ROI selection are needed.

In addition to determining ROI location and extent tangentially to the cortical surface, the manual ROI selection employed by Finn et al. (2019) also entailed gray matter segmentation and the delineation of superficial and deep layers. This second aspect, while not ideal for reproducibility, can potentially offer more precise results. One disadvantage is the significant effort required, which, although still manageable in this study, could become burdensome in larger studies or those involving multiple or larger ROIs, or targeting the entire cortex. Furthermore, we caution against assuming the absolute superiority of manual segmentations. While manual segmentation combined with ROI selection may lead to avoidance of regions affected by imaging artifacts or venous formations, our experience suggests that this also may introduce a bias towards geometrically simpler ROIs. For instance, in our manual ROI segmentations, we adhered to an automatically generated location and extent, and observed that some geometrically complex ROIs might not have been selected manually, particularly when relying on predefined orthogonal cuts. The orientation of these cuts can also affect the accuracy of manual layer delineations. Lastly, manual segmentation may lead to false confidence in subjective interpretations of images affected by noise and artifacts, potentially overestimating its advantages compared to automatic segmentation results.

Lastly, given the wide range of analysis approaches currently used in layer fMRI and the above discussed issue of ROI selection, we believe it to be beneficial that the pioneering experiments are being followed up by studies in which hypotheses and analytic choices have been pre-registered. The prevalence of feasibility studies in current layer fMRI research, that focus on confirmation of basic models of laminar involvement, renders them particularly conducive to pre-registration.

Conclusion

Here, we were not able to fully reproduce the results of Finn et al. (2019). While it can be debated whether our methods were adequate to detect these layer-specific findings, our results appear biologically plausible given invasive data from non-human primates and methodological considerations of layer fMRI acquisition in analysis. Furthermore, quality assurance and bootstrap analyses support that the data quality and the power of our study was sufficient to detect the previously reported effects. In the end, evidence about the type of laminar activity in the human dlPFC during working memory tasks remains inconclusive. It is necessary to further investigate which methodological differences may have caused differences in results between studies. Focusing on replicability and reproducibility, methods for consistent ROI selection, and a better understanding of influencing factors may help layer fMRI become a routine and dependable tool in cognitive neuroscience research.

References

Akbari, A., Bollmann, S., Ali, T.S., Barth, M., 2022. Modelling the depth - dependent VASO and BOLD responses in human primary visual cortex. Hum Brain Mapp. https://doi.org/10.1002/hbm.26094

Assem, M., Glasser, M.F., Essen, D.C.V., Duncan, J., 2020. A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. Cereb Cortex New York Ny 30, 4361–4380. https://doi.org/10.1093/cercor/bhaa023

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Méd. Image Anal. 12, 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bastos, A.M., Loonis, R., Kornblith, S., Lundqvist, M., Miller, E.K., 2018. Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory. P Natl Acad Sci Usa 115, 1117–1122. https://doi.org/10.1073/pnas.1710323115

Brent, R.P., 1973. Algorithms for minimization without derivatives. Courier Corporation.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R., 2011. Laminar differences in gamma and alpha coherence in the ventral

stream. Proc National Acad Sci 108, 11262–11267. https://doi.org/10.1073/pnas.1011284108

Buschman, T.J., Miller, E.K., 2022. Working Memory Is Complex and Dynamic, Like Your Thoughts. J. Cogn. Neurosci. 35, 17–23. https://doi.org/10.1162/jocn_a_01940

Carricarte, T., Iamshchinina, P., Trampel, R., Chaimow, D., Weiskopf, N., Cichy, R.M., 2024. Laminar dissociation of feedforward and feedback in high-level ventral visual cortex during imagery and perception. iScience 27, 110229. https://doi.org/10.1016/j.isci.2024.110229

Chaimow, D., Weiskopf, N., Lorenz, R., 2022. Estimating voxel-wise laminar depth from surface meshes, in: 28th Annual Meeting of the Organization for Human Brain Mapping (OHBM).

