*We would like to thank all reviewers for their time and careful reading of the manuscript. A point-by-point list of the amendments we have made is provided below. All changes in the manuscript are highlighted in the annotated version.*

# Reviewer 1

##### Using a gradient-echo line-scanning technique, Heij et al. investigated laminar BOLD responses to local and surround stimuli in human V1. While the local stimuli in the RF center showed a stimulus driven response biased to the superficial layer, a characteristic draining vein effect in BOLD fMRI, contextual surround stimuli (especially the far surround) induced double peak responses in the superficial and deep layers. The authors interpreted the laminar pattern as consistent with descending feedback processes from higher cortical areas.

##### The line scanning approach has been used in animal fMRI, and it has never to be used to address neuroscience questions in humans. The results also seem to be consistent with previous electrophysiological study. I have some concerns about the interpretation of results and method details.

We thank the reviewer for careful reading of our manuscript and valuable input.

## Major points

##### 1. The authors claimed that the superficial-and-deep laminar response pattern is consistent with the electrophysiological study by Bijanzadeh et al. 2018. But the previous study only showed a latency difference across layers, there was no LFP different across layers after 150 ms. The stimulus duration in the current study is 2 seconds and measured sluggish BOLD responses, so overall there should be no difference across layers according to the previous study? Please explain this discrepancy.

The reviewer correctly points out that the Bijanzadeh, et al., (2018) study mainly refers to latency differences, whereas the results in the current study present magnitude differences. Our interpretation is based on the similarity in response *pattern* across cortical layers for the large annulus stimulus. In both the spiking activity (MUA) and BOLD, responses to such stimuli were reserved for superficial and deep layers. Although the BOLD response is slow, high-frequency modulations can be observed with sufficiently fast sampling1. This would make investigations of latencies across cortical depth fascinating. In the current study, we were unfortunately hampered by SNR limitations from an acquisition (repetition time) and experimental (number of trials) perspective, obscuring any latency effects. We have emphasized the difference between the Bijanzadeh, et al., (2018) study and the current work in the discussion and suggested strategies for fMRI that might bridge the gap further (p. 7):

“*One important distinction is, however, that Bijanzadeh, et al.16 mainly reported differences in latencies between stimulus-driven and contextual processes, whereas the current study reports on amplitude across cortical depth. The speed with which these signals are transmitted makes it complicated to capture using BOLD fMRI10,18,75, though faster acquisitions with more trials may be able to reveal such effects more robustly25,44,76,77.*”

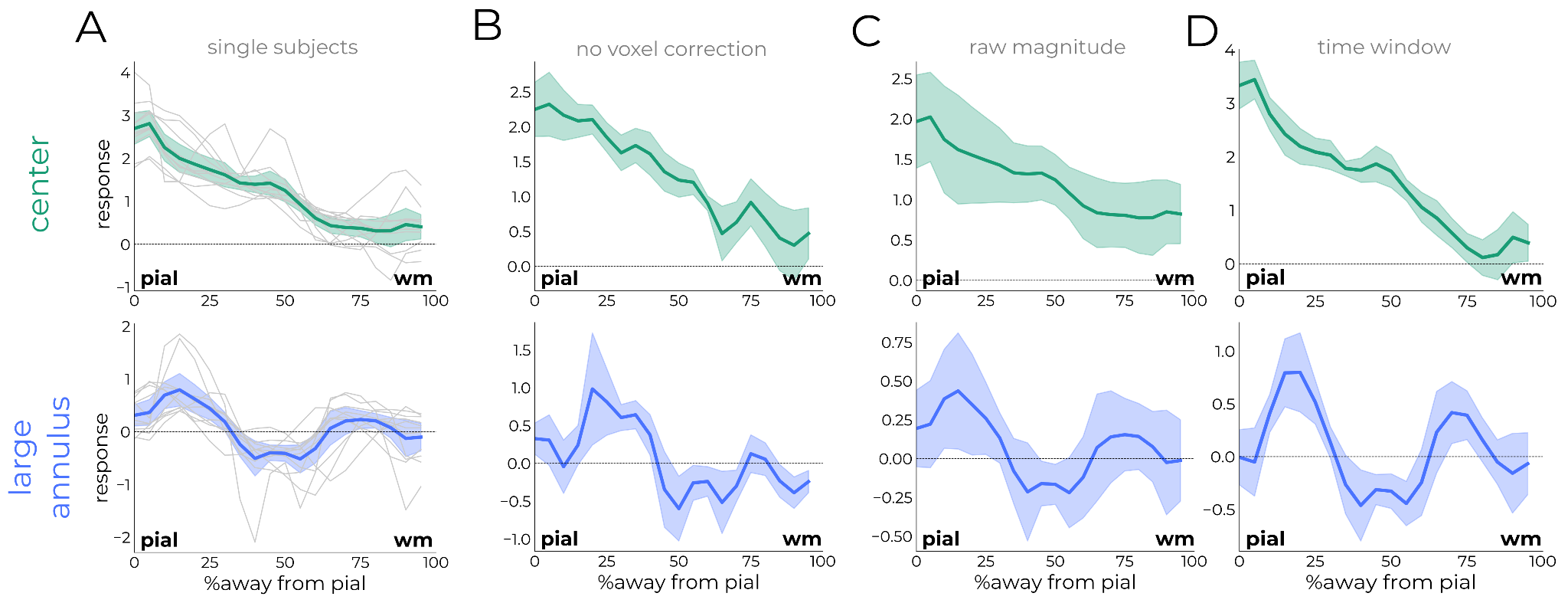
##### 2. The shapes of timecourse are very different in center and surround conditions (Fig. 2D). There was no explanation in the main text about the difference in shape of time course. Does this reflect a difference in SNR, or suggest different engagement in excitatory and inhibitory activity?

We indeed attribute the different responses to the center and surround conditions to a difference in neuronal populations that are stimulated. In the words of the reviewer, a different engagement in excitatory and inhibitory activity. The data from Figure 2D are an aggregate response over the entire cortical depth. This by itself shows a negative deflection for the large annulus condition but is uninformative regarding cortical depth-specific responses. We have now added a sentence regarding the implication that these stimuli elicit different responses in the text (p. 3):

“*Based on the size-tuning curve, we derived three stimuli centered on the pRF (Figure 2C) differentially eliciting stimulus-driven and contextual processes (Figure 1, middle panels). These stimuli indeed generated distinct responses collapsed over cortical depth, suggesting differential engagement of stimulus-driven and contextual processes (Figure 2D):*”

##### 3. It is known that negative BOLD will influence the shape of HRF. If the latter is true (reflect distinct neural dynamics), it is not appropriate to use the center response as a template to calculate the laminar profile of surround stimulus.

While the exact underlying mechanism of negative BOLD remains elusive, modeling approaches have shown that variations in negative BOLD response can be modeled by a negative weighting of the canonical HRF2,3. The negative weighting of the positive BOLD response is reflected by the negative deflection in middle depths (Figure 3, bottom-right panel) in the surround condition. That said, we also applied a more data-driven analysis approach similar to the approach during the quality assessment stage (Figure S11). Here, we used a 2s window between 5-7s after stimulus onset, showing similar results. These results have now been added to Figure S3 as panel D. To avoid issues with arbitrary selection of a time window, we opted to use the response to the center stimulus as references. Figure S3 now looks like this:



and we have now clarified this rationale in the text (p. 4):

“*To visualize this effect more clearly, we collapsed the responses across depth over time using the group response to the center stimulus averaged across depth as a template. This approach is favored over using a time window (as used in Figure S11) due to variability in response latency across participants.*”

##### 4. The line scanning technique has extremely high spatial resolution (0.25mm along the line). While the laminar results were presented as continuous profiles, the interpretations are based on a three-layer compartments of superficial, middle and deep. In fact, layer 1, 2/3 in the superficial layers, and the upper and lower part of deep layers have very different functions. The electrophysiological study showed different laminar patterns in the local surround (layer 2/3 and 5) and far surround (layer 1 and 6) conditions. Your data also suggest different laminar patterns in medium and far surround conditions. It will make much more sense to segment and discuss the results in 6-layer compartments.

