nature portfolio

Peer Review File

Spatial attention selectively alters visual cortical representation during target anticipation

Corresponding Author: Dr Ekin Tünçok

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this study, Tünçok et al. investigated how covert spatial attention impacts both spatial encoding properties and spatial activation patterns in retinotopic cortex. Inside the scanner, participants performed a visual psychophysics task in which they attended to either one or four peripheral locations to perform an orientation discrimination task at threshold (estimated based on attend-four trials). During the interval between attention cue and target array onset, a bar-shaped 'mapping' stimulus appeared at a random location, allowing for estimates of visual population receptive field (pRF) properties during each attention condition. Behaviorally, they found that sensitivity was highest for trials with a valid cue to one of four possible target locations as compared to trials with an uninformative cue requiring distributed attention across all target locations. Looking at the temporal dynamics of attentional modulation in BOLD signal from spatial attention, attentional enhancements spanned the interval from cue to target, suggesting they were sustained in anticipation of the target array. Furthermore, pretarget BOLD responses showed a baseline shift in visual field maps, where responses were more positive at the attended location compared to unattended locations relative to the distributed attention condition. pRF centers also shifted towards the attended location (mostly).

Overall, this manuscript is a well-written and thorough exploration of how focal and distributed covert spatial attention impact task performance and neural processing of visual stimuli. However, there are some important issues that preclude a positive recommendation at this time. It seems that the primary novelty of this study compared to many previous studies comparing pRF properties across manipulations of covert spatial attention is the use of a distributed condition, which the authors treat throughout as a neutral baseline. This condition is maybe the most interesting part of the paper, but it's hard to fully understand the results of this condition on their own without the much more 'typical' baseline often used in pRF studies (including in most mapping studies by the senior author's lab): the attend-fixation (or, sometimes, attend-mapping-stimulus) condition. This and other important issues are elaborated below:

Major comments:

1. Behavioral task: d' is very high for the valid trials of the focused cue condition, and very low (basically at chance for several participants) for the invalid cue condition. (a) does this mean that participants were only attending the cued location, and essentially ignoring the uncued locations? It would appear so, at least for some participants. I don't think this would substantially impact the interpretation of the neural results, but it may be worth considering. (b) is it possible that participants learned that the discrimination task was much easier on focused trials (on average), and so weren't as engaged as they were on distributed cue trials? The psychophysics are elegant and do an overall nice job of showing the impact of focused attention on behavioral discrimination performance, but one consequence of the equated discrimination thresholds between conditions is that the task is easier/harder across conditions of interest, which could plausibly lead to differences in fMRI signals.
2. 'Suppression': throughout the manuscript, the authors discuss their results from the focused-cue condition as compared to the distributed-cue condition as evidence for 'suppression' of activation, etc, due to focused attention. Why is this suppression relative to distributed attention, rather than distributed attention is enhancement relative to focused attention

elsewhere? I've thought carefully about the Supp Figure (3) and discussion paragraph discussing this issue, and I can't agree with the authors' argument that distributed attention is a sensible baseline condition for fMRI studies. And this gets at a central issue identified in the summary above: it's hard to compare these results to the many previous studies that have often used a more typical attend-fixation condition against which focused attention results have been compared. Perhaps the authors are right and those comparisons are incorrect for some reason - the point stands that the literature at present is based primarily on those comparisons, and so it seems important to at least acknowledge that the terminology used throughout is deeply contingent on the baseline state used, and that this terminology should not be used interchangeably with that from other papers using different baseline conditions. My preference (and suggestion) would be to drop the use of 'suppression' throughout, unless its existence can be more concretely demonstrated.
3. Did behavioral performance vary as a function of the relative location of the mapping stimulus to the reported target stimulus? When the mapping stimulus was nearby, was there any evidence for forward masking?
4. Related to the points above: the authors talk about the uncued stimuli during the focal attention condition as though they should be suppressed - but shouldn't they just be enhanced to a lesser degree than the cued stimulus? On a substantial number of trials, each of these stimuli is relevant for task performance (though see the point about near-chance behavioral performance for those trials), so suppressing their representations doesn't seem useful. Instead, wouldn't the mapping stimulus - which is always task-irrelevant - be suppressed? As mentioned above, are there behavioral consequences of such suppression, which likely can't be disengaged particularly quickly? Zooming out, a better theoretical treatment for how the mapping stimuli should be considered within the task structure would be helpful to the reader. For the study purposes, it's clear that the goal is to use the mapping stimuli to probe visual cortex sensitivity during the cue-target interval, but at a psychophysics level, how should we be thinking of these stimuli?
5. Several additional studies testing the interaction between visual stimulation and attention in fMRI from the deYoe lab are not cited (e.g., Datta & deYoe 2009, Vision Research; Puckett & deYoe, 2015, J Neurosci), despite (I think) being fairly relevant to this result. Additionally, given that this study is somewhat unique in its investigation of distributed attention using pRF mapping methods, I was surprised it didn't address previous fMRI results showing, at least in some cases, similar enhancement of neural responses for each attended stimulus, regardless of how many are attended (e.g., White et al, 2017, JOV; Chen & Seidemann, 2012, Neuron)
6. While I understand that this constitutes a small number of trials, it seems particularly important to show results from the analysis illustrated in Fig. 5 for trials with no mapping stimulus. This would reflect the actual, true, 'raw' attentional modulation in the absence of the distracting visual stimulus, which I believe is the point of this figure.
7. Related to the above, I think the method the authors are using the quantify the 'width' of the spread in Fig. 5C is a bit strange - based on the description in the Methods (lines 974-980), it seems that spread is computed as the width of the function as it crosses 0 - but this seems completely arbitrary, and like it would conflate mean amplitude changes (shifting the function up/down) and changes in the width itself. My sense is the shape of the function - that is, its actual width rather than width at a specific (and variable) y value - is what should be reported. If there's a strong argument for keeping the 'spread' value as computed, it should be made in the manuscript.

Minor comments:

1. D-prime measures for the lower vertical meridian trials seem low - is this due to visual field asymmetries in attention/perception? Is this possibly due to the use of a horizontal reference orientation? Or, instead, is this possibly the result of something like the presence of an eyetracker illuminator/camera nearby that location on the screen?
2. Retinotopic maps seem to be defined using the task data itself. One example dataset is shown (Fig. 3D), but it seems like the eccentricity/polar angle values are only illustrated for voxels included within a subset of ROIs, rather than all voxels which pass a VE threshold, as is typically displayed in figures like this. Additionally, the definition of hV4 seems to differ somewhat from previous studies by this lab (I think), where what's drawn as hV4 looks like it might actually be called hV4, VO1, and VO2 (possibly - it's hard to tell since only included voxels are colored; I'm thinking of the maps and tentative recommendations made in Winawer & Witthoft, 2015, specifically). Regardless, commenting directly about the strategy used to define hV4 would be helpful and provide clarity.
3. For completeness, I think it would be helpful to expand the data presented in Fig. 4 to show (a) both attend-in and attend-out data from focal cue trials, and (b) data from distributed cue trials. The data at present only show the modulation, so it's hard to get a sense for what the 'input' data used to compute the modulation timecourse looks like
4. In the manuscript, the authors refer to the illustration of beta weights sorted according to bar position as a "pseudo-timeseries". As someone quite familiar w/ the implementation of pRF mapping methods, I understand why this term is used - but anyone who isn't deep in the weeds of these methods would be very confused by this terminology, I expect. Given the broad readership of this journal, I recommend using language that's more accessible, like 'beta weight profile' or something like this.
5. For Fig. 7B-C, it would also be helpful to include data from all ROIs, even if the results are primarily significant in the ROIs shown in the current version of the figure
6. Perhaps I just missed it - but what are the intertrial intervals (ITIs)? They seem like they must be relatively short (~<5 s).
7. I don't understand the sentence on lines 858-861 that describes the "triangular function" used to smooth the GLM coefficients. Is this smoothing the bars of beta weights shown in e.g. Fig. 3B/C, 6A? Why is this step necessary? It seems that this was only for 'visualization purposes', but I would prefer the authors include data in the form that was used for model-fitting
8. The paragraph starting on line 954 describes how the 1D polar angle functions are computed and mentions averaging over eccentricities between 4 and 9 deg - why is there a greater distance away from fixation (3 deg) than towards fixation (2 deg)? Cortical magnification? The vertex inclusion threshold defined in lines 982-988 is different (4-8 deg). If there are good reasons for these being different, please include this explanation in the paper. If not, keeping these choices consistent throughout would be easier for the reader to understand.

9. Line 971 describes a parameter “A” in Eq 3, but I don’t see this parameter (perhaps S in the equation?). There’s also a blank space where a parameter (kappa?) should be in that line (“B is an offset, <missing parameter here> is the von Mises center…”)
10. Why does the spatial shift analysis only include voxels foveal of the target stimuli? I know previous studies (e.g., Klein et al 2014) have done this, but they also required participants to attend to stimuli at the extreme edges of the screen. I imagine readers would be interested in seeing these results for voxels with pRF centers at similar eccentricity as the stimuli (e.g., CW/CCW), and those more eccentric (I certainly would be)
11. Given that the pRF mapping approach here is somewhat unconventional (as compared to the smoothly-moving bars more typically used), it would be helpful to readers to include data showing the measured size vs eccentricity relationship in this dataset with this method to verify it aligns with values typically observed using more traditional methods.

(Remarks on code availability)
I've briefly looked through the code. This lab has extensive track record of sharing easy-to-use code. The manuscript directly points to specific files used to run analysis scripts. I have not been able to try running the code myself.

Reviewer #2

(Remarks to the Author)
This paper investigates how attention modulates population receptive fields measured using fMRI. The authors develop a nice behavioral paradigm that reduces some of the artifacts that may have confounded prior studies of prfs, and which allow study of attentional modulation. The psychophysical methods used here are solid, as I would expect from this first-rate group of researchers. Some aspects of the task and model are notable. The use of a random prf mapping method is clearly preferable to the conventional sweeping method. This should minimize the sorts of hysteresis effects that contaminate typical prf measurements. The use of a glm before the prf also probably increases SNR.

Several results are reported. (1) Attention improves performance. (2) The prf model provides an "accurate" fit to the data. (3) The data provide a coarse view of the timecourse of attention effects as measured using BOLD signals. (4) The prfs shift toward the attended location. None of these results are particularly novel or controversial, and most of them have been reported previously in dozens of prior fMRI and neurophysiology studies. That said, the study is well designed and the results fairly clear as far as they are presented, so this is solid work overall.

There are some issues that should be addressed in revision. As the authors make clear in the discussion (but not in the introduction), the issue of receptive field shifts is actually an old question, both in neurophysiology and fmri. Prior studies have shown modulation of firing rates and bold responses from the cue alone in many brain areas, and they have also shown modulation of tuning both before and during stimulation. Some of these studies were cited (e.g., Treue 2008, Silver 2006). In particular, the work of Connor 1997 long precedes the Treue work and seems to be relevant. (In fact the whole paradigm used here is remarkably similar to the Connor neurophysiology study. Given that it seems quite a serious oversight that the connor study was not cited or discussed.) Other relevant studies are Mazer 2004, Motter 1998, Roelfsema 1995 (all neurophysiology), and Hansen 2006 (fMRI). I don't begrudge the authors motivation to maximize the apparent novelty of their results (we all have to do that to some extent) but it would be good to see a more balanced presentation of the current study in the light of prior studies.

The paper talks about attentional suppression of bold response, but offers no explicit information about how this likely affects neuronal responses. This is important here because attentional suppression isn't something one typically sees in neurophysiology studies of attention in visual cortex. What is the assumption about hemodynamic coupling here? How are we to interpret this sort of response? (I could make the same argument about attention related BOLD enhancement of course.) Without some explicit discussion of hemodynamic coupling and/or analysis of the current results relative to what is found in neurophysiology, the value of the data presented in Table 1 is severely reduced.

This problem regarding hemodynamic coupling appears elsewhere in the paper as well. The paper implies throughout that fmri fmri voxels are equivalent to neurons, but they are not. Voxels are collections of many neurons whose activity is filtered through some complicated, nonlinear, possibly nonstationary hemodynmic functions. Therefore, fmri signals cannot be interpreted without making some assumption -- either explicit or implicit -- about hemodynamic coupling.

The accuracy reported here is ~30-40% of variance (and that may be overfit – unclear what steps were taken to avoid overfitting). I am not sure how the authors determine that this is "accurate". The authors claim that this is "relatively high variance" explained, but not relative to what.