Compte, A., Brunel, N., Goldman-Rakic, P.S., Wang, X.-J., 2000. Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. Cereb Cortex 10, 910–923. https://doi.org/10.1093/cercor/10.9.910

Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. Trends Cogn Sci 7, 415–423. https://doi.org/10.1016/s1364-6613(03)00197-9

Degutis, J.K., Chaimow, D., Haenelt, D., Assem, M., Duncan, J., Haynes, J.-D., Weiskopf, N., Lorenz, R., 2024. Dynamic layer-specific processing in the prefrontal cortex during working memory. Commun. Biol. 7, 1140. https://doi.org/10.1038/s42003-024-06780-8

D'Esposito, M., Postle, B.R., 2015. The Cognitive Neuroscience of Working Memory. Annu Rev Psychol 66, 1–28. https://doi.org/10.1146/annurev-psych-010814-015031

Dickie, E.W., Anticevic, A., Smith, D.E., Coalson, T.S., Manogaran, M., Calarco, N., Viviano, J.D., Glasser, M.F., Essen, D.C.V., Voineskos, A.N., 2019. Ciftify: A framework for surface-based analysis of legacy MR acquisitions. Neuroimage 197, 818–826. https://doi.org/10.1016/j.neuroimage.2019.04.078

Dresbach, S., Huber, R., Gulban, O.F., Goebel, R., 2023. Layer-fMRI VASO with short stimuli and event-related designs at 7T. NeuroImage 279, 120293. https://doi.org/10.1016/j.neuroimage.2023.120293

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn Sci 14, 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Faes, L.K., Martino, F.D., Huber, L. (Renzo), 2023. Cerebral blood volume sensitive layer-fMRI in the human auditory cortex at 7T: Challenges and capabilities. Plos One 18, e0280855. https://doi.org/10.1371/journal.pone.0280855

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods 41, 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fedorenko, E., Ivanova, A.A., Regev, T.I., 2024. The language network as a natural kind within the broader landscape of the human brain. Nat. Rev. Neurosci. 25, 289–312. https://doi.org/10.1038/s41583-024-00802-4

Finn, E.S., Huber, L., Jangraw, D.C., Molfese, P.J., Bandettini, P.A., 2019. Layer-dependent activity in human prefrontal cortex during working memory. Nat. Neurosci. 22, 1687–1695. https://doi.org/10.1038/s41593-019-0487-z

Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J. - P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189-210. https://doi.org/10.1002/hbm.460020402

Frost, M.A., Goebel, R., 2012. Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. NeuroImage 59, 1369–1381. https://doi.org/10.1016/j.neuroimage.2011.08.035

Gaser, C., Dahnke, R., Thompson, P.M., Kurth, F., Luders, E., Initiative, the A.D.N., 2024. CAT: a computational anatomy toolbox for the analysis of structural MRI data. GigaScience 13, giae049. https://doi.org/10.1093/gigascience/giae049

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., Essen, D.C.V., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178. https://doi.org/10.1038/nature18933

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Essen, D.C.V., Jenkinson, M., Consortium, for the W.-M.H., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Goldman-Rakic, P.S., 1995. Cellular basis of working memory. Neuron 14, 477–485. https://doi.org/10.1016/0896-6273(95)90304-6

Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U.F., 2017. Precision Functional Mapping of Individual Human Brains. Neuron 95, 791-807.e7. https://doi.org/10.1016/j.neuron.2017.07.011

Gulban, O.F., Nielson, D., Poldrack, R., lee, john, Gorgolewski, C., Vanessasaurus, Ghosh, S., 2019. poldracklab/pydeface: v2.0.0. https://doi.org/10.5281/zenodo.3524401

Havlicek, M., Uludağ, K., 2020. A dynamical model of the laminar BOLD response. Neuroimage 204, 116209. https://doi.org/10.1016/j.neuroimage.2019.116209

Heinzle, J., Koopmans, P.J., Ouden, H.E.M. den, Raman, S., Stephan, K.E., 2016. A hemodynamic model for layered BOLD signals. Neuroimage 125, 556–570. https://doi.org/10.1016/j.neuroimage.2015.10.025

Huber, L., Finn, E.S., Chai, Y., Goebel, R., Stirnberg, R., Stöcker, T., Marrett, S., Uludag, K., Kim, S.-G., Han, S., Bandettini, P.A., Poser, B.A., 2021. Layer-dependent functional connectivity methods. Prog. Neurobiol. 207, 101835. https://doi.org/10.1016/j.pneurobio.2020.101835

Huber, L., Handwerker, D.A., Jangraw, D.C., Chen, G., Hall, A., Stüber, C., Gonzalez-Castillo, J., Ivanov, D., Marrett, S., Guidi, M., Goense, J., Poser, B.A., Bandettini, P.A., 2017. High-Resolution CBV-fMRI Allows Mapping of Laminar Activity and Connectivity of Cortical Input and Output in Human M1. Neuron 96, 1253-1263.e7. https://doi.org/10.1016/j.neuron.2017.11.005