We thank the reviewer for the suggestion. However, we prefer to refer to the data points as “cortical depths”, as we are unable to resolve individual layers of the cytoarchitecture using fMRI. Though we have high resolution, partial voluming and spatial hemodynamic pooling determine the actual resolution limit. To acknowledge these limitations, we refer to the three compartment model4.

Moreover, even though we discuss the results using the terminology from the 3-compartment model, the actual number of data points varies across subjects. Due to individual differences in local cortical thickness, we acquired 6-10 data points along the cortical patch of interest. Rather than averaging the signals to 3 compartments, we upsampled the data so that 20 data points covered the cortical patch (see Figure S5 for such an approach based on a 5-compartment model). To refrain from referring to specific layers (due to the reason outlined above), we use terms like *superficial* and *deep depths* (see also ref5 for discussion on this terminology). We have now clarified this rationale in the text (p. 4):

“*Because of the increased resolution in the laminar direction, this study has reduced statistical dependency across depth compared to previous studies that sampled 2-3 data points to 20 cortical depths28,32,33,56–58. As we are not measuring individual layers of the cytoarchitecture, we discuss the results in terms of the three compartment model (superficial, middle, and deep depths)14*.”

##### 5. The limitation of the draining vein effect of BOLD signals, which was obvious in your results, need to be discussed in the limitations. In your analysis (Fig.4 and 6), using a linear trend to account for the draining vein effect need better justifications. This is not consistent with the physiological model of vasculature.

The reviewer is correct to state that a linear trend is not consistent with the physiological model of vasculature. We included a linear trend for simplicity and due to the limited data points across cortical depth. Nevertheless, we included two control analyses accounting for vasculature in different ways. We have included the results in the section “*Modeling the laminar profile for large annuli*” (p. 5-6):

“*We evaluated the impact of draining vasculature in two ways. First, instead of using a linear component to account for the draining vein effect, we used a negative exponential component increasing towards the surface - better reflecting the macrovascular contribution (Figure S6). Second, we deconvolved the profiles across cortical depth using the vascular model from Markuerkiaga and colleagues65,69 (Figure S7). Based on histological data70 and vascular modeling67, they derived how each layer affects subsequent layers (see Table 1 in Marquardt, et al.69 or Figure 3F in Markuerkiaga, et al.65). The deconvolved profiles were then entered in the same model as described above. Neither of these vascular correction methods altered the observed activation differences or conclusions.*”

Moreover, we have added a substantial section to the discussion (section “*BOLD response reflects metabolic needs*”), highlighting the impact of the draining vein effect on the interpretation of laminar fMRI results (p. 8):

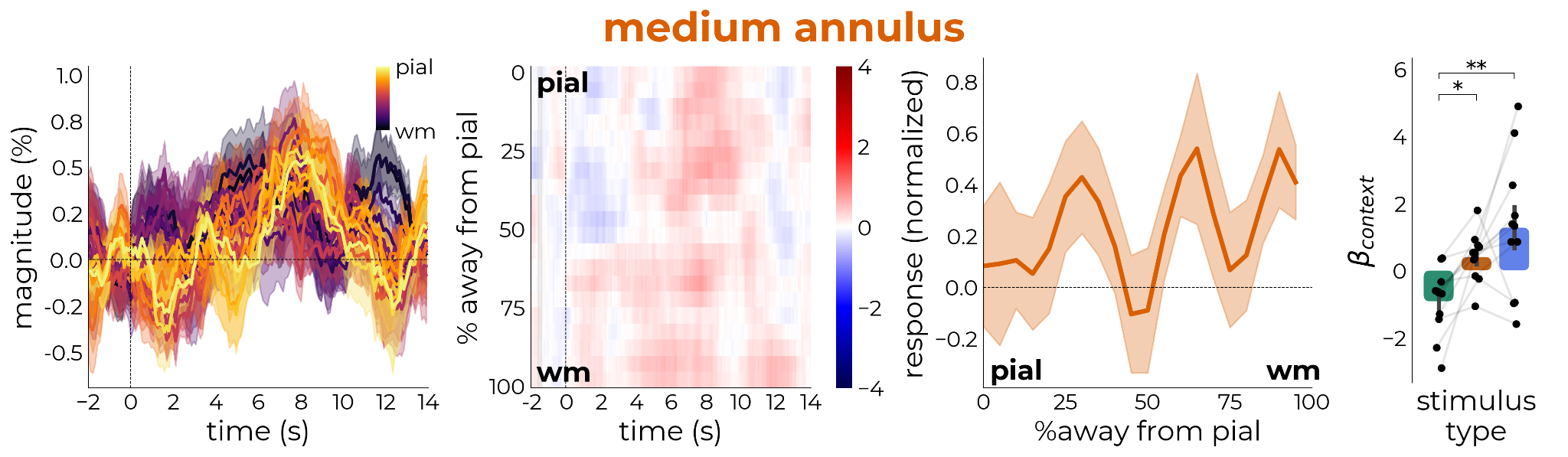
“***BOLD response reflects metabolic needs***

*The BOLD effect arises from a change in relative amounts of oxygenated and deoxygenated blood. It is therefore an indirect measure of neuronal activity, linked to particular neuronal firing patterns such as local field potentials (LFP)85–89, or spiking activity (in case of pRFs)90. Upon neuronal activation, metabolic demands trigger extra delivery of blood towards the site of activation91. Deoxygenated blood, which distorts the magnetic field due to unpaired iron atoms, is pushed away, resulting in increased signal92,93. Within the cortex, penetrating arteries branch off the pial network to supply the parenchyma of nutrients68,83, while veins drain the deoxygenated paramagnetic blood unidirectionally towards the pial surface94–96. The corollary of this process is twofold: 1) BOLD signal changes are larger at superficial cortical depths where the deoxygenated blood is pooled97–102 and 2) signals from deeper depths influence the signal at superficial depths, which has been referred to as the “draining vein effect”, “carry-over effects”, or “leakage problem”37,65–68. Strategies have since been developed to mitigate the effect of large pial veins, including acquisition103–106, analysis37,65,97,107–110, and experimental design51,111,112 (for further details, see these laminar fMRI reviews28,32,34,75,84,113–116).*

*Why, then, is the carry-over effect less visible for contextual stimuli? The center stimulus is designed to stimulate the center of the pRF, maximally eliciting the stimulus-driven processes. Such processes drive feedforward inputs into the middle layers of V110,14–16. The transmission of these inputs involves dense excitatory synapses and high-frequency spiking activity to propagate sensory input89,91,117. The metabolically demanding processes increase local oxygen consumption and blood flow, producing a robust BOLD response89,91,118. In contrast, the large annulus primarily engages lateral and feedback inputs facilitating contextual processes119. Contextual processes are modulatory, rather than driving, i.e., modulate existing activity rather than generate new action potentials4,5,120–123. These factors are thought to be metabolically less demanding compared to ascending feedforward inputs120,122,124,125. This difference in metabolic demand between these inputs might explain the presence of the draining vein effect in the stimulus-driven condition but not the contextual condition as well as overall magnitude differences.*”

##### 6. Results in Figure 4 and 6 are highly duplicated, please better organize your results.

We have merged figure 5 and 6 into a single figure to reduce duplication. This figure is now figure 5:



and the updated caption reads (p. 13):

“*Figure 5. Response evolution as time courses (first column) and depth-by-time (second column) for the medium annulus, a stimulus halfway between the center and large annulus. The third column represents the projection of the profile of the center stimulus averaged over cortical depth and participants (Figure 3, top right panel). Whereas the large annulus elicited responses close to superficial depths (~15% away from pial surface, Figure 3, bottom right panel), the medium annulus elicited responses in multiple sites across cortical depth. The last column represents the model outcome of all stimulus events. The medium annulus fell in between the center stimulus and large annulus (center and large annulus stimuli taken from Figure 4D). Shaded error represents 95% confidence intervals.*”

##### 7. In your modeling analysis of double peak response, how do you define the superficial and deep depths. I couldn’t find anywhere in the manuscript about the definition.