The shift magnitude appears higher in v1 than in hv4. This is inconsistent with what we would expect from neurophysiology, so I suspect that it is due to some fMRI-related artifact (is it difficult to measure signals accurately in hv4?). The authors should offer some explanation of this.

It appears that only group-level data are shown here. This is disappointing, because vision studies usually show data from individual participants. Showing only group data without individual participants is a step backwards. In revision the authors should show all the data, at least in the supplementary materials. They might also consider showing cortical surface maps of shift magnitude. Those might give a better sense of the quality of the data than the simple summary figures shown here.

In revision the authors should also clarify some methodological issues, as follows:

Its not clear why invalid cues were required in the focused attention condition. Wouldn't this reduce participant's motivation to focus? Why was the mapping stimulus discretized into one or two seconds? Wouldn't uncertainty about the target onset be maximized if the mapping duration was continuous? And given the calibration procedure, what was the effective eye position accuracy of the eyelink system?

Why was multiband factor 6 chosen? That is a large amount of accelleration. The guidelines that most groups follow now is that one should not exceed MB factor 3 or 4 at most.

As I recall GLMdenoise estimates a single hemodynamic response function for the participant. But it is known that voxels (and so the associated vertices) have widely variable HRFs. Could this have influenced the results? (On the authors side, I'm not sure that this would matter much. This is an attention study and so compares across attend conditions.)

A 2 degree eye window seems enormous… is this diameter or radius? It seems far larger than necessary given the statement that average gaze position was within 0.05 deg of fixation (a number that is probably lower than the accuracy of the eye tracker).

The prf analysis enforces a fairly strong prior on the shape of the prf. Given that the parameters for the prf come from the glm, why couldn't the authors just use the glm itself to get the same information as the prf, but with a less restrictive prior?

Was the prf estimation regularized or not? Given the noise in fmri data regularization would seem to be important.

The attentional latency analysis seems to have been done for each retinotopic map separately. But that assumes that the attentional delay is constant per map. This is a strong assumption. Some analysis should be provided to show that this assumption is correct.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The manuscript describes an interesting extension to the senior authors' recent work. Overall I am enthusiastic about the results, but have several comments (listed below) that the authors might wish to consider.

1. Hemisphere: How did the authors deal with hemisphere? One assumes responses were averaged across hemispheres, but this is not specified clearly.

2. Sample size: N=8. I did not see any mention of a power analysis for this sample size. Is this sample commensurate with prior work looking at similar attention effects?

3. Behaviour: For completeness it would be appropriate to report the post-hoc tests between cue types.

4. Eye-movements: Although rate (~2% of sampled frames) were time-points with fixation breaks included in the analysis?

5. Retinotopic maps:
- What do the individual attention condition estimates look like? This is important because it is not clear if a single attention condition is driving the average responses.
- The dorsal and ventral components of V2/V3 should be labelled. Even if these are averaged for subsequent analyses.
- Report the R2 of the maps used in Figure 3D.

6. Rise and Fall time analysis:
- Line 224 – what was the rationale for operationalising this as 10% of the max response before the peak?
- Line 267 – Again, why 10%? Does this effect hold if you change the threshold (e.g., 25%)?
- Why weren't statistics run on these data?
- The pattern across ROIs between cue-locked and target-locked (notwithstanding the difference in absolute value) appear very similar (i.e., Figure 4D top-panel vs bottom-panel) – could this suggest overall differences between ROIs in terms of SNR?

7. Focal attention affects:
- For completeness, include the main effects of mixed-ANOVA.

- hV4 shows the strongest negative responses. The senior author has previously demonstrated the impact that the dural sinus can have on BOLD estimates in and around the borders of hV4 – One assumes this was looked at in these data? From Figure 3D the polar angle representation of hV4 extends only to the horizontal meridian (green), which would be consistent with distortion from the dural sinus.

8. Focal attention independent of stimulus location: Again, for completeness, can the authors include the significance of these correlation coefficients?

9. pRF shifts:
- Why were statistics not computed for A?

- The basic effect of pRF centres moving towards the attended location is not novel, but it is nicely replicated here. One question that springs to mind is what happens to the pRF size estimate? If constant between attend-in and attend-out then the shift towards the attended location will place more of the stimulus in the pRF of the voxel. An increase in pRF size plus a shift will place even more of the stimulus within the pRF of the voxel, but a decrease and a shift would not necessarily result in any more of the stimulus being in the pRF. Particularly, when the pRF centre shifts are so small (0.4 deg in LO1).

- Figure 7C – why only these ROIs? Why is LO1 now averaged with V3A/B?

- Missing -5 from the x-axis of 7C (bottom-left).

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Overall, I appreciate the authors' careful attention to the comments on the previous manuscript, and think the incorporated changes enhance the paper overall.

There is an extensive response concerning the use of the phrase "suppression" that focuses on what the authors mean by suppression relative to a given baseline (here, distributed attention). If that's the way the authors wish to use that term, that's fine, but it continues to feel very non-intuitive. Most of the time, we're attending to the thing we're looking at (attend-fixation). Sometimes, we need to sample the whole display to decide where to look (distributed attention), and other times, we need to focus attention on a specific location away from fixation. I see the argument as presented in the manuscript that one case is out/out (attend fixation), the other is in/out (attend one stimulus), and the other is in/in (attend all) – but I'll reiterate that it feels a bit off to me. That said, this mostly would just serve to confuse some readers, and doesn't really impact the conclusions themselves (aside from terminology).

The previous comment concerning pRF shifts outside the stimulus eccentricity remains. Yes, the predictions are less straightforward, but that doesn't mean the data shouldn't be shown. The authors overall have done a great job transparently illustrating their results across ROIs and steps of the analysis, with this being an exception. From my perspective, an unclear result from those voxels wouldn't impact the clarity of the result foveal to stimulus locations (as shown now), and would allow readers to relate these findings to other animal and human studies. I'd encourage the authors to reconsider this choice, even

if it would just be for a supplemental figure.

Concerning the impact of the mapping stimulus on behavior – this seems interesting, and worth reporting. There appears to be a substantial change in performance when the mapping stimulus is nearby the reported location, suggesting there is some interaction between its processing and behavioral performance. (the new line, on pg 6, line 141-143, says there is "little effect on performance", but I wouldn't agree with this characterization – the difference in d' appears to be ~0.25-0.75 for the neutral and valid datapoints). I've been thinking about this finding extensively, and overall I don't think it's a problem with the paper (especially because performance seems impacted equivalently across conditions). But it's interesting, and may be worth including in supplemental figures should future studies be interested in elaborating on this finding.

Additional comments:
1. Fig. 8 – I think I misunderstood this figure during the previous submission. Looking over the code on GitHub, which the authors should be commended for sharing so freely, it appears that the simulation was conducted with a single voxel in each ROI. However, the figure is referred to throughout the text as through it aggregates over voxels with different pRFs and shift directions/magnitudes (as in the analyses shown in Fig. 6). I'm a bit confused what this is meant to show – a shift in pRF position should change the set of stimuli that cause the voxel to respond (doesn't seem to be the case here). Additionally, for hV4, why is the response to vertical stimuli greater than to horizontal mapping stimuli? Is this due to pRF coverage extending beyond the aperture? Either the explanation of this figure should be clarified, or the simulations should be brought into alignment with the text.
2. I still am hesitant to support the use of the term "distractor" to refer to the non-cued Gabor locations. These regions are relevant in 1 of 4 trials – so it's not the case that they're ignored, or considered distracting; they're just de-prioritized. I see the authors' new argument that people seem to treat a 75/25% valid/invalid cue as 100/0 (accepting chance-level performance on invalid trials) – but that's speculation on the part of the authors. The fact is those locations were relevant sometimes, just less so. Thus, they're not 'distractors' that can only impair performance when erroneously attended – they're target locations with lower priority.
3. fMRI pulse sequence: in the response to reviewer 2, the authors mention that they're using a multiband sequence with 6x simultaneous multi slice acceleration to scan 64 slices (mentioned in response letter, pg 15) – but I don't believe this would be possible, as I was under the impression that the number of slices was required to be an integer multiple of the multiband acceleration factor. The authors should verify this number, and should add this detail to the manuscript, as the number of slices is only presented in the response letter at present.
4. Significant vertices for ROI definition: on line 1132 of the revised manuscript, the authors mention the vertices visualized for defining ROIs are thresholded according to the GLM variance explained, but there's no mention as to whether pRF variance explained is incorporated in the thresholding. Additionally, in the above section, the authors mention using voxels based on 3% GLM VE, again with no mention of pRF variance explained (nor an explanation as to why different thresholds are used throughout the analyses). This should be justified. (also, I was surprised that there's no mention of the visualization threshold in Fig. 3's retinotopic map in the manuscript – it's reported as > 0.2 in the response to reviewers)
5. The use of Bayesian statistics for the HRF latency analysis in Fig. 4 is a nice addition, though it is unclear why Bayesian statistics are appropriate for this analysis but not others. This figure is also missing a description of error bars for panels B and C – I assume these are CIs, as mentioned in the Methods, but readers would likely benefit from clear description of error bars in each figure caption
6. Fig. 5 stats are missing (the LME described in the text, I believe, just verifies there's no 3-way interaction, doesn't establish other effects of interest), and no definition of error bars (CIs?)
7. New stats lines 461-465 in which shuffled vs intact model are used as a factor in ANOVA seems like a roundabout way to reject the null hypothesis that an intact model explains a different amount of variance than a shuffled one. Also, did this analysis involve only one shuffling iteration?
8. In Fig. 6, the caption mentions "left target meridian-selective vertices", but the cartoon above panel A highlights the right meridian target location. Please check that these are correct

(Remarks on code availability)
Where appropriate, I reviewed the code (see comments). This lab has an excellent track record of sharing code and data to reproduce results.

Reviewer #2

(Remarks to the Author)
I thank the authors for their very thorough and clear reply to reviewers, and to their revisions to the manuscript. The current paper is much improved over the original submission.

Positive aspects of the revised manuscript:

- Appreciate that the authors have revised text to address hemodynamic coupling. Don't agree completely with their points, but they are arguable and no one really knows the ground truth, so I am happy to leave the authors' statements as given.

- I thank the authors for correcting the visual area boundaries. The new boundaries look more plausible.

- I also thank the authors for adding the missing data for individual participants.

Issues that should be addressed in revision:

- Because different regions of the brain have different SNR it isn't really appropriate to use one brain area to try to estimate the noise ceiling or noise floor of another area. So as far as I'm concerned it would be less misleading to simply eliminate supplementary figure 1a.

- I also have to quibble with the authors' claim that enforcing a hard pRF model is equivalent to regularization. Regularization is a method for separating signal from noise, by enforcing a prior noise model. The pRF model does not model the noise, it provides a prior form for modeling the signal. The model of the noise and the model of the signal are two different things. Others have written extensively about this issue in the regression literature...

Other comments about issues that should not impede publication.

- As the authors note the pRF model is quite complicated, and involves lots of prior assumptions. In such circumstances, the best route forward is to compare several versions of a model that differ in the aspects of the model form that are most likely to lead to confounds, artifacts, or incorrect conclusions. If all of these multiple variants lead to the same conclusion, then one can be certain that the prior assumptions did not confound the results. It would be straightforward to test several competing pRF models, and I've always been confused about why this approach isn't common. Given that there isn't really anything in this report that conflicts with prior studies I am not going to request that the authors do that this time. But I would recommend that they do it in the future.

- Many of the authors' responses to reviewer's complaints about experimental parameters (e.g., 1 or 2 s jitter) or model fitting (e.g., using a canonical HRF, aggregating vertices across ROIs) were based on an argument that it was done this way because it was easiest. That isn't a great argument, and it doesn't inspire confidence. There isn't really anything in this report that conflicts with prior studies, and I like this group and have a high opinion of their work in general, so I am not going to litigate all the points. But my advice is that in the future the authors should consider trying to improve their computational methods to avoid these problems.

- I would advice the authors to avoid the use of MB6 in the future. MB6 is too high to provide optimal BOLD SNR, and it is even worse if combined with other forms of acceleration. This is widely known in the field.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have responded to my comments. I am still unconvinced by their novelty argument, but I appreciate the work that has been put into clarifying the other points raised.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the authors' careful response to my concerns. They are completely right that the d' difference I commented on reflects a very very small difference in behavior, and I should have caught that – my mistake. I have no further concerns, and look forward to citing this work in the future.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)
The current study investigates how covert spatial attention cues alter population receptive field properties across visual cortex, measured via fMRI and psychophysics. Tünçok and colleagues mapped pRFs during anticipation, the moment between the cue and target stimulus onset, revealing spatially-tuned push-pull dynamics in pRFs along the visual hierarchy. More specifically, relative to distributed attention, focal attention elicited similar levels of enhancement of the task-relevant location across visual areas, with increased suppression of unattended regions along the visual hierarchy. This is an important study for the on-going quest to elucidate the neural circuits that control attention-related enhancement and suppression of visual representations.