Huber, L., Ivanov, D., Krieger, S.N., Streicher, M.N., Mildner, T., Poser, B.A., Möller, H.E., Turner, R., 2014. Slab-selective, BOLD-corrected VASO at 7 Tesla provides measures of cerebral blood volume reactivity with high signal-to-noise ratio. Magnetic resonance in medicine 72, 137–148. https://doi.org/10.1002/mrm.24916

Huber, L., Kassavetis, P., Gulban, O.F., Hallett, M., Horovitz, S.G., 2023. Laminar VASO fMRI in focal hand dystonia patients. Dystonia 2, 10806. https://doi.org/10.3389/dyst.2023.10806

Huber, L. (Renzo), Poser, B.A., Bandettini, P.A., Arora, K., Wagstyl, K., Cho, S., Goense, J., Nothnagel, N., Morgan, A.T., Hurk, J. van den, Müller, A.K., Reynolds, R.C., Glen, D.R., Goebel, R., Gulban, O.F., 2021. LayNii: A software suite for layer-fMRI. NeuroImage 237, 118091. https://doi.org/10.1016/j.neuroimage.2021.118091

Huber, L., 2020. Quality assurance measures for layer-fMRI time series: How to obtain them in LAYNII. layer fMRI blog. URL http://web.archive.org/web/20230601163246/https://layer-fmri.com/2020/04/06/qa/ (accessed 6.1.21).

Hussar, C., Pasternak, T., 2010. Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. Proc. Natl. Acad. Sci. 107, 21842–21847. https://doi.org/10.1073/pnas.1009956107

Kashyap, S., 2021. srikash/presurfer: ondu.

Kok, P., Bains, L.J., van Mourik, T., Norris, D.G., de Lange, F.P., 2016. Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. Curr Biol 26, 371–376. https://doi.org/10.1016/j.cub.2015.12.038

Koopmans, P.J., Barth, M., Norris, D.G., 2010. Layer - specific BOLD activation in human V1. Hum Brain Mapp 31, 1297-1304. https://doi.org/10.1002/hbm.20936

Koopmans, P.J., Barth, M., Orzada, S., Norris, D.G., 2011. Multi-echo fMRI of the cortical laminae in humans at 7T. Neuroimage 56, 1276–1285. https://doi.org/10.1016/j.neuroimage.2011.02.042

Kritzer, M.F., Goldman - Rakic, P.S., 1995. Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in