We thank the reviewer for pointing this out. The starting point of the Gaussian distribution representing the superficial peak was placed 25% and the distribution representing the deep peak was placed at 75%. Due to the addition of derivative components, the exact location of these peaks could vary across subjects. We have added this to the caption of Figure 4 (p. 12):

“*(A) Based on invasive anatomical and functional findings, we expected strong responses in superficial (peak at 25%) and deep depths (peak at 75%) (red shaded boxes).*”

and section “*Modeling the laminar profiles for large annuli*” (p. 5):

“*To quantify this effect, we defined a model with two Gaussian distributions representing the peaks at superficial (peak at 25%) and deep (peak at 75%) cortical depths (Figure 4AB).*”

# Reviewer 2

##### The authors have examined neural responses to a central stimulus, and compared that to the response to a surrounding stimulus. The authors use an innovative line-scanning MRI technique to obtain higher resolution BOLD information than typical studies. They show that V1 cortical voxels that preferentially respond to the center target have strongest responses in pial layers, decreasing as the voxel is closer to the white matter. This is consistent with previous studies of BOLD, but not consistent with single unit studies which suggest that middle layers should be activated by a ‘bottom up’ stimulus. Interestingly, an annulus centered at the same location produces responses in the superior and inferior layers, consistent with the presence of top-down signals, which extensive anatomical work shows are preferentially projecting to superior and inferior layers. The authors interpret that result in the context of the divisive normalization model. The work is interesting in that it suggests that there is preferential activation in areas of ‘top-down’ signals when a ‘surround’ stimulus is shown. This is consistent with divisive normalization.

We would like to thank the reviewer for their time and constructive comments. Note that with fMRI, BOLD responses are measured. These are an indirect measure at the end of a complex neurovascular coupling cascade most closely linked to gamma LFP6–10 and spiking activity11. The vasculature of the cortex is such that BOLD signals drain towards the surface, so that largest signal changes might be observed near the pial surface while most neural activation occurs in the middle layers. Several approaches to deal with this in laminar fMRI have been proposed (e.g., refs3,12–17) and are now discussed in more detail in this manuscript too. We refer to the points below for a more detailed answer.

## Major points

##### 1. The authors’ argument rests on figures showing the normalized response as in the right side of Figure 3. It is very unclear what is on the y axes of all these graphs. The description in the text was not clear, and is quoted here: “To visualize this effect more clearly, we projected the responses to the center stimulus averaged across cortical depth to the response profiles of the other stimuli. In other words, we aimed to find a scaling between 121 the cortical depth-dependent profiles and the participants’ average profile to the center stimulus 122 across cortical depth (Figure S3, green profile in bottom right panel, average).” The authors should work on extensively clarifying this method, and consider including an equation to explain what they did, in order to make this work reproducible.

We thank the reviewer for the suggestion and have made clarifications to the sections. The specific section now reads (p. 4-5):

“*To visualize this effect more clearly, we collapsed the responses across depth over time using the group response to the center stimulus averaged across depth as a template. This approach is favored over using a time window (as used in Figure S11) due to variability in response latency across participants. We calculated a weighted average to quantify how well the participants’ depth-dependent response profiles for other stimuli matches the profile for the group-averaged response to the center stimulus. This was achieved using the expression:*

*where hrfdepth represents the cortical depth-dependent response profiles for the other stimuli, and hrftemplate corresponds to the profile of the center stimulus (averaged over participants and cortical depth). The result provides a scaling factor that indicates how closely the depth-dependent responses of other stimuli align with the characteristic shape of the center stimulus profile (Figure S3, green profile in bottom-right panel labeled as “average”)*.”

Moreover, the code will be released upon publication, further improving the reproducibility of the work (see section “Data and code availability”, p. 10): <https://github.com/spinoza-centre/holeresponse>

##### 2. The authors are using a BOLD sequence, which is known to emphasize signals near the pia — in the limitations section, the authors should discuss how this bias would influence their results. (or potentially, why their experimental design might side-step that problem). They mention briefly in the discussion that the pattern of results they find for stimuli (strongest near pia) is consistent with the BOLD literature, but not the single unit literature. More extensive discussion of this general point, is needed.

We have added a section to the discussion (section “*BOLD response reflects metabolic needs*”), highlighting the impact of the draining vein effect on the interpretation of laminar fMRI results (p. 8).

“***BOLD response reflects metabolic needs***

*The BOLD effect arises from a change in relative amounts of oxygenated and deoxygenated blood. It is therefore an indirect measure of neuronal activity, linked to particular neuronal firing patterns such as local field potentials (LFP)85–89, or spiking activity (in case of pRFs)90. Upon neuronal activation, metabolic demands trigger extra delivery of blood towards the site of activation91. Deoxygenated blood, which distorts the magnetic field due to unpaired iron atoms, is pushed away, resulting in increased signal92,93. Within the cortex, penetrating arteries branch off the pial network to supply the parenchyma of nutrients68,83, while veins drain the deoxygenated paramagnetic blood unidirectionally towards the pial surface94–96. The corollary of this process is twofold: 1) BOLD signal changes are larger at superficial cortical depths where the deoxygenated blood is pooled97–102 and 2) signals from deeper depths influence the signal at superficial depths, which has been referred to as the “draining vein effect”, “carry-over effects”, or “leakage problem”37,65–68. Strategies have since been developed to mitigate the effect of large pial veins, including acquisition103–106, analysis37,65,97,107–110, and experimental design51,111,112 (for further details, see these laminar fMRI reviews28,32,34,75,84,113–116).*

*Why, then, is the carry-over effect less visible for contextual stimuli? The center stimulus is designed to stimulate the center of the pRF, maximally eliciting the stimulus-driven processes. Such processes drive feedforward inputs into the middle layers of V110,14–16. The transmission of these inputs involves dense excitatory synapses and high-frequency spiking activity to propagate sensory input89,91,117. The metabolically demanding processes increase local oxygen consumption and blood flow, producing a robust BOLD response89,91,118. In contrast, the large annulus primarily engages lateral and feedback inputs facilitating contextual processes119. Contextual processes are modulatory, rather than driving, i.e., modulate existing activity rather than generate new action potentials4,5,120–123. These factors are thought to be metabolically less demanding compared to ascending feedforward inputs120,122,124,125. This difference in metabolic demand between these inputs might explain the presence of the draining vein effect in the stimulus-driven condition but not the contextual condition as well as overall magnitude differences.*”

Moreover, we included two control analyses accounting for vasculature in different ways. We have included the results in the section “*Modeling the laminar profile for large annuli*” (p. 5-6):

“*We evaluated the impact of draining vasculature in two ways. First, instead of using a linear component to account for the draining vein effect, we used a negative exponential component increasing towards the surface - better reflecting the macrovascular contribution (Figure S6). Second, we deconvolved the profiles across cortical depth using the vascular model from Markuerkiaga and colleagues65,69 (Figure S7). Based on histological data70 and vascular modeling67, they derived how each layer affects subsequent layers (see Table 1 in Marquardt, et al.69 or Figure 3F in Markuerkiaga, et al.65). The deconvolved profiles were then entered in the same model as described above. Neither of these vascular correction methods altered the observed activation differences or conclusions.*”

## Minor points

##### 3. The focus of the paper, including the first sentence of the introduction emphasizes the concept of the PRF. That sentence reads “Neurons in early visual cortex respond to a particular part of the visual field, ie the population receptive field.” Of course, a given neuron responds to the cell’s receptive field. This definition seems to be a bit slippery, and is offputting for those of us who sometimes think about individual cells.

We apologize for the imprecise phrasing. We have now altered the first sentence to focus solely on the concept of the receptive field (RF). Later in the fMRI section, we describe that fMRI measures large populations of neurons resulting in the concept of the population receptive field (pRF). We believe this improves the separation between the fields of neurophysiology (measuring RFs) and neuroimaging (measuring pRFs).

The RF section now reads (p. 2):

“*Individual neurons in the early visual cortex respond to specific parts of the visual field, referred to as the receptive field (RF)1,2*”

and the pRF section (p. 3):

“*Rather than measuring activity from single neurons, fMRI samples from large groups of neurons (20.000 to 30.000 neurons per cubic millimeter in cerebral cortex47,48). Similar to single neurons, these groups (or populations) respond preferentially to specific parts of the visual field as well49,50, referred to as the population receptive field (pRF)51.*”

##### 4. Similarly, the text focuses on divisive normalization models — an excellent framework and set of models. But the expertimental design seems to focus on stimulation of the surround. There is no need to invoke the ‘normalization’ part here — why not just refer to the surround?