Past reviewers had several concerns, a couple of them on the authors' definition of "suppression" and "distractor". In my opinion, the authors have appropriately addressed these concerns by improving the consistency of word choice and placement; but (and if possible) would the authors be opposed to incorporating Supplementary Figure 8 as a subpanel in one of the main figures, maybe Figure 1 or Figure 5? Having that visualization in the main text could facilitate an understanding in more readers that the attend-all condition serves as the more intuitive baseline for quantifying suppression than an attend-fixation condition, where resources are already withdrawn.

Additional comments include requests to clarify behavioral and pRF results, methodology and statistical analyses. The authors have refined and supplemented all of the above. I was brought on to evaluate whether Reviewer 2's concerns were addressed specifically. Major issues from Reviewer 2 were (1) that the noise floor/ceiling of one brain area shouldn't be used to estimate the noise floor/ceiling in another, so comparing the variance explained between brain areas in Supplementary Figure 1 might be misleading; and (2) enforcing a hard pRF model is not equivalent to regularization.

The authors state that "noise floor" was a mischaracterization on their part, as they agree that each brain area has varying levels of SNR to begin with. If anything, this point strengthens their findings in Supplementary Figure 1 that the visual areas have more similar distributions of variance explained than they do with the pre-central gyrus. Additionally, the authors omit regularization, avoiding the previous false equivalency altogether. The authors acknowledge the remainder of Reviewer 2's recommendations for the future.

Overall, I believe the authors have appropriately addressed all Reviewer concerns and I enjoyed reading their manuscript.

Minor comments:
- Is ref. 26 missing? It does not appear in the References section on my end.
- Figure 8 has a very low resolution on my end as well.
- In Figure 1A, along the time dimension, I believe there is a small Adobe Illustrator artifact: a red square with a plus, which usually means a text box that is too small.
- Potential typo on Lines 661–662; my correction suggestion in brackets: "The task-relevant target (the Gabor) was low contrast, small and briefly presented to minimize its effect on the BOLD [signal]."
- Do all supplementary figures need to be referenced in text? If so, Supplementary Figures 5 and 6 are not.

(Remarks on code availability)

Reviewer #1 (Remarks to the Author):

In this study, Tünçok et al. investigated how covert spatial attention impacts both spatial encoding properties and spatial activation patterns in retinotopic cortex. Inside the scanner, participants performed a visual psychophysics task in which they attended to either one or four peripheral locations to perform an orientation discrimination task at threshold (estimated based on attend-four trials). During the interval between attention cue and target array onset, a bar-shaped 'mapping' stimulus appeared at a random location, allowing for estimates of visual population receptive field (pRF) properties during each attention condition. Behaviorally, they found that sensitivity was highest for trials with a valid cue to one of four possible target locations as compared to trials with an uninformative cue requiring distributed attention across all target locations. Looking at the temporal dynamics of attentional modulation in BOLD signal from spatial attention, attentional enhancements spanned the interval from cue to target, suggesting they were sustained in anticipation of the target array. Furthermore, pretarget BOLD responses showed a baseline shift in visual field maps, where responses were more positive at the attended location compared to unattended locations relative to the distributed attention condition. pRF centers also shifted towards the attended location (mostly).

Overall, this manuscript is a well-written and thorough exploration of how focal and distributed covert spatial attention impact task performance and neural processing of visual stimuli. However, there are some important issues that preclude a positive recommendation at this time. It seems that the primary novelty of this study compared to many previous studies comparing pRF properties across manipulations of covert spatial attention is the use of a distributed condition, which the authors treat throughout as a neutral baseline. This condition is maybe the most interesting part of the paper, but it's hard to fully understand the results of this condition on their own without the much more 'typical' baseline often used in pRF studies (including in most mapping studies by the senior author's lab): the attend-fixation (or, sometimes, attend-mapping-stimulus) condition. This and other important issues are elaborated below:

– We thank the reviewer for their comments and suggestions.

Major comments:

1. Behavioral task: d' is very high for the valid trials of the focused cue condition, and very low (basically at chance for several participants) for the invalid cue condition. (a) does this mean that participants were only attending the cued location, and essentially ignoring the uncued locations? It would appear so, at least for some participants. I don't think this would substantially impact the interpretation of the neural results, but it may be worth considering. (b) is it possible that participants learned that the discrimination task was much easier on focused trials (on average), and so weren't as engaged as they were on distributed cue trials? The psychophysics are elegant and do an overall nice job of showing the impact of focused attention on behavioral discrimination performance, but one consequence of the equated discrimination thresholds between conditions is that the task is easier/harder across conditions of interest, which could plausibly lead to differences in fMRI signals.

(A) Indeed, the goal of our cueing protocol was to elicit focal attention (for the trials with focal cues), meaning to direct attention to the cued location and not to other locations. To the degree that this

manipulation was successful, we see it as a strength of the experiment, not a confound or problem. We used a cue validity of 75%. There is psychophysical evidence that participants perform tasks with ≥75% cue validity as if they only receive 'valid' cues (e.g., Sperling & Melchner, 1978; Kinchla, 1980; Giordano, McElree & Carrasco, 2009). Our fMRI analysis focused on the changes in brain activity evoked by these attentional cues. Participants attending the cued locations while ignoring the uncued locations increased both the behavioral accuracy and the BOLD response amplitude.

(B) We do not have a measure of engagement in our study so we cannot test the possibility that differences in engagement, rather than attention, account for our results. If, however, we make a simple (but non-quantitative) assumption that reduced engagement leads to reduced BOLD responses and poorer performance, then the engagement account would make opposite predictions to the selective attention account: If participants were less engaged in trials with focal cues, this would have resulted in *decreased* performance and brain activity. We report the opposite pattern, in line with the attentional manipulation account rather than decreased engagement. Finally, we note that the question of whether deploying distributed or divided attention is more effortful or taxing than deploying focal attention in some well-defined sense is indeed an interesting question, but beyond the scope of this study.

Action:

(1) We now discuss how the fraction of cue validity affects performance (in *Discussion: Importance of the neutral condition in an attentional experiment,* **Page** 19-20, **Lines** 527-530).
(2) We also discuss distributed/divided attention (in *Discussion: Importance of the neutral condition in an attentional experiment,* **Page** 20, **Lines** 564-570).
(3) We also contrast an effort vs attention account (in *Results: Focal attention affects the BOLD amplitude in a spatially tuned manner,* **Page** 12, **Lines** 349-355*)*. See also our reply to Comment (5).

2. 'Suppression': throughout the manuscript, the authors discuss their results from the focused-cue condition as compared to the distributed-cue condition as evidence for 'suppression' of activation, etc, due to focused attention. Why is this suppression relative to distributed attention, rather than distributed attention is enhancement relative to focused attention elsewhere? I've thought carefully about the Supp Figure (3) and discussion paragraph discussing this issue, and I can't agree with the authors' argument that distributed attention is a sensible baseline condition for fMRI studies. And this gets at a central issue identified in the summary above: it's hard to compare these results to the many previous studies that have often used a more typical attend-fixation condition against which focused attention results have been compared. Perhaps the authors are right and those comparisons are incorrect for some reason - the point stands that the literature at present is based primarily on those comparisons, and so it seems important to at least acknowledge that the terminology used throughout is deeply contingent on the baseline state used, and that this terminology should not be used interchangeably with that from other papers using different baseline conditions. My preference (and suggestion) would be to drop the use of 'suppression' throughout, unless its existence can be more concretely demonstrated.

– There are effectively two issues raised in this question: (1) why distributed attention as a baseline

and (2) what is meant by "suppression". These questions underlie most of the major issues raised by the reviewer. We address them here (and make corresponding updates to the manuscript).

(1) First, **why distributed attention as the baseline?** Attention is the prioritized processing of some kinds of information over others. An important question is whether the additional processing comes for free – i.e., whether one can obtain enhanced performance for some location (in the case of spatial attention) *with no cost* to processing anywhere else, or whether there is a tradeoff. This has been considered a fundamental question in the field (see, for example, the following reviews, Kinchla 1992; Desimone & Duncan, 1995; Carrasco 2011, 2014; Anton-Erxleben 2013; Reeves, 2020). Whether or not the benefits of attention come for free has implications for theories of attention and interpretation of experimental results and has implications for the neural mechanisms underlying attention. Assessing both the costs and benefits of selective attention is a key part of our study, both at the neural and the behavioral level. Doing so requires three conditions: attend-in, attend-out, and a baseline. The attend-in and attend-out are straightforward in our experimental design: For behavior, attend-in are the trials with valid cues and attend-out are the trials with invalid cues. For the fMRI data, attend-in is analysis of vertices whose pRFs in the attended location, and attend-out is analysis of the vertices whose pRFs are outside the attended location. Better performance with valid cues than invalid cues (or larger BOLD responses in the attend-in vs attend-out regions) could be due to enhancement at the attended location, suppression at the unattended location, or both. A baseline is needed to distinguish these. The reviewer suggests that the baseline condition should be attend-fixation. Fixation, however, is itself a location (x=0, y=0), and it is a location that differs from the peripheral test stimulus location. Hence *attend-fixation is attend-out*, and using this condition as the baseline would mean we would have one attend-in and two attend-out conditions, rather than one attend-in, one attend-out, and one baseline. That would leave us unable to test for behavioral or neural tradeoffs and defeats the purpose of the baseline condition. With the attend-fixation baseline, we would be unable to assess whether improved performance and increased BOLD response due to focal attention comes at a cost. Note that prior fMRI work on attentional modulation in visual cortex acknowledged this ambiguity: *"It is important to note that the phase-mapping technique used here to identify retinotopic organization [of attention] does not distinguish cyclic enhancement (increased activation) from cyclic suppression (decreased activation) or from alternating enhancement and suppression"* (Brefczynski and DeYoe, 1999, Nature).

Finally, a clarification about the baseline. The reviewer wrote, "I can't agree with the authors' argument that distributed attention is a sensible baseline condition for fMRI studies." We do not argue that, in general, one should use this baseline for fMRI studies. This baseline is optimal for our research questions and experimental design. We now clarify this in the text **(Page** 20, **Lines** 558-559).

(2) Second, **what do we mean by "suppression"**? We use the word "suppression" in the typical neuroscience way, as a *phenomenological* rather than *mechanistic* descriptor. Specifically, we use suppression to refer to a decrease in neural activity or performance relative to a baseline. We use "enhancement" in the parallel way: an increase in neural response or performance relative to a baseline. As the reviewer correctly points out, one could switch the reference point and describe baseline as suppression relative to attend-in (or as enhancement relative to attend-out). Neither reference point is more correct than the other. But for consistency throughout the paper, we refer to performance and neural responses in the attend-in and attend-out conditions relative to a baseline,

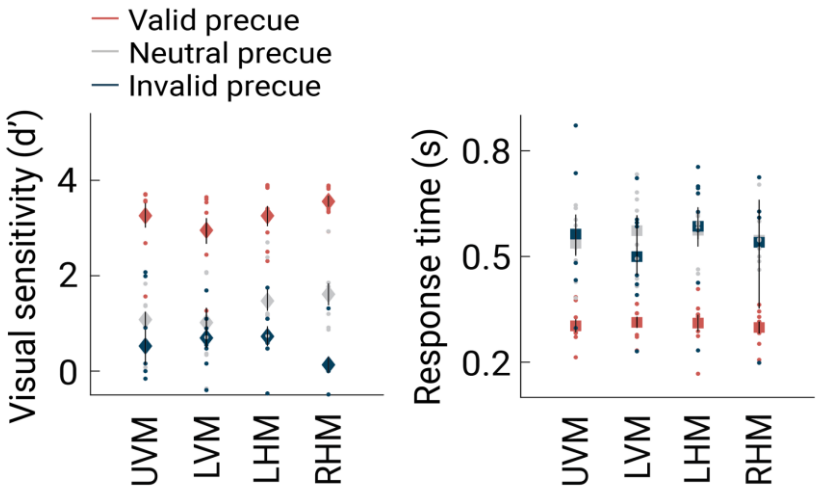
rather than describing the baseline relative to the two focal conditions. As explained above, we chose a distributed attention condition as a baseline for concrete reasons. By this definition, BOLD responses that are lower than this baseline are suppressed, and BOLD responses that are higher are enhanced. We show multiple results that support the claim of suppression according to this definition (**Figures 5** and **6** and a new **Supplementary Figure 2**). The reviewer suggests dropping the word suppression, "unless its existence can be more concretely demonstrated". This implies that the reviewer defines suppression differently than we do, presumably as a kind of neurobiological mechanism such as hyperpolarization due to inhibitory neurotransmitters. For the meaning implied by the reviewer, the more typical word would be "inhibition", which we do not use. As but one of many examples, consider this paper, *Suppression and facilitation of human neural responses* (Schalmo et al, 2018, eLife), which argues that the phenomenon of suppression need not arise from the neural process of inhibition. The question of the neurobiological cause of suppression – such as shunting inhibition versus a reduction in recurrent amplification – is interesting but beyond the scope of this paper.