the rhesus monkey. J Comp Neurol 359, 131‑143. https://doi.org/10.1002/cne.903590109
Lawrence, S.J., Norris, D.G., Lange, F.P. de, 2019. Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. Elife 8, e44422. https://doi.org/10.7554/elife.44422
Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157. https://doi.org/10.1038/35084005
Maier, A., Adams, G.K., Aura, C., Leopold, D.A., 2010. Distinct Superficial and Deep Laminar Domains of Activity in the Visual Cortex during Rest and Stimulation. Frontiers Syst Neurosci 4, 31. https://doi.org/10.3389/fnsys.2010.00031
Markowitz, D.A., Curtis, C.E., Pesaran, B., 2015. Multiple component networks support working memory in prefrontal cortex. Proc National Acad Sci 112, 11084–11089. https://doi.org/10.1073/pnas.1504172112
Markuerkiaga, I., Barth, M., Norris, D.G., 2016. A cortical vascular model for examining the specificity of the laminar BOLD signal. Neuroimage 132, 491–498. https://doi.org/10.1016/j.neuroimage.2016.02.073
Merriam, E.P., Kay, K., 2022. The need for validation in layer-specific fMRI. https://doi.org/10.31219/osf.io/f9vqc
Miller, J.A., Tambini, A., Kiyonaga, A., D'Esposito, M., 2022. Long-term learning transforms prefrontal cortex representations during working memory. Neuron 110, 3805-3819.e6. https://doi.org/10.1016/j.neuron.2022.09.019
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual Variability in Functional Connectivity Architecture of the Human Brain. Neuron 77, 586–595. https://doi.org/10.1016/j.neuron.2012.12.028
Olman, C.A., Harel, N., Feinberg, D.A., He, S., Zhang, P., Ugurbil, K., Yacoub, E., 2012. Layer-Specific fMRI Reflects Different Neuronal Computations at Different Depths in Human V1. Plos One 7, e32536. https://doi.org/10.1371/journal.pone.0032536
Opris, I., Hampson, R.E., Stanford, T.R., Gerhardt, G.A., Deadwyler, S.A., 2011. Neural Activity in Frontal Cortical Cell Layers: Evidence for Columnar Sensorimotor Processing. J Cognitive Neurosci 23, 1507–1521. https://doi.org/10.1162/jocn.2010.21534
Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., 2019. PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods 51, 195–203. https://doi.org/10.3758/s13428-018-01193-y
Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Essen, D.C.V., Jenkinson, M., 2014. MSM: A new flexible framework for Multimodal Surface Matching. Neuroimage 100, 414–426. https://doi.org/10.1016/j.neuroimage.2014.05.069
Sawaguchi, T., Matsumura, M., Kubota, K., 1990. Catecholaminergic effects on neuronal activity related to a delayed response task in monkey prefrontal cortex. J Neurophysiol 63, 1385–1400. https://doi.org/10.1152/jn.1990.63.6.1385
Sawaguchi, T., Matsumura, M., Kubota, K., 1989. Depth distribution of neuronal activity related to a visual reaction time task in the monkey prefrontal cortex. J Neurophysiol 61, 435–446. https://doi.org/10.1152/jn.1989.61.2.435
Sharoh, D., Mourik, T. van, Bains, L.J., Segaert, K., Weber, K., Hagoort, P., Norris, D.G., 2019. Laminar specific fMRI reveals directed interactions in distributed networks during language processing. Proc National Acad Sci 116, 21185–21190. https://doi.org/10.1073/pnas.1907858116
Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N.K., Ugurbil, K., 2007. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. NeuroImage 35, 539–552. https://doi.org/10.1016/j.neuroimage.2006.12.030
Smith, M.A., Jia, X., Zandvakili, A., Kohn, A., 2013. Laminar dependence of neuronal correlations in visual cortex. J Neurophysiol 109, 940–947. https://doi.org/10.1152/jn.00846.2012
Stirnberg, R., Stöcker, T., 2021. Segmented K‑space blipped‑controlled aliasing in parallel imaging for high spatiotemporal resolution EPI. Magn. Reson. Med. 85, 1540‑1551. https://doi.org/10.1002/mrm.28486
Viessmann, O., Polimeni, J.R., 2021. High-resolution fMRI at 7Tesla: challenges, promises and recent developments for individual-focused fMRI studies. Curr Opin Behav Sci 40, 96–104. https://doi.org/10.1016/j.cobeha.2021.01.011
Zhang, K., Chen, L., Li, Y., Paez, A.G., Miao, X., Cao, D., Gu, C., Pekar, J.J., Zijl, P.C.M. van, Hua, J., Bakker, A., 2023. Differential Laminar Activation Dissociates Encoding and Retrieval in the Human Medial and Lateral Entorhinal Cortex. J. Neurosci. 43, 2874–2884. https://doi.org/10.1523/jneurosci.1488-22.2023

Acknowledgements

We thank Emily Finn and Renzo Huber for helpful discussions and for providing additional information about analyses related to Finn et al. (2019). J.K.D. was supported by the Max Planck Society and BMBF. N.W. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project no. 347592254 (WE 5046/4-2); the Federal Ministry of Education and Research (BMBF) under support code 01ED2210; the European Union's Horizon 2020 research and innovation programme under the grant agreement No 681094. R.L. was funded by the Wellcome Trust (209139/Z/17/Z). This project also received funding from the Klaus Tschira Stiftung gGmbH foundation (support code GSO/KT18). This paper was typeset with the bioRxiv word template by @Chrelli: www.github.com/chrelli/bioRxiv-word-template

Competing interests

The Max Planck Institute for Human Cognitive and Brain Sciences and Wellcome Centre for Human Neuroimaging have institutional research agreements with Siemens Healthcare. N.W. holds a patent on acquisition of MRI data during spoiler gradients (US 10,401,453 B2). N.W. was a speaker at an event organized by Siemens Healthcare and was reimbursed for the travel expenses.

Methods

Participants and sample size rationale

22 healthy participants were recruited from the Max Planck Institute for Human Cognitive and Brain Sciences (MPI CBS) subject database. All participants had normal or corrected to normal vision, were native German speakers, and right-handed as determined by the Edinburgh handedness inventory. Participants had no history of neuropsychiatric disorders and were not currently taking psychoactive medications. All participants gave written and informed consent. The study was approved by the ethics committee of the University of Leipzig (441/20-ek).