Indeed, the concept of surround is more important here. Nevertheless, divisive normalization (DN) models are central to understanding how neural circuits balance excitatory and inhibitory inputs across populations. The surround is a specific context within this framework, as it represents the “normalization pool” in divisive computations. By invoking normalization, we emphasize the mechanism (contextual integration) via the anatomical or spatial properties of the surround. We have altered the text throughout to emphasize the interest in contextual processes and the use of DN as operationalization.

##### 5. From a biological perspective normalization (or any model that takes into account the surround), “cannot be explained by ascending connections alone,” true. But it is not at all clear that there must be descending connections between neurons and neighboring areas — center-surround inhibition, for example, famously only requires lateral connections, not necessarily top-down. Your first paragraph in the introduction is trying to motivate why to look at the responses to the center and surround of the PRF, but its emphasis on motivating this because of DN models in particular seems forced. Why not motivate more generally as ‘context’?

We thank the reviewer for raising this important point. The reviewer is absolutely correct that center-surround inhibition can be explained by lateral connections that do not necessarily require descending connections. Our intention was not to imply that DN models exclusively depend on top-down pathways. Instead, we aimed to highlight the broader biological and computational context of how responses to center and surround stimuli can reflect contextual modulations. We have tailored the text such that the reader is aware that the topic of interest is contextual processing across cortical depth, with divisive normalization as operationalization because of its capability to unify disparate responses (p. 2).

“*The responses of these neurons are not solely reliant on stimulus-driven input; rather, they are complemented by context-related responses integrated from neighboring neurons3–6 (Figure 1, top panels). These responses can be modeled using surround-suppression or divisive normalization (DN). Specifically, DN has been shown to unify disparate responses including surround-suppression and compression6,7.*”

We have also added a section to the discussion that, while the results align with the idea of divisive normalization, the results can be interpreted in different frameworks (e.g., center-surround configurations, excitation/inhibition, or vasculature) (p. 7-8).

“*While the observed effects align with the idea of DN, alternative explanations may also account for or contribute to the observed responses. Similar to the studies previously discussed3,5,16, the results could be interpreted in many context-dependent processes such as lateral inhibition, of surround-suppression or differences in temporal dynamics between the stimuli. For example, the large annulus might engage lateral inhibitory circuits that suppress activity in the middle layers, while enhancing processing in superficial and deep layers4,10,78. Alternatively, the large annulus might recruit distinct populations of excitatory and inhibitory neurons across layers, with inhibition dominating in the middle layers and excitation driving responses in superficial and deep layers79,80. The large annulus might also elicit non-linear response patterns, where neurons in the middle layer are less responsive due to saturation or competitive suppression5,9. From a vascular perspective, the BOLD signal in superficial and deep layers may be amplified due to their proximity to larger veins and the influence of feedback inputs. These factors, combined with potential suppression of middle-layer activity (layer 4) by the large annulus, could account for the observed depth-specific BOLD response patterns57,81–83. However, different frameworks are unified by the specific circuits that underlie them: ascending, descending, and horizontal projections. Complementary approaches perturbing these processes may inform us further about how the cortex resolves them using the same circuit architecture75,84.*”

##### 6. The use of the term “Biased” is confusing in several places, and led this reader to misunderstand your goals on the first read of the paper. For example in figure 1’s description, it states “stimuli falling in the classical pRF are biased toward the activation pool.” A stimulus is presented by the experimenter, and thus cannot be biased? Does this statement mean the same thing as “Stimuli that fall within the classical pRF preferentially activate cells within the activation pool.” Further, you absolutely need to define the term “activation pool.”

We regret the confusion caused by our terminology. The reviewer is correct that we mean the same thing with “biased towards the activation pool” and “preferentially targeting the activation pool”. Given that we design each stimulus uniquely based on the participants’ target pRF estimates, we as experimenters are biasing the stimuli towards one or the other neuronal pool. To improve the clarity of the manuscript, we have defined the activation and normalization pool more explicitly in the introduction as well as changed the term “bias” to “to target the different populations of neurons facilitating stimulus-driven or contextual processes” (or alike) wherever applicable. The introduction now reads (p. 2):

“*Individual neurons in the early visual cortex respond to specific parts of the visual field, referred to as the receptive field (RF)1,2. The responses of these neurons are not solely reliant on stimulus-driven input; rather, they are complemented by context-related responses integrated from neighboring neurons3–6 (Figure 1, top panels). These responses can be modeled using surround-suppression or divisive normalization (DN). Specifically, DN has been shown to unify disparate responses including surround-suppression and compression6,7. In DN, the responses of neurons are modeled as the ratio of the activation and normalization pool8,9. The activation pool refers to the population of neurons predominantly driven by direct, feedforward input through ascending connections to the classical RF (blue arrows). The normalization pool consists of neurons that integrate contextual information from surrounding areas and contribute to balancing the responses of the activation pool via divisive normalization processes. While the activation pool primarily reflects direct input, the normalization pool mediates surround suppression and other forms of contextual modulation6,8,9. From a biological perspective, contextual computations cannot be explained by ascending connections alone. Such computations require descending and lateral connections between neurons and neighboring areas4,5,10–13.*”

and an updated caption for Figure 1 (p. 12):

“*Figure 1. Stimuli falling within the classical receptive field (RF) primarily elicit stimulus-driven processes mediated by ascending feedforward (FF) connections (red). These signals are then propagated through connections spanning multiple cortical layers (interlaminar) toward the surface and to neighboring regions. In contrast, stimuli on the flanks of the classical RF predominantly drive contextual processes, where information from neighboring areas is integrated via lateral and descending feedback (FB) connections (blue) to the superficial and deep layers. Cortical layers thus provide a unique window into the interplay between stimulus-driven and context-related computations. Using line-scanning fMRI (purple rectangle), we can capture depth-resolved responses to custom-designed stimuli weighted differentially towards stimulus-drive or context with unprecedented detail.*”

##### 7. There are quite a few grammar errors, which should be fixed. Their presence made the document more difficult to read. Examples are on page 8, line 186, “This work was done by… and present stimuli” should be “presents.” page 8, line 207 “constraint” should be “constrained.” These kinds of grammar errors may seem trivial but they do make an already complex document even harder to parse.

We thank the reviewer for pointing these out. We have made numerous changes to the manuscript to improve grammar and sentence flow throughout the text.

##### 8. Page 4, line 113 notes that spatial autocorrelation is not as much of a problem for the current work with 250micron resolution than for previous work that has 800 micron resolution and attempts to interpolate 20 depths in a ~2 mm thick cortical sheet. At least I think that is the point that the authors are making. It is hard to tell from the text saying “spatial autocorrelation is still less compared to interpolating 2-3 data points” — we need more context.

The reviewer’s conclusions are correct. Conventional laminar analyses interpolate 20 depths from profiles with 2-3 data points. This introduces substantial statistical dependency between depths because the interpolated values are not truly independent and rely heavily on assumptions about the underlying data structure (see main text p. 2-3). The current study benefits from a higher spatial resolution of 250µm in the laminar direction. With 250 μm resolution, the depth profile is sampled with a much higher density of data points (6–10), reducing the reliance on interpolation and improving the independence of data points. This results in a more accurate representation of laminar responses. We have clarified the section to reflect the explanation (p. 4):

“*The number of data points across the cortical depth of the target patch varied across participants due to differences in cortical thickness, ranging from 6 to 10 data points. To standardize the depth profiles for analysis, we regridded the data so that 20 evenly spaced points covered the cortical depth for all participants. Because of the increased resolution in the laminar direction, this study has reduced statistical dependency across depth compared to previous studies that sampled 2-3 data points to 20 cortical depths28,32,33,56–58*”

and have emphasized the assumptions underlying the data structure (p. 2-3):

“*This method requires extremely precise segmentation of the cortical sheet and accurate co-registration of anatomical and functional data34. It also assumes that the depth-dependent fMRI signal across cortical depth (i.e., the tangential direction) remains consistent within the ROI33; blood vessels oriented 90° with respect to the B0-field can cause significant signal reduction35–37. Given the highly complex folding pattern of the cortex38, this results in high variability across the ROI. These issues translate to data with high statistical dependency and low variability across cortical depth and require averaging over a large ROI32–34. Subpopulations of this ROI might exhibit variability in functional properties (e.g., RFs or center-surround configurations), reducing the specificity of the employed paradigm or acquisition scheme39,40.*”

##### 9. There is a strong emphasis on the relationship to previous data collected in animal models. But then there is very little about comparison to animal models in the discussion. The relevant sentence is “Similarly, they found that responses to stimuli biased toward normalization were constraint to superficial and deeper layers of the cortex; layers where descending connections carrying context-related information from neighboring regions terminate17,18. “ So the point seems to be that the authors have found that context-related information is focused on superficial and deep layers (Figure 3). This is also consistent with the anatomical literature, showing feedback projections in superficial and deep layers.