Actions:

(1) We now explain the logic of the attend-distributed baseline more explicitly **(Page** 20, **Lines** 544-552).
(2) We have clarified the terms *enhancement* and *suppression* (in Results: *Focal attention affects the BOLD amplitude in a spatially tuned manner,* **Page** 11**, Lines** 316-323) and Discussion **(Page** 24, **Lines** 667-672).
(3) We have added BOLD time series plots for the three conditions, attend-in, attend-out, and attend-distributed, showing that the latter is intermediate between the other two **(Supplementary Figure 2)**.

3. Did behavioral performance vary as a function of the relative location of the mapping stimulus to the reported target stimulus? When the mapping stimulus was nearby, was there any evidence for forward masking?

– We chose ISIs to be long enough to avoid forward masking (≥300 ms on 90% of trials). Nonetheless, to check, we separated trials according to mapping stimulus position. We then analyzed the behavioral sensitivity and reaction time at each target location for trials where any part of the mapping stimulus was near the reported target (within the 4.5° to 7.5° eccentricity band around the stimulus location of 6°, resulting in 6 bar positions per polar angle location out of 48 total locations). Even though this leaves only 10% of the total trials, sensitivity remains very high for the valid condition, much lower for neutral, and lower still for invalid. If there is any masking effect, it is small and constant across the valid, invalid and neutral conditions.



Action: We now note that visual inspection of the data showed that the bar stimulus location had little effect on performance (**Page** 6**, Lines:** 141-143). We did not include the new plot in the revised manuscript, however, as we do not see any way in which tiny performance differences as a function of mapping stimulus location confounds any of our results or conclusions.

4. (A) Related to the points above: the authors talk about the uncued stimuli during the focal attention condition as though they should be suppressed - but shouldn't they just be enhanced to a lesser degree than the cued stimulus? On a substantial number of trials, each of these stimuli is relevant for task performance (though see the point about near-chance behavioral performance for those trials), so suppressing their representations doesn't seem useful.

(B) Instead, wouldn't the mapping stimulus - which is always task-irrelevant - be suppressed? As mentioned above, are there behavioral consequences of such suppression, which likely can't be disengaged particularly quickly? Zooming out, a better theoretical treatment for how the mapping stimuli should be considered within the task structure would be helpful to the reader. For the study purposes, it's clear that the goal is to use the mapping stimuli to probe visual cortex sensitivity during the cue-target interval, but at a psychophysics level, how should we be thinking of these stimuli?

(A) The goal of focal cueing is to elicit attention to the cued location. Whether or not suppression is found, as in reduced BOLD amplitude or poorer performance relative to neutral, is an empirical question, not one that can be answered from first principles. An ideal observer model (with resource limits), might or might not show suppression at the uncued locations, depending on the relative size of increases in performance due to attending toward vs decreases in performance due to attending away, and the frequency of each trial type. But our goal was not to compare performance to an ideal observer model, but rather to understand how focal attention affects performance. For this goal, it is necessary to get subjects to deploy focal attention. If they had not done so, this would have been a failure of our experimental design, rather than a finding. (Note that our findings might be useful in the future to someone who wants to implement an optimal performance model in with realistic tradeoffs due to limited resources). As we replied in (1.A) and (1.B), participants relied heavily on the attentional cues to perform well on our task, which shows the effectiveness of our attentional manipulations. It is

correct that across the experiment each of these locations bears importance to perform the task well, however, this relevance changes on a trial-by-trial basis. The participants therefore know that a central focal cue is highly valid and helps them perform better in an otherwise difficult task.

(B) As the reviewer notes, the function of the mapping stimulus is to map the visual cortical responses after the attentional cue is shown, with no relation to the task performed. In line with this, we explicitly informed our participants that the mapping stimulus is *task-irrelevant*. As shown in the previous response, the bar stimulus had little effect on behavioral performance. The behavioral results were reported here to validate our experimental manipulation to better interpret the fMRI results. The behavioral results were not reported to test a new idea about spatial attention and performance. It is clear from the behavioral results, that our goal of eliciting focal attention from focal cues was successful, irrespective of bar position, and that any small effects of bar position were swamped out by the large cueing effects. Hence exploring these small effects in greater depth would go beyond the scope of the paper.

Action: Please refer to our reply to Comment (1) about why we expected the focal cues to elicit focal attention, and to our reply to Comment (3), which shows little effect of mapping stimulus location on attentional trade-offs.

5. Several additional studies testing the interaction between visual stimulation and attention in fMRI from the deYoe lab are not cited (e.g., Datta & deYoe 2009, Vision Research; Puckett & deYoe, 2015, J Neurosci), despite (I think) being fairly relevant to this result. Additionally, given that this study is somewhat unique in its investigation of distributed attention using pRF mapping methods, I was surprised it didn't address previous fMRI results showing, at least in some cases, similar enhancement of neural responses for each attended stimulus, regardless of how many are attended (e.g., White et al, 2017, JOV; Chen & Seidemann, 2012, Neuron)

Action:

(1) We added a section in our discussion on distributed/divided attention (in *Discussion: Importance of the neutral condition in an attentional experiment,* **Page** 20, **Lines** 564-570), citing the White et al and Chen and Seidemann papers.
(2) We also cited studies from DeYoe lab in our introduction (**Page** 3**, Lines** 65).

6. While I understand that this constitutes a small number of trials, it seems particularly important to show results from the analysis illustrated in Fig. 5 for trials with no mapping stimulus. This would reflect the actual, true, 'raw' attentional modulation in the absence of the distracting visual stimulus, which I believe is the point of this figure.

Action: We ran the same analysis in Figure 5 only using the beta estimates from null trials (no mapping stimulus shown) and added **Supplementary Figure 4**. We note that the pattern of findings was remarkably similar whether the analysis was conducted on all trials or only the blank trials.

7. Related to the above, I think the method the authors are using the quantify the 'width' of the spread in Fig. 5C is a bit strange - based on the description in the Methods (lines 974-980), it seems that

spread is computed as the width of the function as it crosses 0 - but this seems completely arbitrary, and like it would conflate mean amplitude changes (shifting the function up/down) and changes in the width itself. My sense is the shape of the function - that is, its actual width rather than width at a specific (and variable) y value - is what should be reported. If there's a strong argument for keeping the 'spread' value as computed, it should be made in the manuscript.

– The data from which we derive tuning widths are differences between conditions, not response amplitudes for a single condition. The *y*-values are not arbitrary. A *y*-value of zero means focal attention responses are equal to distributed attention responses. Hence the width at y=0 defines the spatial extent of enhancement. Shifting the function up or down, as the reviewer notes, would indeed change our calculation of the width. That is a feature, not a bug, of our metric. If, for example, the entire curve was shifted below 0, that would mean that there was no part of cortex where the focal condition caused a larger BOLD response than distributed, and hence the width should be 0.

Now consider an alternative. We might find the halfway point between the top of the curve and bottom of the curve and report that width, as the reviewer appears to suggest. However, in our view, the halfway point between the top and bottom of the curve is arbitrary: it could indicate suppression (if it is below the x-axis), enhancement (if it is above the x-axis), or neither (if it is on the x-axis).

Action: We had previously reported our logic in the methods **(Page** 35, **Lines** 1118-1121 in revised text). We have now added text to the Results as well to explain why we use the x-intercept method in the spatial extent estimation (in *Results: The pattern of enhancement and suppression varies across visual cortex,* **Page** 14, **Lines** 375-381).

Minor comments:

1. D-prime measures for the lower vertical meridian trials seem low - is this due to visual field asymmetries in attention/perception? Is this possibly due to the use of a horizontal reference orientation? Or, instead, is this possibly the result of something like the presence of an eye tracker illuminator/camera nearby that location on the screen?

– The overall accuracy was about the same across the 4 locations. Breaking down the locations by cue type, the lower vertical meridian differs from the other locations in the neutral cue condition (distributed attention). We don't have an explanation other than the measurement noise, which is unavoidable, especially with limited numbers of trials. Because an effect of horizontal orientation would also affect performance at the upper vertical meridian location (tangential orientation at both locations), we do not interpret the lower vertical meridian pattern due to the stimulus orientation. Given that d' in this location is close to 4 for the focally cued condition, it seems unlikely that a physical obstruction imposed a meaningful visibility problem. We attempted to equate performance across location by a pre-scan calibration, but the estimates from that calibration have noise, as do the measurements in the scanner.
Action: Given that differences between locations were small and inconsistent across cue conditions, we prefer not to speculate about what the sources of the variability are.

2. Retinotopic maps seem to be defined using the task data itself. One example dataset is shown (Fig. 3D), but it seems like the eccentricity/polar angle values are only illustrated for voxels included within a subset of ROIs, rather than all voxels which pass a VE threshold, as is typically displayed in figures like this. Additionally, the definition of hV4 seems to differ somewhat from previous studies by this lab (I think), where what's drawn as hV4 looks like it might actually be called hV4, VO1, and VO2 (possibly - it's hard to tell since only included voxels are colored; I'm thinking of the maps and tentative recommendations made in Winawer & Witthoft, 2015, specifically). Regardless, commenting directly about the strategy used to define hV4 would be helpful and provide clarity.

Action: We made corrections to manually drawn ROI borders after the reviewer pointed out the inconsistency. Now in all participants the borders of hV4 are delineated based on Winawer & Witthoft, (2015). ROIs for a few subjects had been drawn by a different lab member for a different project on the same participants. For consistency, all the ROIs have now been drawn by the first author and checked by all the coauthors. We repeated our analysis with the new ROI borders and now report those results. This affects all figures with fMRI data, but the findings are the same.

We also updated **Figure 3** to show the eccentricity and polar angle estimates of all voxels that pass a threshold of variance explained by the model (> 0.2). We also show the average variance explained across vertices for these maps.

Action: Some ROIs were redrawn with corresponding updates to all relevant figures. Variance explained summary was added to **Figure 3**.

3. For completeness, I think it would be helpful to expand the data presented in Fig. 4 to show (a) both attend-in and attend-out data from focal cue trials, and (b) data from distributed cue trials. The data at present only show the modulation, so it's hard to get a sense for what the 'input' data used to compute the modulation timecourse looks like

Action: Done. We have now updated **Figure 4B** to show the separate attend-in, distributed and attend-out time courses from V3. We now also show the time courses separately for each visual field map in **Supplementary Figure 2**.

4. In the manuscript, the authors refer to the illustration of beta weights sorted according to bar position as a "pseudo-timeseries". As someone quite familiar w/ the implementation of pRF mapping methods, I understand why this term is used - but anyone who isn't deep in the weeds of these methods would be very confused by this terminology, I expect. Given the broad readership of this journal, I recommend using language that's more accessible, like 'beta weight profile' or something like this.

Action: Done. We changed the term "pseudo time series" to "beta weight profile" throughout the paper.

5. For Fig. 7B-C, it would also be helpful to include data from all ROIs, even if the results are primarily significant in the ROIs shown in the current version of the figure

Action: Done. We now updated **Figure 7B** and **7C** to include all ROIs reported in **Figure 7A**.

6. Perhaps I just missed it - but what are the intertrial intervals (ITIs)? They seem like they must be relatively short (~<5 s).

– The ITI was the response period, either 1350 or 1700 ms, with no additional delay between trials. The variability in the response period was opposite to the variability in the ISI following the mapping stimulus presentation, to ensure that the total trial length was an integer number of seconds (and hence, integer number of TRs).

Action: We now have made this information explicit in Methods (**Page** 27, **Lines** 817-821).