Data from one subject was excluded from further analysis due to excessive head motion as prescribed in the preregistration (relative motion in at least two runs, as estimated by retrospective motion correction, exceeded the mean plus 3 standard deviations from all runs in all subjects).

The final sample consisted of 21 participants (ages 18 – 36 years, 10 female), which was our predefined target sample size and had been determined as follows. First, we performed a power analysis in G*Power (Faul et al., 2009), using partial eta squared values calculated from the ANOVA results reported in Finn et al. (2019) (personal communication). Specifically, we took the smallest interaction effect of the below described across layers analysis ($\eta_p^2 = 0.681$ for the interaction between layer and contrast during the response period). We decreased this effect size by ⅓ to 0.45 to account for a fewer planned number of runs per subject in our study. Due to time constraints imposed by the maximum allowable duration for scanning per subject in one session (90 min), we limited the data acquisition to 4 instead of 5-6 runs; this decision was made in accordance with local institutional regulations and ethical guidelines specified for our study. We set the effect size specification option in G*Power to "as in Cohen (1988)". Note that this gave us the most conservative results. For the non-sphericity correction, we used the smallest sphericity index estimate from the across layers analysis (ε=0.612; communicated to us by the authors). We further set α=0.05, the desired power to 1-β=0.8, the number of groups to 1 and the number of measurements to 4. This approach resulted in a required sample size of 21 subjects. We performed a bootstrapping simulation using all single subject results from Finn et al. (2019) provided to us by the authors. From that data we randomly generated 1000 sets of single subject datasets for each tested subject sample size (sampling with replacement). We then ran the same across layers ANOVA as in Finn et al. (2019) and counted the proportion of statistically significant interaction effects ($p<0.05$). This procedure resulted in a minimum sample size of only 6 subjects to achieve a power of at least

80%. Note that in contrast to our G*Power analysis, the bootstrapping analysis did not consider the reduced amount of data per subject. Nonetheless, the bootstrapping analysis indicates the strength of effects reported in Finn et al. (2019). However, to be conservative we chose the larger sample size of 21 subjects, as estimated above, as our target sample size.

Stimuli and procedure

Participants performed a WM task inside the scanner. The task was created with PsychoPy3 (Peirce et al., 2019) and was made identical to the description in Finn et al. (2019) (see their "axial protocol"). Identical to Finn et al. (2019) there were two types of experimental runs (alphabetization vs. remembering and action vs. non-action), each consisting of 20 pseudo randomized trials. Every subject performed up to two runs of each type in an interleaved manner for a total of 4 runs. Finn et al. (2019) used up to 6 runs per subject, a number which we could not achieve, due to maximum scanning time constraints at the MPI CBS. Note that this was accounted for in our target sample size calculation. Before the acquisition of the four experimental runs with the SS-SI-VASO sequence, identical to Finn et al. (2019), subjects underwent a 6-min whole-brain functional localizer run using a GE-BOLD sequence with 2 mm isotropic resolution. This localizer consists entirely of "ALPHABETIZE" trials and slightly altered timing (the ITI was shortened from 16s to 5s). The location of delay period peak activity within the dlPFC from the analysis of the functional localizer run is then used to position the SS-SI-VASO slab for the experimental runs consisting of 24 slices with a resolution of 0.8 x 0.8 x 0.9 mm^3.

fMRI protocol

MRI data was acquired using a Siemens MAGNETOM Terra 7T MRI pTx system (Siemens Healthineers, Erlangen, Germany) at the MPI CBS, using an 8Tx/32Rx channel head coil (Nova Medical Inc., Wilmington, MA, USA). Structural T1 weighted images were acquired using MP2RAGE sequence (Marques et al., 2010) (TR = 5000 ms, TE = 2.27 ms, 0.75 mm isotropic voxels) at two inversion times (TI of 900 ms, 2700 ms with a flip angle of 3°, 5°, respectively) that were combined to yield a T1-weighted image with uniform contrast (UNI). High-resolution partial-brain functional data were acquired using an SS-SI-VASO sequence with 3D-EPI readout (Huber et al., 2021; Stirnberg and Stöcker, 2021; version 822d59f4). The resolution was 0.8 x 0.8 x 0.9 mm^3, with a FOV of 150 x 150 x 12.6 mm^3. The volume TR encompassing the acquisition of a nulled and not-nulled image was 3.7 s in total. Further parameters were: TE=19 ms, TI=1.4 s, GRAPPA factor 3, 6/8 partial Fourier, readout bandwidth=1064 Hz/pixel). A lower resolution whole-brain 2D EPI scan (2 mm isotropic resolution, TR = 2 s, TE = 18 ms) was used for intra-session functional localization and slab positioning.