Indeed, given the striking similarities in experimental setups between the Bijanzadeh, et al. (2018) study and our work, we relate our findings to animal models and anatomical literature. The similarities in response patterns is encouraging for the field of laminar fMRI.

# Reviewer 3

##### In this study, the authors tested the divisive normalization model of receptive fields in the visual cortex by presenting subjects with a simple checkerboard stimulus within the central RF and an annulus within the surround of the RF. BOLD fMRI responses were measured across cortical depths, and different cortical depth profiles were observed in response to the center stimulus and a large-annulus stimulus: the center stimulus elicited a standard BOLD fMRI response that decreased monotonically from the pial surface to the white matter, suggesting a form of large-vein bias dominated the responses, whereas the large-annulus stimulus elicited a much weaker and unusual BOLD fMRI response that instead peaked somewhat below the pial surface, with a smaller peak in the lower depths and, surprisingly, a decrease in the middle cortical depths. This was interpreted as being consistent with descending pathway input to V1. An intermediate stimulus consisting of a medium-sized annulus, elicited an even weaker BOLD response with a puzzling cortical depth profile.

##### Overall the data are interesting, insofar as they are consistent across subjects. My main concerns are over the interpretation and the acquisition method.

We thank the reviewer for careful reading of our manuscript and valuable input. We agree with the reviewer that the medium annulus evokes a puzzling response, which we have aimed to address through vascular correction and by separating the participants based on stimulus configuration. We refer below for more details.

## Major points

##### 1. First, while the data are interesting, it is unclear how far they can be interpreted. While I understand that the experiment was motivated by an interest in divisive normalization, this is not the only explanation for the observed responses, and the conclusions should be toned down. For example, statements like the following should be reconsidered: “These results highlight the similarity of the cortical signature evoked by stimuli that drive context-related inputs compared to stimuli that drive stimulus-driven inputs.” I do understand how the large annulus is tapping into “context” in a way, however there are other possible explanations for these observations. This statement is far too general and is not supported by the data.

We explicitly relate to context-related processing as this captures a large range of processing, including lateral inhibition, surround suppression, or differences in temporal dynamics between the stimuli. We have added a paragraph to the discussion detailing alternative explanations. This section reads as follows (p. 7-8):

“*While the observed effects align with the idea of DN, alternative explanations may also account for or contribute to the observed responses. Similar to the studies previously discussed3,5,16, the results could be interpreted in many context-dependent processes such as lateral inhibition, of surround-suppression or differences in temporal dynamics between the stimuli. For example, the large annulus might engage lateral inhibitory circuits that suppress activity in the middle layers, while enhancing processing in superficial and deep layers4,10,78. Alternatively, the large annulus might recruit distinct populations of excitatory and inhibitory neurons across layers, with inhibition dominating in the middle layers and excitation driving responses in superficial and deep layers79,80. The large annulus might also elicit non-linear response patterns, where neurons in the middle layer are less responsive due to saturation or competitive suppression5,9. From a vascular perspective, the BOLD signal in superficial and deep layers may be amplified due to their proximity to larger veins and the influence of feedback inputs. These factors, combined with potential suppression of middle-layer activity (layer 4) by the large annulus, could account for the observed depth-specific BOLD response patterns57,81–83. However, different frameworks are unified by the specific circuits that underlie them: ascending, descending, and horizontal projections. Complementary approaches perturbing these processes may inform us further about how the cortex resolves them using the same circuit architecture75,84.*”

and the quoted sentence has been adjusted to the following to merely describe the observed response patterns (p. 6):

“*Together, these results show that stimuli designed to elicit contextual processes result in responses that differ from stimuli eliciting stimulus-driven processes through ascending feedforward connections.*”

##### 2. Also regarding the interpretation, the observed decrease in activity in middle cortical depths is somewhat counterintuitive and does not seem to fit the proposed model. Can the authors account for this?

We agree that this observation merits more space in the discussion section. While the decrease in activity in middle cortical depths may appear counterintuitive at first, it is consistent with certain biological and computational mechanisms (e.g., inhibitory feedback or contextual modulation). Middle cortical depths receive strong feedforward excitatory input (e.g., from ascending connections via the LGN). They are also influenced by inhibitory feedback or lateral connections from neighboring neurons or higher-order areas. Stimuli targeting the normalization pool, such as the large annulus, may preferentially modulate neurons in superficial and deep layers via feedback and lateral connections. These feedback layers interact with the middle layer through excitatory or inhibitory interneurons, potentially reflecting a redistribution of neural processing across cortical depth.

##### 3. Regarding the acquisition, in the introduction the authors state that line-scanning fMRI provides an advantage over conventional methods in that it provides ultra-high spatial resolution of 250 microns along the line. (As an aside, it would be helpful if the authors could include the other two dimensions of this acquisition up front in the Introduction, rather than postponing this to the Methods section.) However, the main results divide the cortex into three depth bins, superficial, middle, and deep, as indicated in Figure 1. Line-scanning fMRI also comes with major disadvantages, such as an inability to sample from multiple locations, as mentioned in the Discussion, and more sensitivity to motion and prescription errors. Could this study have been performed more reliably with a more conventional acquisition? The manuscript alludes to reduced partial voluming and effects from large veins, however it is unclear whether the 0.8 mm isotropic resolution mentioned in the Introduction would not have sufficed. Given that this acquisition approach precludes sampling both the targeted neurons whose RFs overlap with the center stimulus and nearby neurons whose RFs overlap with, say, the medium and large annulus stimuli, a potentially informative control experiment is impossible. While the authors were careful regarding the prescription of the line and aligning this with the stimulated region, perhaps it would have been preferable to sample from the entire stimulated region of V1 and performed a less potentially circular ROI selection. In short, the authors should more convincingly justify this unusual acquisition, and demonstrate that a 0.8 mm resolution would not suffice, given the constraints the line-scanning fMRI approach placed on the experimental design and ROI selection.

The reviewer expressed concerns regarding our acquisition strategy. We have described the sequence, placement, and motivation to use line-scanning rather than 0.8mm isotropic resolutions elaborately in previous work18–21. Here, we opted to use the available scan time to maximize SNR by only acquiring line-scanning data, enabling us to answer neuroscientific questions at specific locations in visual cortex. Moreover, center-surround organizations are unique per cortical location. A whole brain acquisition contains cortical locations with a mixture of center-surround stimulation and little clean distinction. No center-surround stimulation can exist for the whole of V1.