7. I don't understand the sentence on lines 858-861 that describes the "triangular function" used to smooth the GLM coefficients. Is this smoothing the bars of beta weights shown in e.g. Fig. 3B/C, 6A? Why is this step necessary? It seems that this was only for 'visualization purposes', but I would prefer the authors include data in the form that was used for model-fitting

– Yes, the triangular function was used to smooth the data for better visualization of the GLM output, and as was expressed in the manuscript, we used the beta weight profile as it is to estimate the pRF of each vertex.

Action: We now added the unsmoothed beta weight profile in the form that it was used for model fitting in **Supplementary Figure 6**.

8. The paragraph starting on line 954 describes how the 1D polar angle functions are computed and mentions averaging over eccentricities between 4 and 9 deg - why is there a greater distance away from fixation (3 deg) than towards fixation (2 deg)? Cortical magnification? The vertex inclusion threshold defined in lines 982-988 is different (4-8 deg). If there are good reasons for these being different, please include this explanation in the paper. If not, keeping these choices consistent throughout would be easier for the reader to understand.

– The reported 4º to 9º range was a typo. An eccentricity range of 4°- 8° was used for all the amplitude analyses pertaining to target-selective cortical ROIs.

Action: We have corrected the typo for the eccentricity range of 1D polar angle functions from 4-9° to 4-8° **(Page** 34, **Lines** 1099).

9. Line 971 describes a parameter "A" in Eq 3, but I don't see this parameter (perhaps S in the equation?). There's also a blank space where a parameter (kappa?) should be in that line ("B is an offset, <missing parameter here> is the von Mises center…")

Action: We now fixed the typo in text on parameter *S,* as it should be. The missing parameter was due to an error perhaps in Word to PDF conversion and referred to μ parameter for the Von Mises center.

10. Why does the spatial shift analysis only include voxels foveal of the target stimuli? I know previous studies (e.g., Klein et al 2014) have done this, but they also required participants to attend to stimuli at the extreme edges of the screen. I imagine readers would be interested in seeing these results for voxels with pRF centers at similar eccentricity as the stimuli (e.g., CW/CCW), and those more eccentric (I certainly would be)

– We chose the locations in the visual field constrained by the target stimuli because the predictions are clearest for pRFs in these locations. For any pRF whose eccentricity is <6º, the comparison between attend-left and attend-right predicts shifts in the opposite direction with respect to the x-axis, and for attend-up vs attend-down along the y-axis. For pRFs greater than the target eccentricity, the difference in predicted shifts can be ambiguous. For example, for a vertex with a pRF at [10, 0], attend-left and attend-right might both cause a rightward shift.

Anticipatory shifts of receptive fields have been investigated more in the presaccadic remapping than the spatial attention literature. Two separate accounts have been proposed as to how the presaccadic remapping occurs, the target selection account and the retinal image displacement account (Zirnsak & Moore, 2014). These two accounts would have opposite predictions on the behavior of vertices outside of the target range, and more similar (but not identical) predictions for vertices with pRF inside the target range. Resolving the difference in predictions of these two accounts, however, goes beyond the scope of this paper. Limiting our analyses to vertices within the 6º target eccentricity is hence less dependent on this debate.

Action: We now explain our rationale for the eccentricity restriction (in *Methods*: *Directional vector graphs of spatial shifts***; Page** 36, **Lines** 1162-1168).

11. Given that the pRF mapping approach here is somewhat unconventional (as compared to the smoothly-moving bars more typically used), it would be helpful to readers to include data showing the measured size vs eccentricity relationship in this dataset with this method to verify it aligns with values typically observed using more traditional methods.

Event-related designs for pRF experiments are indeed less common than designs with smoothy moving stimuli. Nonetheless, they have been used in several impactful studies, each showing the usual size vs eccentricity relationship (e.g., Kay, Winawer et al, 2013; Kay, Weiner, Grill-Spector, 2015; Poltoratski et al, 2021).

Action: We now added the pRF size-eccentricity estimates as **Supplementary Figure 1B**. We also cite the aforementioned papers using event-related designs for pRF mapping **(Page** 32, **Lines** 1012-1014).

Reviewer #1 (Remarks on code availability):
I've briefly looked through the code. This lab has extensive track record of sharing easy-to-use code. The manuscript directly points to specific files used to run analysis scripts. I have not been able to try running the code myself.

—

Reviewer #2 (Remarks to the Author):

This paper investigates how attention modulates population receptive fields measured using fMRI. The authors develop a nice behavioral paradigm that reduces some of the artifacts that may have confounded prior studies of prfs, and which allow study of attentional modulation. The psychophysical methods used here are solid, as I would expect from this first-rate group of researchers. Some aspects of the task and model are notable. The use of a random prf mapping method is clearly preferable to the conventional sweeping method. This should minimize the sorts of hysteresis effects that contaminate typical prf measurements. The use of a glm before the prf also probably increases SNR.

Several results are reported. (1) Attention improves performance. (2) The prf model provides an "accurate" fit to the data. (3) The data provide a coarse view of the timecourse of attention effects as measured using BOLD signals. (4) The prfs shift toward the attended location. None of these results are particularly novel or controversial, and most of them have been reported previously in dozens of prior fMRI and neurophysiology studies. That said, the study is well designed and the results fairly clear as far as they are presented, so this is solid work overall.

– We thank the reviewer for their comments and suggestions.

There are some issues that should be addressed in revision.

1. As the authors make clear in the discussion (but not in the introduction), the issue of receptive field shifts is actually an old question, both in neurophysiology and fmri. Prior studies have shown modulation of firing rates and bold responses from the cue alone in many brain areas, and they have also shown modulation of tuning both before and during stimulation. Some of these studies were cited (e.g., Treue 2008, Silver 2006). In particular, the work of Connor 1997 long precedes the Treue work and seems to be relevant. (In fact the whole paradigm used here is remarkably similar to the Connor neurophysiology study. Given that it seems quite a serious oversight that the connor study was not cited or discussed.) Other relevant studies are Mazer 2004, Motter 1998, Roelfsema 1995 (all neurophysiology), and Hansen 2006 (fMRI).

Action: We now cite Connor et al, 1997 and Hansen et al, 2006 in our introduction (**Page** 3, **Line** 66).

The other mentioned studies are all interesting, but please note that our study investigates covert spatial attention. Both Mazer & Gallant, 2004 and Motter & Belky, 1998 investigate saccadic planning. Covert attention and presaccadic attention (related to saccadic planning) can be dissociated (Li, Hanning & Carrasco, 2021). Roelfsema's work from 1995 focuses on neural synchronization in cat visual cortex, and a first-author 1998 paper investigates object-based attention. Neural mechanisms and behavioral signatures of both object-based attention and saccadic planning differ from those of spatial attention (e.g., Soto & Blanco, 2004; Mazer, 2011; Li Pan & Carrasco, 2021).

2.The paper talks about attentional suppression of bold response, but offers no explicit information about how this likely affects neuronal responses. This is important here because attentional

suppression isn't something one typically sees in neurophysiology studies of attention in visual cortex. What is the assumption about hemodynamic coupling here? How are we to interpret this sort of response? (I could make the same argument about attention related BOLD enhancement of course.) Without some explicit discussion of hemodynamic coupling and/or analysis of the current results relative to what is found in neurophysiology, the value of the data presented in Table 1 is severely reduced.

– Throughout the paper, we refer to *an increase from the baseline measurement* (distributed attention) as *enhancement*, and a *decrease* from the baseline measurement as *suppression.* The term suppression here does not denote a negative BOLD signal, nor assigns an *inhibitory* mechanism to these response profiles. Rather, it denotes a decrease in response relative to the distributed condition. The response relative to *no stimulus and no task* can be seen in **Figure 6A**. With the exception of the V1 responses to a few stimuli in one condition, all of the data points in all visual areas in all cue conditions are greater than or equal to 0. So, while it is certainly possible that hemodynamic coupling differs across cue conditions, we do not see any reason to think that this is likely. There are several studies in animal models and humans suggesting that hemodynamic coupling to neural activity can change as a function of global state changes, such as expectation or arousal (Sirotin and Das, 2009; Das, Murphy and Drew, 2020); however, this is manifested as a global BOLD response (Burlingham et al., 2022). Although attend in and attend out are considered different conditions for the purposes of analysis, with the latter not the former, showing suppression, they are in fact the same task - i.e., attend focally, distinguished only by the relation between the locus of attention and the pRFs of the voxels being analyzed. Focal attention results in increased BOLD signal in some locations (the locus of attention) and decreased signal in other locations, inconsistent with the kind of state-dependent global signal changes that have been associated with atypical neurovascular coupling.

Action:

(1) We have clarified what we mean by the terms *enhancement* and *suppression* (in Results: *Focal attention affects the BOLD amplitude in a spatially tuned manner,* **Page** 11**, Lines** 316-323) and Discussion **(Page** 24, **Lines** 667-672; See also response to R1, Comment 2)

(2) We also now discuss neurovascular coupling and global state changes (in Discussion: *A multiplicity of effects of attention on BOLD amplitude,* **Page** 21-22**, Lines** 607-622).

3.This problem regarding hemodynamic coupling appears elsewhere in the paper as well. The paper implies throughout that fmri fmri voxels are equivalent to neurons, but they are not. Voxels are collections of many neurons whose activity is filtered through some complicated, nonlinear, possibly nonstationary hemodynmic functions. Therefore, fmri signals cannot be interpreted without making some assumption -- either explicit or implicit -- about hemodynamic coupling.

–We agree that making inferences about neural responses from fMRI measurements requires assumptions about hemodynamic coupling. Perhaps the most pertinent question is whether attentional modulations of the BOLD signal are accompanied by the same kind of neural activity as stimulus-driven modulations. We have no way of answering this directly from our study, however there is some evidence from the literature that can help with interpretation. First, as noted in the response to the previous question, state-dependent shifts in the BOLD signal that are decoupled from neural activity

tend to be spatially global, whereas our attention-related effects are not. Moreover, there is single unit data from macaque in both V1 and V4 showing baseline shifts in neural activity due to attention, consistent with our fMRI data (Williford & Maunsell, 2006 in V4; Thiele et al., 2009 in V1). Finally, our observations are almost entirely positive BOLD signals relative to a fixation baseline, even though they are sometimes negative relative to the distributed attention condition. Altogether, this suggests that there is no special reason to assume that the neurovascular coupling is very different across conditions in our study.

Action: We now discuss neurovascular coupling and global signals (in Discussion: *A multiplicity of effects of attention on BOLD amplitude,* **Page** 21-22**, Lines** 607-622).

4.The accuracy reported here is ~30-40% of variance (and that may be overfit – unclear what steps were taken to avoid overfitting). I am not sure how the authors determine that this is "accurate". The authors claim that this is "relatively high variance" explained, but not relative to what.

The main steps taken to avoid overfitting were to (1) use a cross-validated procedure to select the number of nuisance regressors in the GLM fitting, and (2) employ a low dimensional pRF model, fitting only the x, y, and sigma of the receptive fields.

Action: We now report the variance explained of the averaged pRF model in a sensory brain area that is not visual (the precentral gyrus) to estimate the noise floor and to compare to the fits in visual cortex in **Supplementary Figure 1A.**

5.The shift magnitude appears higher in v1 than in hv4. This is inconsistent with what we would expect from neurophysiology, so I suspect that it is due to some fMRI-related artifact (is it difficult to measure signals accurately in hv4?). The authors should offer some explanation of this.

– Note that there are two kinds of shifts, a shift in the response amplitude and a shift in position. The *position* shifts are larger for hV4 than V1 **(Figure 7A)**. We now refer to other literature that also showed larger position shifts in higher visual areas. The absolute modulation in amplitude is also larger in hV4 than in V1 **(Figure 5C, top, black data)**. This is consistent with electrophysiology findings which typically compare attend-in and attend-out. However, in the initial manuscript, the enhancement effect (increase in BOLD amplitude from attend-in compared to distributed attention) was indeed larger in V1 than in hV4. As noted in the responses to R1, we redrew some of the retinotopic map boundaries. And while this did not have large effects on any major pattern of data, it did result in larger estimated BOLD enhancement in hV4, which is no longer lower than that of V1.

Actions: We removed the comments about hV4 having a low target enhancement signal. In addition, we now report the BOLD time courses in all visual field maps, including hV4 in **Supplementary Figure 2,** which shows that the signal modulation in hV4 has similar profile to other visual field maps, and is consistently above zero, in contrast to what is expected in the case of transverse sinus artifacts (Winawer et al., *JoV*, 2009).