Experimental procedure

After instruction and a short behavioral training to familiarize with the task, participants were first scanned using the low-resolution localizer after which the anatomical image was acquired. The functional localizer scan was analyzed and the high-resolution slab was positioned to maximize coverage of activated locations within and around the left dlPFC. Participants then performed 4 experimental runs while high-resolution VASO data was acquired. Alphabetization vs. remembering runs were interleaved with action vs. no-action runs.

Differences in data acquisition to Finn et al. 2019:

Our MRI scanner (Siemens 7T Magnetom Terra) differed from the one used in Finn et al. 2019. Although both scanners were produced by Siemens Healthineers, they ran different non-compatible software versions. Therefore, we were not able to use the exact SS-SI-VASO pulse sequence that was used in Finn et al. (2019). Instead we used a newer, slightly different implementation (Huber et al., 2021; version number 822d59f4) with a 3D-EPI readout (Stirnberg and Stöcker, 2021). Among other things, this sequence differed in the implementation of the 3D-EPI read-out, in that the temporal spacing of nulled and non-nulled read-out blocks is not symmetric and that it does not use a phase-skip parameter. Consequently, possible and recommended parameter protocols differed somewhat between both sequences. Our employed resolution was 0.8 x 0.8 x 0.9 mm^3, compared to 0.9 x 0.9 x 1.1 mm^3 in Finn et al. (2019), which is expected to yield more thermal noise but also better resolvability between layers. Our TR, including the measurement of 1 nulled and 1 non-nulled volume, was 3.7s compared to 4s in Finn et al. (2019).

Preprocessing and co-registration

Before further processing all data was converted from DICOM to Nifti format, anonymized, and structural data was defaced (Gulban et al., 2019). In subject 1, the high-resolution functional runs consisted of 188 instead of 178 volumes. These additional volumes were discarded. All data was organized into a BIDS compatible structure, adding meta information as necessary (e.g. task event files).

Preprocessing of structural data

We developed a custom processing pipeline around Freesurfer's recon-all (Fischl, 2012). This addressed the problem that residual noise in processed MP2RAGE data results in difficulties with skull stripping and that high-resolution 7T T1w images are difficult for estimating gray matter without being affected by dura. Using pilot data sets, we informally explored several different approaches before settling on what produced the most accurate gray matter and white matter surfaces. The MP2RAGE INV2 images (second inversion, mostly PD weighted) were bias-corrected using SPM12 (Friston et al., 1994) before multiplying it with the MP2RAGE UNI images (uniform T1-weighted images with salt-and-pepper noise in regions of low signal) to yield MPRAGE like images (Kashyap, 2021). These images were fed into the first stage (autorecon1) of Freesurfers (v7.1) recon-all processing stream without skull-stripping using the highres flag. The same images were fed into the CAT12 (Gaser et al., 2024) segmentation stream and a brain mask was defined as the union of resulting gray and white matter segmentations. This brain mask was then used to substitute Freesurfer's skull stripping steps, specifically applying it to orig.mgz, writing the result as brainmask.mgz and brainmask.auto.mgz. After that the remaining stages (autorecon2 and autorecon3) of Freesurfers recon-all processing stream were run, using the highres flag and setting the mris_inflate iterations to 100 using an expert option file.

For subsequent use of the surface based HCP MMP 1.0 atlas within each individual subject, the segmented structural surfaces of each participant were brought into fsLR164k CIFTI space using ciftify_recon_all, which is part of the Python-based ciftify package (Dickie et al., 2019) with expert settings to register to the high-resolution 0.5 mm FSL MNI152_T1 template. Ciftify adapts the post-Freesurfer portion of the HCP minimal preprocessing pipeline (Glasser et al., 2013) to non-HCP acquired data and includes surface based alignment (Robinson et al., 2014) to HCP's fs_LR space.