We have strengthened our motivation for this acquisition in the introduction with the following section (p. 2-3):

*“High-resolution functional MRI (fMRI) permits access to different layers of the cortex24,25. This cortical depth-resolved fMRI enables researchers to study the flow of information through the cortex26,27. Contemporary fMRI acquisitions typically sample the cortex with ~0.8 millimeter isotropic resolutions24,25,28–31, which is still large compared to the spatial dimensions of the cortical layers: at this resolution, the entire cortical depth is covered by about 2–3 data points (voxels). To infer underlying laminar specific responses, data from a region-of-interest (ROI) is typically upsampled28,32,33.* *This method requires extremely precise segmentation of the cortical sheet and accurate co-registration of anatomical and functional data34. It also assumes that the depth-dependent fMRI signal across cortical depth (i.e., the tangential direction) remains consistent within the ROI33; blood vessels oriented 90° with respect to the B0-field can cause significant signal reduction35–37. Given the highly complex folding pattern of the cortex38, this results in high variability across the ROI. These issues translate to data with high statistical dependency and low variability across cortical depth and require averaging over a large ROI32–34. Subpopulations of this ROI might exhibit variability in functional properties (e.g., RFs or center-surround configurations), reducing the specificity of the employed paradigm or acquisition scheme39,40. To improve spatial and functional precision across cortical depth, acquisitions with higher resolutions are required. This can be achieved by using anisotropic voxel dimensions and/or reducing spatial coverage33, while ensuring that the ROI is sampled at high resolution perpendicular to the cortical sheet41–43.”*

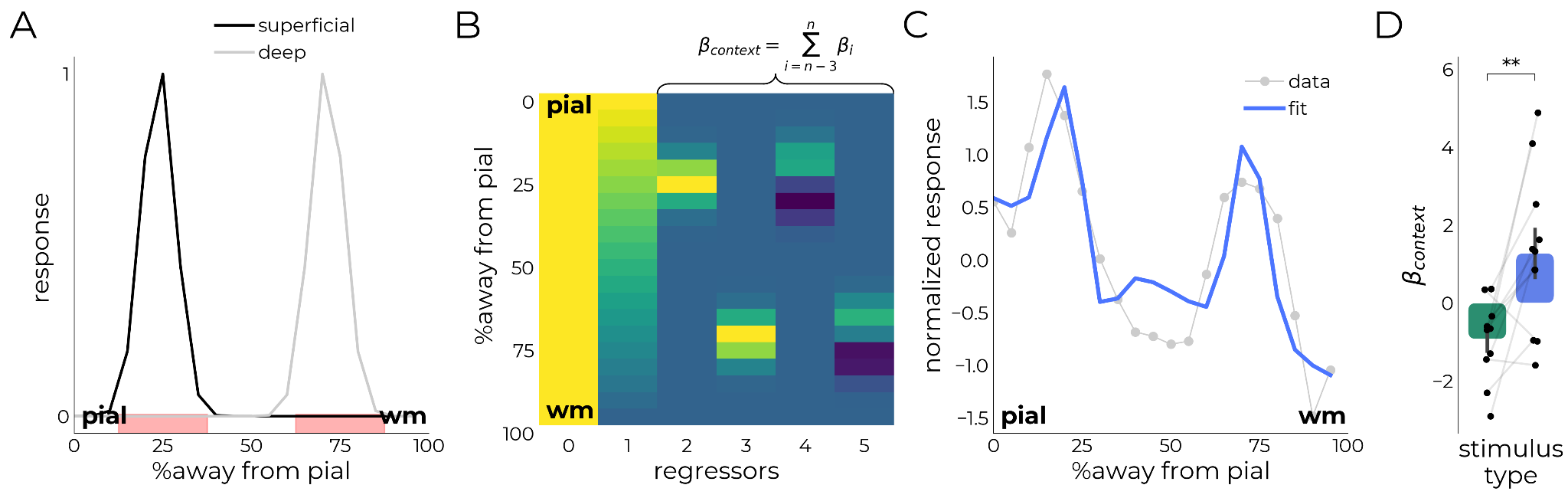
This section is followed by the paragraph introducing line-scanning that mitigates most of these problems at the cost of spatial coverage. We would welcome efforts to replicate these results across larger patches of visual cortex using more conventional imaging approaches.

In the description of line-scanning, the remaining voxel dimensions are now also mentioned (p. 3):

“*Line-scanning uses a single slice (2.5 millimeter thick) where the signal outside the region of interest is suppressed using 2 saturation pulses (4 millimeter gap)44,45 (Figure 1, purple beam).*”

##### 4. Finally, the model that is fit to the data does not seem mechanistic rather it simply captures the BOLD fMRI response amplitude as a function of cortical depth. While this simple model can be used to show that the data as a function of a cortical depth is approximated better by a linear trend plus two gaussian functions peaking in the superficial and deep cortical depths than it is by a simple linear trend, it is unclear how informative this is since this can be appreciated by eye thus this quantitative analysis may not provide much value.

We acknowledge that the current model is descriptive in nature, primarily aimed at capturing and summarizing the depth-dependent profiles in a straightforward and interpretable manner. We do note, however, that the main comparison is not between linear vs biphasic model. Instead, we compare the beta-values from the model components representing contextual processing between conditions. To improve clarity and avoid the impression that the comparison is linear vs biphasic model, we have added two panels (A and B) to Figure 4.



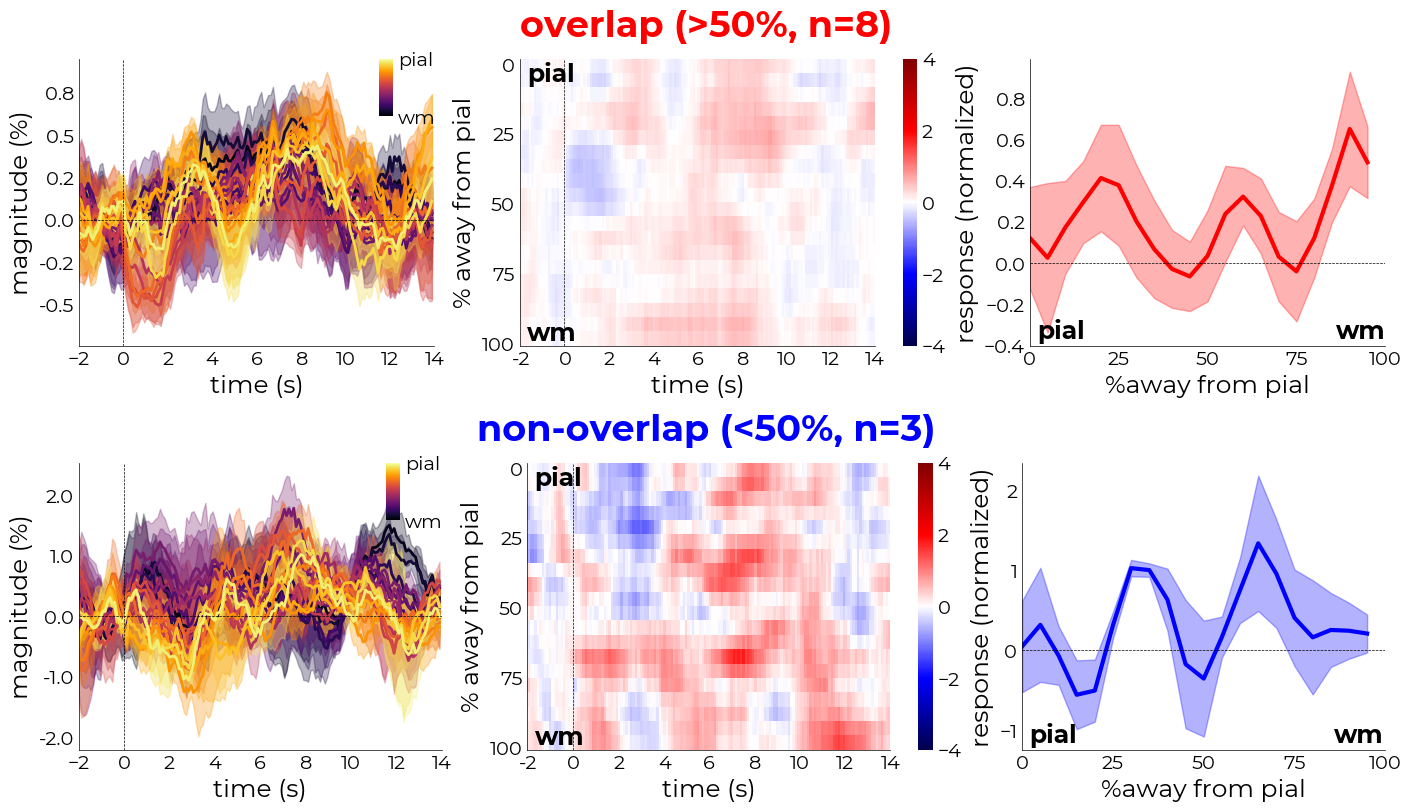
This approach reflects an objective framework for validating observations across participants and groups thereby enabling formal statistics. Crucially, each participant has been presented with a unique set of stimuli, at a unique cortical site. Thus, this approach allows for group analyses in an experimental design where everything is uniquely tailored to each participant.