6.It appears that only group-level data are shown here. This is disappointing, because vision studies usually show data from individual participants. Showing only group data without individual participants is a step backwards. In revision the authors should show all the data, at least in the supplementary materials. They might also consider showing cortical surface maps of shift magnitude. Those might give a better sense of the quality of the data than the simple summary figures shown here.

We had previously reported individual data in **Figures 2**, **3**, and **7A**. Analyses in **Figure 4** and **Figure 5** involve a bootstrapping approach to model fitting, where the model is fit to the group level data, computed across 1000 iterations, with each iteration randomly selecting amongst the participants, with replacement, to ensure that we get a wide representation of the empirical data across participants.

Action: In line with the reviewer's suggestion, we now added individual data for Figure 5 as well, shown in **Supplementary Figure 5.**

In revision the authors should also clarify some methodological issues, as follows:

7a. Its not clear why invalid cues were required in the focused attention condition. Wouldn't this reduce participant's motivation to focus?

– Using invalid cues enabled us to assess both benefits at the attended location and costs at the unattended locations in the behavioral task, a hallmark of selective attention (e.g., review by Carrasco, *Vis Res,* 2011), which provides a useful guide for examining the corresponding changes in BOLD signal. It is correct that the invalid cues do not affect the BOLD analysis, as participants cannot know that the cue is invalid when the bar stimulus is viewed or when the Gabor stimuli are displayed, only once the response cue is presented.

A 25% invalid rate is not problematic for eliciting focal attention. Previous research has shown that a cue validity of ≥75% results in observers performing the task as if the cue was always valid (e.g., Sperling 1980; Kinchla, 1980; Giordano, McElree & Carrasco, 2009). This strategy is consistent with the large behavioral attention effect we observed.

Action: We now further clarify the link between invalid cues for behavior and attend-out pRF analysis for BOLD (Section: *The effect of focal attention on the BOLD amplitude is spatially tuned,* **Page:** 11**, Lines** 305-310).

7b. Why was the mapping stimulus discretized into one or two seconds? Wouldn't uncertainty about the target onset be maximized if the mapping duration was continuous?

– The 1-s and 2-s discretization enabled us to sync the stimulus to the MR acquisition (1-s TR) and to have enough trials for each of the durations. Had the ISI been varied continuously, the GLM analysis would have been more complicated, and we would not be able to easily check basic measures like test-retest reliability across trial types. Cross-validation was part of the GLM fitting routine and would have been complicated by variable durations. The jitter we did use, 1-s and 2-s bar stimuli, as well as variable delay between the bar stimulus and the target stimulus, together induced temporal

uncertainty. We consider that this was enough uncertainty, as our fMRI analyses indicate that attention was sustained throughout the period before the target stimulus appeared, both for 1-s and 2-s bar presentation trials (**Figure 4**). Observers were not told that there were only two possibilities, 1 or 2 s.

Action: We now expanded in the methods why we chose discrete durations to present mapping stimulus bars (**Page:** 27, **Lines** 801-805). Moreover, in the section on "*Attentional modulation of the BOLD amplitude is time-locked to the cue onset and sustained throughout the trial"*, we now connect the finding to the temporal uncertainty in the experimental design (**Page:** 11, **Lines** 301-303).

And given the calibration procedure, what was the effective eye position accuracy of the eyelink system?

Action: We have now reported the accuracy of EyeLink calibration in the Methods (**Page:** 30, **Lines** 933-935).

8.Why was multiband factor 6 chosen? That is a large amount of acceleration. The guidelines that most groups follow now is that one should not exceed MB factor 3 or 4 at most.

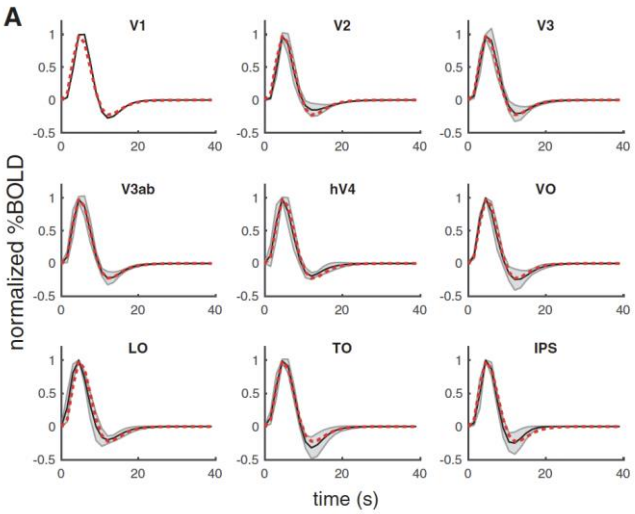
– We piloted the scan parameters when we obtained the new hardware (new Prisma scanner and 64-channel head coil). Many factors affect the ability to separate out the acquired signal into multiple slices using MB acquisitions and algorithms, including coil geometry, head geometry, orientation of the slices, spatial separation between the multiple slices acquired in parallel, and so on. Our pre-testing with this pulse sequence and slice prescription yielded high SNR without obvious multiband artifacts. The reason for 6 specifically is that a multiband factor of 6 allows us to acquire whole brain coverage in a short amount of time with high temporal resolution (64 slices with 2-mm voxels in 1 TR (1-s)).

Action: We now explain our rationale for our MUX factor (Section *MRI Acquisition,* **Page:** 29, **Lines** 875-877).

9.As I recall GLMdenoise estimates a single hemodynamic response function for the participant. But it is known that voxels (and so the associated vertices) have widely variable HRFs. Could this have influenced the results? (On the authors side, I'm not sure that this would matter much. This is an attention study and so compares across attend conditions.)

– Correct. GLMdenoise estimates a single HRF per participant across voxels. As the reviewer notes, HRF differences would not explain differences across attention conditions, as the measurement is repeated per vertex across conditions. Fitting different HRFs per vertex can also lead to the complementary problem: if differences are found between areas, are those differences due to fitting different (possibly incorrect) HRFs per vertex. Moreover, optimizing HRFs per vertex would have been a large computational problem, given that there are 4 sessions of fMRI data per participant, and variable duration events within each trial. In prior work (Zhou, Benson, Kay & Winawer, 2018), we compared results across areas using GLMdenoise, run either once per subject (hence one HRF per participant) or once per ROI (one HRF for each ROI in each participant). The results were similar, indicating that when averaged across a whole ROI, and when using reasonably large voxels (2.5 mm

in that paper), the HRFs are not that variable. As voxels get smaller, they become much more differentially affected by the venous content, and voxel-specific HRFs become more important.



From figure 9 in Zhou, Benson, Kay, and Winawer, 2018 (*J Neurosci)*. The red curves are the same in each plot (subject-specific ROIs, averaged across participants) and the black lines are ROI-specific HRFs.

Action: We clarified the HRF fitting procedure of GLMdenoise (Section *fMRI Data Analysis: General Linear Model,* **Page:** 31, **Lines** 947-948).

10. A 2 degree eye window seems enormous… is this diameter or radius? It seems far larger than necessary given the statement that average gaze position was within 0.05 deg of fixation (a number that is probably lower than the accuracy of the eye tracker).

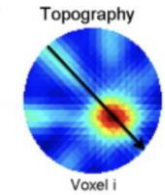
– The eye window of 2° was in radius, and so yes, it is large. But this tolerance was in effect only for the psychophysics testing outside the scanner, while people were trained to do the task and their thresholds were measured. The data we plot in Figure 2C comes from the eye tracker in the scanner and there is no thresholding to remove data.

As for the 0.05° measured gaze position: For analysis of gaze position during the scans, we assumed gaze to be at [0,0] during the fixation period at the start of each trial (before the target location was known), a form of on-the-fly recalibration for each trial. We then measure gaze position at each time point relative to this location. Since many points are measured within each trial, and there are many trials, the average gaze position can be estimated precisely (for example, with a confidence interval that is smaller than the standard deviation of the tracker during calibration). Moreover, the calibration accuracy of about 0.5 to 0.7 deg is estimated from a few large saccades (~10°), whereas the 0.05 estimate is from positions near the center of the screen, where calibration is more precise. Hence for all these reasons, there is no expectation that the fixation accuracy, averaged across time points and trials, and re-calibrated for each trial, should be larger than the reported accuracy of the calibration.

Action: We have now reported the calibration accuracy of the scanner eye tracker (Section *Gaze Position Analysis,* **Page:** 30, **Lines** 933-935).

11.The prf analysis enforces a fairly strong prior on the shape of the prf. Given that the parameters for the prf come from the glm, why couldn't the authors just use the glm itself to get the same information as the prf, but with a less restrictive prior?

– Our GLM estimates 49 beta weights per surface vertex (one for each of 48 bar locations, plus a blank). The linear pRF model is a set of pixel weights, i.e. a 2D image. At the pixel granularity of our experimental display, this would correspond to about a million parameters for a pRF model. Even in the downsampled space we typically use for model fitting pRFs (100 x 100 pixels), there are hundreds of times more parameters than data points. There is no way to get independent estimates for each of the 10,000 parameters from 49 data points alone, and thus no escape from making assumptions about the parameters (i.e., specifying a prior). The assumptions might be explicit or implicit, but they cannot be avoided. Regularized regression (for example, ridge regression) is an alternative to assuming a specific class of shapes. But the prior must show up somewhere. With ridge regression, the prior manifests in pRF solutions that are effectively weighted sums of the stimuli. Hence the pRFs tend to look like this (from Lee, Papanikalou et al, 2009):



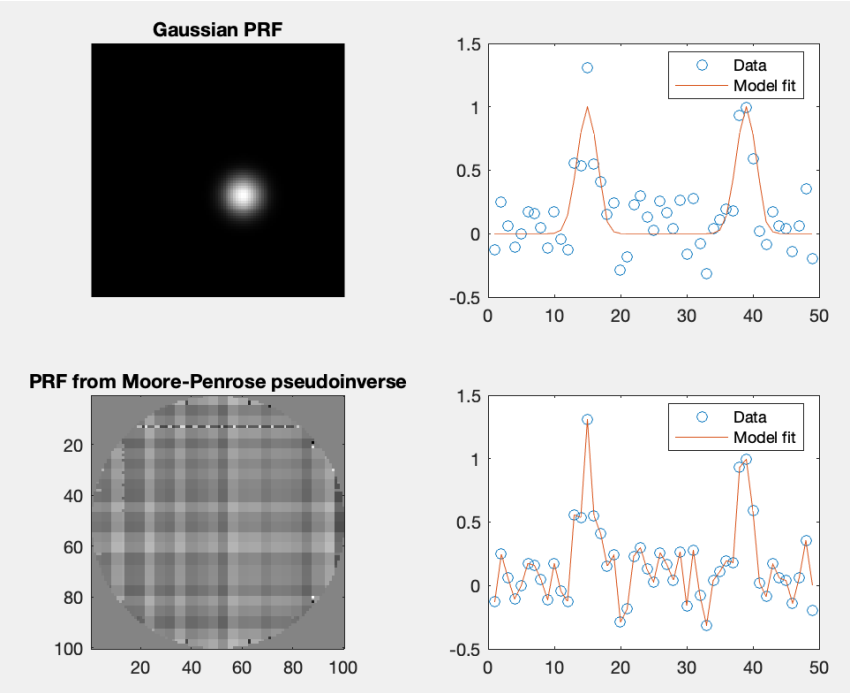
We prefer to make our assumptions about the shape explicit and not linked to the stimulus. (Note that this does not mean that the parameter estimates are unaffected by the choice of stimulus; only that the parameterization itself is unrelated to the stimulus). Why a Gaussian specifically? A small patch of tissue in visual cortex typically contains cells with reasonably similar RF center locations. Averaging many such RF envelopes, even if each individual RF is non-circular, will tend to give something spatially localized and not strongly oriented. There are various ways one could parameterize that shape, such as a raised cosine rather than a Gaussian, or an oriented shape rather than one that is isotropic, and so on. We think there is value in testing such alternatives, but doing so is beyond the scope of this paper, and in any case, we think the solutions will not be dramatically different. See for example, a simulation-based study of oriented pRFs by Lerma- Usabiaga, Winawer & Wandell, 2021 (*J Neurosci*).

Action: We now explain our choice of pRF model in the methods (**Page:** 32, **Lines** 992-1000).

12.Was the prf estimation regularized or not? Given the noise in fmri data regularization would seem to be important.