Preprocessing of functional data

The DICOM images of each run were converted into a nulled and not nulled series. The first two volumes of each series were discarded to account for initial T1 relaxation effects. Motion correction across volumes and runs was performed using AFNI's 3dvolreg and was done separately for the nulled and not-nulled series. Volumes from the same run and with the same index were used as references for both and were determined as those that together had the lowest number of outliers in their voxels as determined by AFNI's 3dToucount. Finally, runs belonging to each type (alphabetization vs. remembering, action vs. non-action) were averaged.

Co-registration and calculation of cortical depths

A T1w image was calculated from all VASO images as follows. All runs were averaged and the inverse coefficient of variation (mean divided by variance) over all volumes (nulled and not-nulled) was computed. This procedure emphasizes the T1 contrast originating from different T1 weighting of nulled and non-nulled read-outs and makes it qualitatively similar to the processed MP2RAGE images. The resulting images were further processed using a custom pipeline that included background denoising using LAYNII (LAYNII's LN2_MP2RAGE_DNOISE; Huber et al., 2021) and bias field correction (SPM12). Next, the Freesurfer processed T1w image (from MP2RAGE acquisition) was registered to the VASO T1w image using ANTS. The registration consisted of 3 stages: rigid, followed by affine and finally non-linear SyN registration . The rigid and affine stages used mutual information as metric and the SyN stage (Avants et al., 2008) used cross-correlation. The resulting transformation was then applied to the Freesurfer gray and white matter surface coordinates, aligning the surfaces with the functional data. From these surfaces, voxel-wise cortical depths were computed using a custom algorithm. In effect, this algorithm calculated for every voxel the relative depth at which a transformation between gray and white matter boundary (under an equivolume model) would intersect with the voxel's center (https://github.com/dchaimow/vdfs; Chaimow et al., 2022). Superficial and deep layers were defined as voxels having a relative cortical depth of <0.5 and >0.5 respectively. In a control analysis that introduced a gap between layers these depth ranges were changed to <⅓ and >⅔. HCP MMP 1.0 atlas labels (Glasser et al., 2016) were transformed from fs_LR to each subject's native freesurfer surface space using HCP's

wb_command (label-resample with ADAP_BARY_AREA method) and the subjects surface transformations previously generated by cifify's recon-all stream. The labels were then projected into functional volume space using HCP's wb_command (label-to-volume-mapping with ribbon-constrained option) using the individual subjects' surfaces in functional volume space.

Definition of ROI

For ROI definition we combined atlas information with functional activation results. First, the BOLD trial averaging F-statistics from alphabetization vs. remembering and action vs. no action runs were averaged, sampled onto the cortical surface (wb_command -volume-to-surface-mapping with ribbon-constrained option) and smoothed (wb_command -metric-smoothing with FWHM = 3 mm).

This map was masked so that only activity within left p9-46v and all surrounding parcels remained. Given a specific threshold value (that was computationally obtained- see description below), the surface ROI was defined as the largest connected cluster that showed any overlap with the center parcel (left p9-46v).

Using an optimization algorithm (scipy brentq; Brent, 1973), we then estimated the threshold value that resulted in an ROI whose surface area (calculated with wb_command metric-weighted-stats) was closest to a specified target surface area. For the main analyses, this target surface area was chosen to be 100 mm^2 after qualitative comparison of the size of ROIs for different target surface areas with the size of ROIs presented in Supplementary Fig. 3 of Finn et al. (2019). Note that we are not able to quantify this as ROIs were not published in the original study. The final surface ROI was then projected into the corresponding gray matter of the EPI volume (using wb_command metric-to-volume-mapping with the ribbon-constrained option). ROIs were split into a superficial and deep layer by binning the voxels according to their relative depth.

Definition of alternative ROIs constrained by group analysis clusters

The individual subjects F-statistics surface maps calculated in 4.6 were transformed to fs_LR space and averaged across subjects. For each vertex a group t- and z-statistics was calculated by dividing the average value by its standard error and considering the number of subjects whose coverage included this vertex. The group averaged z-statistics map was then used to define clusters by first iteratively increasing the threshold until z=2.7. When the threshold increases, individual connected regions either decrease in size or split into multiple smaller regions. Clusters were defined by identifying all connected regions across various thresholds such that they were not further split under higher thresholds, and for each selecting their largest extent under their smallest possible threshold. Group clusters were then back projected into single subject surfaces and used as constraints for the procedure described in 4.6.