##### 5. That said, it is unclear how well this bi-phasic model fit the medium-sized annulus data that appears to exhibit, surprisingly, three peaks across layers. While the large- and medium-sized annuli both yielded peaks across cortical depths, the BOLD response amplitudes are quite small and so there is some concern that the trends may reflect some structured noise. It seems that both the peaks seen in the large-annulus stimulus responses and those seen in the medium-annulus stimulus responses cannot all be in the superficial and deep layers, since the peaks in the medium-annulus data appear to be shifted downwards to deeper depths, and so the medium-sized annulus data are difficult to reconcile with the model and with the other data reported.

The observation of multiple peaks in the medium-sized annulus data across cortical depth is indeed noteworthy. While the biphasic model captures the expected two peaks (superficial and deep layers) observed in the large annulus responses, the additional peak for the medium annulus suggests involvement of more complex dynamics. These may arise from its intermediate spatial configuration, tapping into a mixture of activation and normalization pools (see stimulus configurations for different participants, Figure S1). This might activate a unique combination of lateral, feedback, and local processing circuits. Combined with the limited SNR of our acquisition (e.g., number of trials and noise levels native to line-scanning), these responses may not have been resolved properly. Some of this variability was captured by correcting for carry-over effects (Figure S7). Nevertheless, we do regard the responses in the large annulus condition to reflect underlying neuronal processing given its clear pattern even in the absence of formal quantification through the biphasic model. We have adjusted the discussion to highlight the complications of the model. The section now reads as follows (p. 9):

“*Lastly, future work could develop more advanced definitions of contextual processing. This work operationalized this by using a biphasic model representing termination sites of feedback projections. While this model is relatively simple and allows for some degree of interpretation, it does not provide a mechanistic account*.”

We have also employed a method to clean up the responses to the medium annulus using the zero-crossings of the 1D target pRFs (Figure 1B). Doing so directly relates computational modeling (pRF modeling) to experimental design (stimulus configuration); a unique feature in fMRI studies. We calculated the percentage of pixels of the medium annulus that resided outside the zero-crossing, reflecting a more “true” near-surround and eliciting less stimulus-driven responses. Although the number of participants where there was limited overlap is small (n=3), the response pattern shows clear peaks around but avoiding the middle depths. Such a pattern is in accordance with the response to the “near”-stimulus in the Bijanzadeh, et al. (2018) study, potentially reflecting lateral processing. We have added this figure to the supplement as Figure S9:



With an additional discussion on the mixed nature of the medium annulus response (p. 6):

“*Indeed, the response profile deviated from the other two stimuli. The time courses are noisier and the profile across depth less specific (Figure 5, first and second panel). This could be due to vascular carry-over effects obscuring true responses (Figure S7) or because the stimulus configuration differed across participants (Figure S1, Figure S8); for some participants, the medium annulus may have hit stimulus-driven populations, whereas for others it was fully in the surround, driving contextual processes. The response profile of the medium annulus showed widespread activation with peaks closer to middle depths compared to the large annulus (Figure 5, third panel), potentially reflecting lateral processing.*”

## Minor points

##### 6. The opening sentence states that “Neurons in early visual cortex respond to a particular part of the visual field, i.e., the population receptive field.” Conventionally, the responses of individual neurons are conceptualized as a receptive field, whereas the concept of a \*population\* receptive field (introduced by the senior author) was meant to represent a population of neurons, e.g., the neurons within an fMRI voxel. Perhaps this was written mistakenly. The authors may wish to clarify this, especially given that this imprecise statement appears in the first sentence of the manuscript.

We have altered the first sentence to focus solely on the concept of the receptive field (RF). The sentence now reads (p. 2):

“*Individual neurons in the early visual cortex respond to specific parts of the visual field, referred to as the receptive field (RF)1,2.*”

Later in the fMRI section, we describe that fMRI measures large populations of neurons resulting in the concept of the population receptive field (pRF). This section now reads (p. 3):

“*Rather than measuring activity from single neurons, fMRI samples from large groups of neurons (20.000 to 30.000 neurons per cubic millimeter in cerebral cortex47,48). Similar to single neurons, these groups (or populations) respond preferentially to specific parts of the visual field as well49,50, referred to as the population receptive field (pRF)51.*”

We believe this improves the separation between the fields of neurophysiology (measuring RFs) and neuroimaging (measuring pRFs).

##### 7. Also in the introduction, when the authors propose that divisive normalization cannot be implemented with ascending pathways alone, therefore descending pathways are required, this logic leaves out the possibility of horizontal or lateral pathways. In other words, if the only two options are ascending and descending, then if ascending is ruled out then descending is the only remaining option, however there are three options, ascending, descending and horizontal.

We thank the reviewer for raising this important point. The reviewer is absolutely correct that center-surround inhibition can be explained by lateral connections that do not necessarily require descending connections. Our intention was not to imply that DN models exclusively depend on top-down pathways. Instead, we aimed to highlight the broader biological and computational context of how responses to center and surround stimuli can reflect contextual modulations. We have tailored the text such that the reader is aware that the topic of interest is contextual processing across cortical depth, with divisive normalization as operationalization because of its capability to unify disparate responses (p. 2).

“*The responses of these neurons are not solely reliant on stimulus-driven input; rather, they are complemented by context-related responses integrated from neighboring neurons3–6 (Figure 1, top panels). These responses can be modeled using surround-suppression or divisive normalization (DN). Specifically, DN has been shown to unify disparate responses including surround-suppression and compression6,7.*”

We have also added a section to the discussion that stresses that, while the results align with the idea of divisive normalization, the results can also be interpreted in different frameworks (e.g., center-surround configurations, excitation/inhibition, or vasculature) (p. 7-8).

“*While the observed effects align with the idea of DN, alternative explanations may also account for or contribute to the observed responses. Similar to the studies previously discussed3,5,16, the results could be interpreted in many context-dependent processes such as lateral inhibition, of surround-suppression or differences in temporal dynamics between the stimuli. For example, the large annulus might engage lateral inhibitory circuits that suppress activity in the middle layers, while enhancing processing in superficial and deep layers4,10,78. Alternatively, the large annulus might recruit distinct populations of excitatory and inhibitory neurons across layers, with inhibition dominating in the middle layers and excitation driving responses in superficial and deep layers79,80. The large annulus might also elicit non-linear response patterns, where neurons in the middle layer are less responsive due to saturation or competitive suppression5,9. From a vascular perspective, the BOLD signal in superficial and deep layers may be amplified due to their proximity to larger veins and the influence of feedback inputs. These factors, combined with potential suppression of middle-layer activity (layer 4) by the large annulus, could account for the observed depth-specific BOLD response patterns57,81–83. However, different frameworks are unified by the specific circuits that underlie them: ascending, descending, and horizontal projections. Complementary approaches perturbing these processes may inform us further about how the cortex resolves them using the same circuit architecture75,84.*”

##### 8. While the lines-scanning fMRI method was used to increase resolution, BOLD fMRI still measures hemodynamics and therefore there must be some limits to the resolving power even with such small voxels. This issue has been addressed at length in other high-resolution fMRI studies, and there are ways to extract laminar information from these data, but still the reader may benefit from some discussion of this important limitation.

We have added a section to the discussion (section “*BOLD response reflects metabolic needs*”), highlighting the impact of the draining vein effect on the interpretation of laminar fMRI results (p. 8).