– This question is related to the previous question. The purpose of regularization is to avoid overfitting by constraining the solutions, typically at the cost of goodness of fit. We imposed a hard constraint, requiring that the solutions conform to a circularly symmetric Gaussian, which is a form of regularization. Because of this constraint, the pRF models do not explain all the variance in the data. In

other words, there is no 2D Gaussian model that would predict the exact response amplitudes for each of our 49 stimuli. In contrast, were there no regularization, we would find pRFs that predict the exact beta weights for every stimulus (for example multiplying the beta weights by the pseudoinverse of the stimulus). The two cases are shown below.



Action: We now make explicit that our assumption of a Gaussian pRF is a form of regularization that reduces overfitting (**Page:** 32, **Lines** 992-1000).

13.The attentional latency analysis seems to have been done for each retinotopic map separately. But that assumes that the attentional delay is constant per map. This is a strong assumption. Some analysis should be provided to show that this assumption is correct.

– If we were interested in the sharpness of the rise time or fall time, then yes, averaging across locations within a map that have different rise times or fall times would result in a shallower slope in the average curve than in the separate curves for each location. This kind of mischaracterization arising from averaging sigmoids with different transition points was pointed out in an interesting paper on reinforcement learning by Gallistel, Fairhurst and Balsam, 2004. However, here we are neither trying to quantify the sharpness of the rise time, nor even the mean of the rise time, but rather the *difference* in rise time (or the difference in fall time) between 1-s and 2-s trials. Since the analyses for the 1-s and 2-s trials include the same vertices, estimating the difference in rise time or fall time by subtraction is not dependent on the assumption of homogeneity within each separate analysis. Were we to separate the analysis by retinotopic location, we would greatly lower the SNR and this would impair our ability to characterize differences in rise or fall times.

Action: We now explain why we compute the average response functions for whole maps (**Page:** 33, **Lines** 1063-1069).

Reviewer #3 (Remarks to the Author):

The manuscript describes an interesting extension to the senior authors' recent work. Overall I am enthusiastic about the results, but have several comments (listed below) that the authors might wish to consider.

We thank the reviewer for their comments and suggestions.

1. Hemisphere: How did the authors deal with hemisphere? One assumes responses were averaged across hemispheres, but this is not specified clearly.

– For analyses that show data for the whole visual field, such as Figures 5A and 7B/C, data from the two hemispheres are used (but not averaged). For analyses that show results from sub-ROIs, such as the vertices with pRFs in the attended location, we average across all four wedge-ROIs corresponding to the four target locations (Figure 4, 5B/C, 7A), which includes voxels from both cerebral hemispheres, and in the case of V2 and V3, from both ventral and dorsal portions of the map. Hemispheric differences were not of special interest for the purposes of this study, but polar angle differences were. We report the statistical comparisons of BOLD amplitude comparing four angular targets with each other (**Page:** 12, **Lines:** 332-340), and we find no difference in the attentional modulation across targets. Given that the four target locations include left and right horizontal meridians (represented in the right and left hemispheres, respectively), we can infer that there are also no differences between hemispheres.

Action: We now clarify that the ROIs are all bilateral (**Page:** 12, **Lines** 328-330).

2. Sample size: N=8. I did not see any mention of a power analysis for this sample size. Is this sample commensurate with prior work looking at similar attention effects?

– We checked three previous fMRI studies on visual attention, all of which reported shifts on receptive fields, and found that the number of participants ranged from 3 to 9 (Vo, Sprague & Serences, 2017; Klein, Harvey & Dumoulin, 2014; Kay et al, 2015). Our sample size was equal to the maximum of these values, although one subject was excluded, as noted in the manuscript. Although the number of participants is relatively small, the amount of data per participant is relatively large: 1.5 hours of pre-scan psychophysics, plus 4 hours of concurrent MRI and psychophysics data from each participant, resulting in 32 hours of MRI data collected. This is consistent with the notion of extensive sampling, as advocated by Naselaris, Allen, and Kay (2021).

Action: We have now specified how the sample size was determined (in *Materials and Methods: Participants,* **Page:** 25, **Lines:** 731-733). See also the response to R2 concerning data for individual participants (Comment 6).

3. Behaviour: For completeness it would be appropriate to report the post-hoc tests between cue types.

Action: Done. We now report the post-hoc tests for cue validity in results (**Page:** 6, **Lines:** 147-150).

4. Eye-movements: Although rate (~2% of sampled frames) were time-points with fixation breaks included in the analysis?

We did not exclude trials with fixation breaks from further analysis for two reasons. First, they were rare and so we expect them to make very little impact on the results. Second, the intertrial interval was short (less than 2 seconds), and so the hemodynamic response from one trial continues into the next trial. Omitting trials from the GLM would lead to bias in the estimates of the parameters, since the unmodeled events affect the BOLD response. We can (and did) check that omitting these trials from *behavioral analysis* (which does not depend on the hemodynamics) has no effect. But our main analyses retain these trials so that the behavior and brain data are analyzed from the same set of trials.

Action: We now report that omitting these trials does not affect the pattern of behavioral responses (**Page:** 30, **Lines:** 922-923).

5. Retinotopic maps:
- What do the individual attention condition estimates look like? This is important because it is not clear if a single attention condition is driving the average responses.

– Regarding the maps themselves, the variance explained by the pRF models did not differ across five attention conditions (**Page:** 7, **Lines:** 184-186), which suggests that the pRF estimates in the averaged pRF maps is driven about equally by all five conditions.

As for the effect of attention on the BOLD amplitude, we now report the results of a linear mixed effects model showing that results do not differ across the four attended locations.

Action: We have now clarified both points in the results (pRF models: **Page:** 8, **Lines:** 194-195; linear mixed effects analysis: **Page:** 12, **Lines:** 332-340).

- The dorsal and ventral components of V2/V3 should be labelled. Even if these are averaged for subsequent analyses.
- Report the R2 of the maps used in Figure 3D.

Action: Done. We have labeled the ventral and dorsal portions of V1, V2 and V3. We have also added the variance explained of the maps in **Figure 3** underneath the flat maps of cortex.

6. Rise and Fall time analysis:

- Line 224 – what was the rationale for operationalising this as 10% of the max response before the peak? Line 267 – Again, why 10%? Does this effect hold if you change the threshold (e.g., 25%)?

– The model we fit is a logistic function, which never reaches the absolute minimum or maximum, so some threshold is required. A low threshold is typically used to characterize rise time, though exactly how low is arbitrary. We used 10% following the temporal response analysis reported in Sit, Chen, Geisler, Miikkulainen and Seidemann, *Neuron,* 2009 for temporal responses in local portions of V1. The fact that the threshold is arbitrary is not of great concern here, since we are computing *the difference* in rise time across two conditions, rather than the absolute rise time. Nonetheless, we checked and found that the results are the same whether we use a threshold of 5%, 10% or 25%, as expected. Were we interested in the absolute rise time per condition, the threshold would matter more, as higher thresholds would lead to later rise times.

Action: We repeated the attentional latency analysis with 2 new threshold values to define the rise and fall time of attention response: 5% and 25%. We added these figures as **Supplementary Figure 3.** We note that our main conclusions remain the same on this analysis, independent of the percentage of maximum response taken as the rise and fall.

- Why weren't statistics run on these data?

– We thought the patterns were clear from the statistics we reported (means and bootstrapped confidence intervals on the difference in rise-times and fall-times for 2-s vs 1-s trials): when the data are analyzed relative to the cue onset, the differences were about 0 seconds for rise times, and about 1 s for fall times, which is expected if the rise times are time-locked to the cue and the fall-times are time-locked to the target. The plots in the revision have been simplified (see response to next question), and in the simplified plots, the patterns are even clearer. The difference in rise time is about 0, and the difference in fall time is clearly above 0, about 1-s. We did not think null hypothesis testing would strengthen this observation.

Action: Nonetheless, we have computed the Bayes Factor for likelihood, which confirms what the plots show, namely that the difference in rise times (2-s trials minus 1-s trials), aligned to cue onset, are not significantly different from 0, whereas the difference in fall times are (**Page:** 11 **Lines:** 284-293: Methods: **Page:** 34-35 **Lines:** 1071-1080).

- The pattern across ROIs between cue-locked and target-locked (notwithstanding the difference in absolute value) appear very similar (i.e., Figure 4D top-panel vs bottom-panel) – could this suggest overall differences between ROIs in terms of SNR?

– It is correct that the patterns are very similar, and indeed we should not have plotted the differences for both target-locked and cue-locked analyses, as they are redundant. This is because the temporal jitter was small compared to the 1-s difference between short trials and long trials. Hence the time series analyzed for cue-locked and for target-locked were in fact the same but just shifted by one second.

Action: We have removed the redundant plots. We now only plot results time locked to the cue (**Figure 4D**).

7. Focal attention affects:
- For completeness, include the main effects of mixed-ANOVA.

Action: We have added statistical results from a linear mixed effects model on BOLD data from different visual field maps, attention conditions and target locations (**Page:** 12, **Lines:** 332-340).

- hV4 shows the strongest negative responses. The senior author has previously demonstrated the impact that the dural sinus can have on BOLD estimates in and around the borders of hV4 – One assumes this was looked at in these data? From Figure 3D the polar angle representation of hV4 extends only to the horizontal meridian (green), which would be consistent with distortion from the dural sinus.

– The pRF models required positive Gaussians. And the analyses such as position shifts of pRFs were thresholded at a relatively high variance explained (25%). Hence vertices strongly influenced by the sinus artifact that responded abnormally to the bar stimuli would generally have been excluded by our analyses.

Moreover, in response to R1's comment about the hV4 example shown in Figure 3, we redrew the boundaries of some visual field maps. The hV4 results now do not look so different from other areas. And in the example map, the hV4 coverage now extends to the lower vertical meridian.

Attentional suppression observed in hV4 is quantified by subtracting the BOLD signal measured during distributed attention from BOLD measured during focal attention. Therefore, suppression here is *relative* from the baseline levels, rather than *absolute* negative mean BOLD signal, as in Winawer et al., 2009. In the new time course plots for each ROI, the hV4 response in the attend-in and the distributed conditions are above 0. In the attend-out condition, the responses go a little below 0, but because the same vertices are used for all analyses, the positive responses in two conditions suggest little influence from artifactual signals, which Winawer et al (2010) showed were often in counterphase to the stimulus.

Action: In **Supplementary Figure 2A** we show the time courses from each visual field map, including hV4, separately for three attentional cue conditions. Response levels of hV4 are like the rest of the visual field maps, and in general, above zero across time and attention conditions. See also responses to R1 about suppression and baseline issues, especially Comment 2(2).

8. Focal attention independent of stimulus location: Again, for completeness, can the authors include the significance of these correlation coefficients?

Action: We have added the p-statistics obtained from Pearson's correlation tests (**Page**: 15, **Line**: 421).

9. pRF shifts:

- Why were statistics not computed for A?

Action: We now report the results of an ANOVA on the data plotted in **Figure 7A** (**Page:** 17, **Lines:** 461-465) to provide p-statistics.

- The basic effect of pRF centres moving towards the attended location is not novel, but it is nicely replicated here. One question that springs to mind is what happens to the pRF size estimate? If constant between attend-in and attend-out, then the shift towards the attended location will place more of the stimulus in the pRF of the voxel. An increase in pRF size plus a shift will place even more of the stimulus within the pRF of the voxel, but a decrease and a shift would not necessarily result in any more of the stimulus being in the pRF. Particularly, when the pRF centre shifts are so small (0.4 deg in LO1).

– First, on novelty: It is true that pRF shifts have been reported before, but we feel that our report goes beyond simple replication. There are not many fMRI studies that investigated the pRF center shift effect with attention, therefore, the magnitude of the attention induced pRF shifts and what it depends on, ie., task and stimulus properties, are yet to be discovered. There are discrepancies in the magnitude of the shift effect in two papers that investigated this effect. For instance, whereas Vo, Sprague and Serences (2017) report a shift magnitude of ~0.3° in V3A/B, Klein, Harvey and Dumoulin (2014) report the magnitude to be ~1° in the same visual field map.

We did not look at the effects of focal attention on pRF sizes as 1) we did not have a hypothesis as to how focal vs distributed attention would change pRF sizes and 2) size estimates from pRF model are less reliable than the center estimates (Benson et al., *JoV,* 2018; Lerma-Usabiaga et al.,*PlosCompBio,* 2020), and hence examining size estimates for the 5 separate models would likely be highly noisy due to the limited amount of data per condition.

Action: We have added the pRF size changes for focal and distributed attention as **Supplementary Figure 9** for curious readers.