Control analysis using manually segmented laminar ROIs

ROIs consisting of a superficial and a deep layer were drawn manually on top of the T1 weighted VASO image. For that, we followed the description provided by Finn et al. (2019). The previously computed automatic ROIs were used to determine the location and extent, but gray matter segmentation and layer assignment was purely based on the T1 weighted VASO image. The T1 weighted MP2RAGE image was occasionally only consulted to provide additional contextual information about the local cortical geometry.

SNR calculation

Time-course SNR (tSNR) was calculated by dividing the mean by the standard deviation of each voxel's motion- and BOLD-corrected VASO time series in each individual run. tSNR values were then first averaged across runs and finally across all voxels within each subject's ROI.

Trial averaging

Different from Finn et al. (2019) the trial duration, which we kept identical to 32s, was not a multiple of our TR (3.7s). Therefore, we used finite impulse response modeling to estimate the average trial time course. We estimated raw signal trial averages separately for the nulled and non-nulled data using AFNIs 3dDeconvolve with the TENTzero model and a separate set of polynomial regressor up to order 5 to account for low temporal frequency drift. For the non-nulled time series, we shifted the model in time to account for the delay between nulled and non-nulled acquisition (1.5s), so that the resulting trial averages were temporally aligned with each other. In addition, this analysis gave us voxel-wise F-statistics that quantified the amount of variance explained by the trial structure of the experiment (for the nulled and not-nulled data, only the latter was used subsequently as a measure of BOLD task involvement). The nulled trial average was then divided by the non-nulled trial average resulting in a BOLD corrected VASO estimate while the non-nulled trial average was from here on considered as the corresponding BOLD estimate. Finally, the VASO and BOLD trial averages for each voxel were transformed to units of percent signal change by dividing by the constant predictor value (BOLD corrected in the case of VASO). Note that the VASO signal is expected to be inversely proportional to cerebral blood volume, therefore all VASO plots depict negative percent change which is interpreted as a positive activation.

Statistical analysis of trial period responses

Statistical analyses followed Finn et al. (2019). Trial averaged percent change time courses from all voxels belonging to the ROI were averaged according to cortical depth (deep vs. superficial) and condition. Contrast time courses were calculated as the alphabetization vs. remembering and action vs. non-action difference. Trial period responses were extracted by averaging time points 3, 4 (0-based) for the delay period (taking hemodynamic delay into account) and 5, 6 for the response period following the probe. Identical to Finn et al. (2019) we compared activity within layers and subsequently differential activity within trial periods using a series of two-way (2 x 2) repeated measures ANOVAs, testing for interaction effects using a threshold of $p<0.05$. To answer how the difference of activity between trial types depends on the layer and the contrast of interest:

1. For the delay period:
   factor 1: layer (superficial, deep)
   factor 2: contrast (alphabetization vs. remembering, action vs. non-action)
   Dependent variable: difference of activity during delay period according to contrast
2. For the response period:
   factor 1: layer (superficial, deep)
   factor 2: contrast (alphabetization vs. remembering, action vs. non-action)
   Dependent variable: difference of activity during response period according to contrast

We consider this set of ANOVAs (1. and 2.) to be our primary analyses of interest. These are the analyses that we have optimized our target sample size for.

To answer how the activity within layers depends on trial type and trial period:

1. For superficial layers and alphabetization vs. remembering runs:
   factor 1: trial type (alphabetization, remembering)
   factor 2: trial period (delay, response)
   dependent variable: activity during trial period
2. For superficial layers and action vs. non-action runs:
   factor 1: trial type (action, non-action)
   factor 2: trial period (delay, response)
   dependent variable: activity during trial period
3. For deep layers and alphabetization vs. remembering runs:
   factor 1: trial type (alphabetization, remembering)
   factor 2: trial period (delay, response)
   dependent variable: activity during trial period
4. For deep layers and action vs. non-action runs:
   factor 1: trial type (action, non-action)
   factor 2: trial period (delay, response)
   dependent variable: activity during trial period

Confidence interval bootstrapping analysis

1000 bootstrap samples of size 21 each were generated by randomly selecting subjects with replacement. For each sample an ANOVA analysis was run as described above (factors: layer and condition contrast). In addition, the direction of the mean layer difference (superficial>deep or deep>superficial) with respect to the alphabetization vs. remembering contrast (in case of delay period) or with respect to the action vs. non-action contrast (in case of response period) was calculated. This resulted in a distribution of ANOVA interaction effect (η_p^2) sizes. For each distribution a 95% confidence interval was calculated.