“***BOLD response reflects metabolic needs***

*The BOLD effect arises from a change in relative amounts of oxygenated and deoxygenated blood. It is therefore an indirect measure of neuronal activity, linked to particular neuronal firing patterns such as local field potentials (LFP)85–89, or spiking activity (in case of pRFs)90. Upon neuronal activation, metabolic demands trigger extra delivery of blood towards the site of activation91. Deoxygenated blood, which distorts the magnetic field due to unpaired iron atoms, is pushed away, resulting in increased signal92,93. Within the cortex, penetrating arteries branch off the pial network to supply the parenchyma of nutrients68,83, while veins drain the deoxygenated paramagnetic blood unidirectionally towards the pial surface94–96. The corollary of this process is twofold: 1) BOLD signal changes are larger at superficial cortical depths where the deoxygenated blood is pooled97–102 and 2) signals from deeper depths influence the signal at superficial depths, which has been referred to as the “draining vein effect”, “carry-over effects”, or “leakage problem”37,65–68. Strategies have since been developed to mitigate the effect of large pial veins, including acquisition103–106, analysis37,65,97,107–110, and experimental design51,111,112 (for further details, see these laminar fMRI reviews28,32,34,75,84,113–116).*

*Why, then, is the carry-over effect less visible for contextual stimuli? The center stimulus is designed to stimulate the center of the pRF, maximally eliciting the stimulus-driven processes. Such processes drive feedforward inputs into the middle layers of V110,14–16. The transmission of these inputs involves dense excitatory synapses and high-frequency spiking activity to propagate sensory input89,91,117. The metabolically demanding processes increase local oxygen consumption and blood flow, producing a robust BOLD response89,91,118. In contrast, the large annulus primarily engages lateral and feedback inputs facilitating contextual processes119. Contextual processes are modulatory, rather than driving, i.e., modulate existing activity rather than generate new action potentials4,5,120–123. These factors are thought to be metabolically less demanding compared to ascending feedforward inputs120,122,124,125. This difference in metabolic demand between these inputs might explain the presence of the draining vein effect in the stimulus-driven condition but not the contextual condition as well as overall magnitude differences.*”

##### 9. Why were the multi-echo BOLD fMRI data combined with a sum-of-squares operation? This would seemingly bias the combination to echoes with higher signal levels, which would correspond to shorter echo times with less BOLD weighting.

In previous research19, we explored several strategies for the combination of multi-echo line-scanning data. These strategies included echo-combination based on sum-of-squares, weighted by SNR, and weighted by T2\*. That work showed higher maximum and mean t-stats as well as higher overall tSNR for the sum-of-squares approach. While shorter echoes contribute less BOLD weighting, they are crucial for maintaining signal stability and improving SNR. We therefore opted for a sum-of-squares operation. We have added this motivation to the text (p. 22):

“*Multi-echo data were subsequently combined using a sum-of-squares operation to maximize signal stability and contrast-to-noise (CNR)46.*”

##### 10. Figure 1 caption: “intralaminar” should be “interlaminar” I believe, although both terms are somewhat awkward.

The reviewer is correct that “interlaminar” is more appropriate than “intralaminar”, as the connections we describe involve interactions between cortical layers rather than within a single layer. We also recognize that both terms can be somewhat technical and may feel awkward. To enhance clarity, we propose revising the terminology to explicitly describe the nature of the connections without relying on specialized terms. The particular sentence of the caption now reads (p. 12):

“*These signals are then propagated through connections spanning multiple cortical layers (interlaminar) toward the surface and to neighboring regions.*”

##### 11. It is odd that the magnetic field strength of the MRI scanner is not mentioned. I assume that this study was conducted at 7 Tesla, and “7T” is a keyword, so perhaps the reader can deduce this. Still, I recommend stating this explicitly.

We thank the reviewer for pointing this out. We have added a line to the section “*Data acquisition*” section, stating that all acquisitions were indeed performed using a 7T MRI machine. The section now reads (page 22):

“*The workflow includes two separate scan sessions typically acquired on different days. The first session is dedicated to the acquisition of anatomical information and whole-brain population receptive field (pRF) estimation (see ref43 for acquisition and experimental paradigm). In the second session, we perform our functional line-scanning experiment, targeting a specific location on the cortical surface. All acquisitions were performed on a Philips Achieva 7T MRI system.*”

##### 12. As stated above, it would be helpful if the authors could include the other two dimensions of the linescanning fMRI acquisition up front in the Introduction and perhaps also the Abstract, rather than postponing this to the Methods section. The 4-mm width of the linescan acquisition is surprisingly coarse and this should be stated more transparently.

The “*Data acquisition*” section describing the line-scanning acquisition contains information on the other dimensions (p. 22):

“*A gap of 4 mm between the two OVS bands was used, resulting in a nominal resolution for the line of 4×2.5×0.25 mm3, with 0.25 mm in the laminar direction.*”

Thus, the line-scanning acquisition is based on a single slice of 2.5 mm thick, with 4 mm between the two saturation bands, and 0.25 mm in the frequency encoding direction. To highlight these characteristics further, we have adjusted Figure 1 to contain more specific information on the dimensions (p. 3) and added these to the main text (p. 3):

“*Line-scanning uses a single slice (2.5 millimeter thick) where the signal outside the region of interest is suppressed using 2 saturation pulses (4 millimeter gap)44,45 (Figure 1, purple beam). Thus, spatial coverage is sacrificed to sample responses from a specific patch of cortex with ultra-high spatial resolution (250 µm along cortical depth)44–46.*”

References

1. Lewis, L. D., Setsompop, K., Rosen, B. R. & Polimeni, J. R. Fast fMRI can detect oscillatory neural activity in humans. *PNAS* **113**, E6679–E6685 (2016).

2. Jorge, J., Figueiredo, P., Gruetter, R. & van der Zwaag, W. Mapping and characterization of positive and negative BOLD responses to visual stimulation in multiple brain regions at 7T. *Hum. Brain Mapp.* **39**, 2426–2441 (2018).

3. Kay, K., Jamison, K. W., Zhang, R. Y. & Uğurbil, K. A temporal decomposition method for identifying venous effects in task-based fMRI. *Nat. Methods* **17**, 1033–1039 (2020).

4. Felleman, D. J. & van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

5. Polimeni, J. R., Renvall, V., Zaretskaya, N. & Fischl, B. Analysis strategies for high-resolution UHF-fMRI data. *NeuroImage* **168**, 296–320 (2018).

6. Ogawa, S., Lee, T. M., Kay, A. R. & Tank, D. W. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *PNAS* **87**, 9868–9872 (1990).

7. Ogawa, S. *et al.* Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys. J.* **64**, 803–812 (1993).

8. Logothetis, N. K. The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal. *J. Neurosci.* **23**, 3963–3971 (2003).

9. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).

10. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).

11. Klink, P. C., Chen, X., Vanduffel, W. & Roelfsema, P. R. Population receptive fields in nonhuman primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. *eLife* **10**, e67304 (2021).

12. Chen, G., Wang, F., Gore, J. C. & Roe, A. W. Layer-specific BOLD activation in awake monkey V1 revealed by ultra-high spatial resolution functional magnetic resonance imaging. *NeuroImage* **64**, 147–155 (2013).

13. Olman, C. A. *et al.* Layer-Specific fMRI Reflects Different Neuronal Computations at Different Depths in Human V1. *PLoS One* **7**, e32536 (2012).

14. Panchuelo, R. M. S., Schluppeck, D., Harmer, J., Bowtell, R. & Francis, S. Assessing the Spatial Precision of SE and GE-BOLD Contrast at 7 Tesla. *Brain Topogr.* **28**, 62–65 (2015).

15. Yu, X. *et al.* Direct imaging of macrovascular and microvascular contributions to BOLD fMRI in layers IV–V of the rat whisker–barrel cortex. *NeuroImage* **59**, 1451–1460 (2012).

16. Markuerkiaga, I., Barth, M. & Norris, D. G. A cortical vascular model for examining the specificity of the laminar BOLD signal. *NeuroImage* **132**, 491–498 (2016).

17. Havlicek, M. & Uludağ, K. A dynamical model of the laminar BOLD response. *NeuroImage* **204**, 1–44 (2020).

18. Raimondo, L. *et al.* A line through the brain: implementation of human line-scanning at 7T for ultra-high spatiotemporal resolution fMRI. *J. Cereb. Blood Flow Metab.* **41**, 2831–2843 (2021).

19. Raimondo, L. *et al.* Robust high spatio-temporal line-scanning fMRI in humans at 7T using multi-echo readouts, denoising and prospective motion correction. *J. Neurosci. Methods* **384**, 109746 (2023).

20. Heij, J. *et al.* A selection and targeting framework of cortical locations for line-scanning fMRI. *Hum. Brain Mapp.* **44**, 5471–5484 (2023).

21. Raimondo, L. *et al.* Towards functional spin-echo BOLD line-scanning in humans at 7T. *Magn. Reson. Mater. Phy.* **36**, 317–327 (2023).