- Figure 7C – why only these ROIs? Why is LO1 now averaged with V3A/B?

– V3A and LO1 have lower variance explained by the pRF models than the other maps. Because of our variance explained threshold for this analysis, these maps have fewer vertices than the others. Hence to increase our signal-to-noise ratio, we combined the two.

Action: We have added all ROIs to Figure 7B and 7C.

- Missing -5 from the x-axis of 7C (bottom-left).

Action: We have fixed the x-axis of Figure 7C.

Reviewer #1 (Remarks to the Author):

Overall, I appreciate the authors' careful attention to the comments on the previous manuscript, and think the incorporated changes enhance the paper overall.

There is an extensive response concerning the use of the phrase "suppression" that focuses on what the authors mean by suppression relative to a given baseline (here, distributed attention). If that's the way the authors wish to use that term, that's fine, but it continues to feel very non-intuitive. Most of the time, we're attending to the thing we're looking at (attend-fixation). Sometimes, we need to sample the whole display to decide where to look (distributed attention), and other times, we need to focus attention on a specific location away from fixation. I see the argument as presented in the manuscript that one case is out/out (attend fixation), the other is in/out (attend one stimulus), and the other is in/in (attend all) – but I'll reiterate that it feels a bit off to me. That said, this mostly would just serve to confuse some readers, and doesn't really gimpact the conclusions themselves (aside from terminology).

– A key result in the paper is that focal attention results in both behavioral and neural tradeoffs. Demonstrating this requires defining a baseline. We operationally defined the terms "suppression" and "enhancement" in the Results section the first time they are used (**Line**: 343-348) so the readers should not be confused.
Note that participants are never asked to attend at the fovea, and that attending to peripheral locations can result in foveal neglect (Walshe and Geisler, *Curr Bio* 2022).

The previous comment concerning pRF shifts outside the stimulus eccentricity remains. Yes, the predictions are less straightforward, but that doesn't mean the data shouldn't be shown. The authors overall have done a great job transparently illustrating their results across ROIs and steps of the analysis, with this being an exception. From my perspective, an unclear result from those voxels wouldn't impact the clarity of the result foveal to stimulus locations (as shown now), and would allow readers to relate these findings to other animal and human studies. I'd encourage the authors to reconsider this choice, even if it would just be for a supplemental figure.

Action: We have plotted additional data outside the stimulus eccentricity in Supplementary Figure 7. Note that we have made all data and code public, such that readers can not only reproduce the figures in the paper, but also explore aspects of the data not reported in the paper. The eccentricity range of these data can be easily changed in *fig7_B_directional_vector_graphs_of_shifts.m* file.

Concerning the impact of the mapping stimulus on behavior – this seems interesting, and worth reporting. There appears to be a substantial change in performance when the mapping stimulus is nearby the reported location, suggesting there is some interaction between its processing and behavioral performance. (the new line, on pg 6, line 141-143, says there is "little effect on performance", but I wouldn't agree with this characterization – the difference in d' appears to be ~0.25-0.75 for the neutral and valid datapoints). I've been thinking about this finding extensively,

and overall I don't think it's a problem with the paper (especially because performance seems impacted equivalently across conditions). But it's interesting, and may be worth including in supplemental figures should future studies be interested in elaborating on this finding.

– A d' of 4 vs 3.5 implies a hit rate of 97.7% vs 96.0%, corresponding to a difference of about 1 hit. Compared to the difference in d' between attentional conditions (4 vs 1.6 vs 0.5 for valid, neutral, invalid), this is a small difference, so our description is accurate, namely that the mapping stimulus location "had little effect on performance, relative to the large and systematic effects of cue validity". Nonetheless, we agree with the Reviewer that there may be interesting questions about the relationship between bar position and target.

Action: We will add a script to the GitHub repository for the interested reader to produce this result.

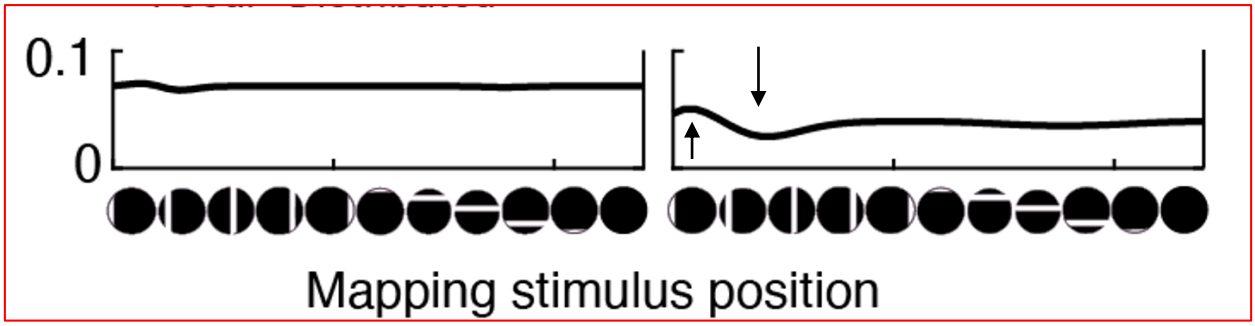
Additional comments:
1. Fig. 8 – I think I misunderstood this figure during the previous submission. Looking over the code on GitHub, which the authors should be commended for sharing so freely, it appears that the simulation was conducted with a single voxel in each ROI. However, the figure is referred to throughout the text as through it aggregates over voxels with different pRFs and shift directions/magnitudes (as in the analyses shown in Fig. 6). I'm a bit confused what this is meant to show – a shift in pRF position should change the set of stimuli that cause the voxel to respond (doesn't seem to be the case here). Additionally, for hV4, why is the response to vertical stimuli greater than to horizontal mapping stimuli? Is this due to pRF coverage extending beyond the aperture? Either the explanation of this figure should be clarified, or the simulations should be brought into alignment with the text.

– Regarding aggregating data: For each ROI, the values used for the simulation are the mean amplitude change and the mean pRF center shifts for voxels in that ROI. Hence the shifts in direction and magnitude plotted in the figure reflect typical effects for each ROI.

Regarding the position shift: The shift in pRF position is manifest as an increase in the response to some bar positions and a decrease in response to others (Figure 8, bottom panels showing the difference in response). To make this clearer, we illustrate these effects with black arrows in the plots below. In these examples, focal attention is directed leftward, which results in a leftward shift. Hence the response increases for the leftmost vertical bars (upward arrow), and decreases for the bars slightly to the right of the left edge of the aperture (downward arrow).

Regarding the vertical vs horizontal amplitudes in hV4: the hV4 pRF is larger than the bar width. For the vertical bars sweeping horizontally, the bar happens to land more squarely in the center of the pRF compared to horizontal bars sweeping vertically. This explains the slightly larger peak for the vertical bars. In V1, where the pRFs are smaller, such asymmetries are smaller.

Action: We clarified that we are using the mean response amplitude and pRF center shift changes in Figure 8 caption (**Line:** 650-652), and added that the focal attention cue is to the left (**Line:** 649). We also added arrows to the plot as in the figure below.



2. I still am hesitant to support the use of the term "distractor" to refer to the non-cued Gabor locations. These regions are relevant in 1 of 4 trials – so it's not the case that they're ignored, or considered distracting; they're just de-prioritized. I see the authors' new argument that people seem to treat a 75/25% valid/invalid cue as 100/0 (accepting chance-level performance on invalid trials) – but that's speculation on the part of the authors. The fact is those locations were relevant sometimes, just less so. Thus, they're not 'distractors' that can only impair performance when erroneously attended – they're target locations with lower priority.

– When describing the experimental procedure and stimuli, we generally use neutral terms: "four Gabor stimuli" (**Line** 110); "four Gabor patches" (Figure 1 caption); "the Gabor patches" (**Line 774**); "the four Gabor targets", (**Line** 969). In contrast, when describing the results, we refer to a "target enhancement region" and "distractor suppression region". These are widely used terms in the literature, and they are consistent with the results and definitions in the paper, as we define enhancement and suppression as positive and negative changes relative to baseline. In reviewing the manuscript, we found one exception to this pattern, in which we had used the words "target and distractor" (**Line** 175) when describing the procedure and stimuli.

Action: We removed the word "distractor" in that case to be consistent with the descriptions above this (**Line** 175).

3. fMRI pulse sequence: in the response to reviewer 2, the authors mention that they're using a multiband sequence with 6x simultaneous multi slice acceleration to scan 64 slices (mentioned in response letter, pg 15) – but I don't believe this would be possible, as I was under the impression that the number of slices was required to be an integer multiple of the multiband acceleration factor. The authors should verify this number, and should add this detail to the manuscript, as the number of slices is only presented in the response letter at present.

Action: The reviewer is correct. This was a typo. We wrote 64 instead of 66. We have added the number of slices (104 x 104 x 66) to the manuscript (**Line** 884).

4. Significant vertices for ROI definition: on line 1132 of the revised manuscript, the authors mention the vertices visualized for defining ROIs are thresholded according to the GLM variance explained, but there's no mention as to whether pRF variance explained is incorporated in the thresholding. Additionally, in the above section, the authors mention using voxels based on 3% GLM VE, again with no mention of pRF variance explained (nor an explanation as to why different thresholds are used throughout the analyses). This should be justified. (also, I was surprised that there's no mention of the visualization threshold in Fig. 3's retinotopic map in the manuscript – it's reported as > 0.2 in the response to reviewers)

–For all analyses throughout the paper, we thresholded the vertices by their variance explained from the GLM. For analysis of pRF parameters (shift of pRF center) we applied an additional threshold based on pRF variance explained (**Line:** 1154).

Action: We have clarified this in the manuscript (**Line:** 952-954). We have added the thresholding information in **Figure 3** (**Line**: 211).

5. The use of Bayesian statistics for the HRF latency analysis in Fig. 4 is a nice addition, though it is unclear why Bayesian statistics are appropriate for this analysis but not others. This figure is also missing a description of error bars for panels B and C – I assume these are CIs, as mentioned in the Methods, but readers would likely benefit from clear description of error bars in each figure caption

Action: We have added the error bar description to Figure 4 (**Line**: 272). Bayesian statistics were used because there are two plausible hypotheses (point estimates): time-locking relative to the cue (0 second difference between 2-second and 1-second mapping stimulus trials) and time-locking relative to the target (1 second difference). A Bayes factor is appropriate for comparing two specific hypotheses. For other results, such as the shift in pRF location, we did not have two specific point estimates to compare. Instead we used null hypothesis testing to ask whether the measured shift differed from 0.

6. Fig. 5 stats are missing (the LME described in the text, I believe, just verifies there's no 3-way interaction, doesn't establish other effects of interest), and no definition of error bars (CIs?)

Action: We have added a table of LME results in Supplementary Table 3 (**Page** 12), and added the error bar description to the caption (**Line**: 370)

7. New stats lines 461-465 in which shuffled vs intact model are used as a factor in ANOVA seems like a roundabout way to reject the null hypothesis that an intact model explains a different amount of variance than a shuffled one. Also, did this analysis involve only one shuffling iteration?

We did not conduct an ANOVA to compare variance explained by an intact vs shuffled pRF model. Rather, we conducted an ANOVA to compare the size of the shift due to focal attention for an intact vs shuffled model. Shuffling was done independently for each participant on a trial

by trial basis. The reason for shuffling was to estimate the amount of shift expected by measurement noise. Because there were over 2000 trials per participant, shuffling the assignment of attention condition to trial will result in groupings each of which contain about an equal number of trials from the five attention conditions. Hence any difference in pRF solutions among the 5 shuffled conditions will be due to measurement noise.
Please note that the statistical tests were requested by R3 in the prior round.

8. In Fig. 6, the caption mentions “left target meridian-selective vertices”, but the cartoon above panel A highlights the right meridian target location. Please check that these are correct

– We fixed this inconsistency.

Reviewer #1 (Remarks on code availability):

Where appropriate, I reviewed the code (see comments). This lab has an excellent track record of sharing code and data to reproduce results.

Reviewer #2 (Remarks to the Author):

I thank the authors for their very thorough and clear reply to reviewers, and to their revisions to the manuscript. The current paper is much improved over the original submission.

Positive aspects of the revised manuscript:

- Appreciate that the authors have revised text to address hemodynamic coupling. Don't agree completely with their points, but they are arguable and no one really knows the ground truth, so I am happy to leave the authors' statements as given.

- I thank the authors for correcting the visual area boundaries. The new boundaries look more plausible.

- I also thank the authors for adding the missing data for individual participants.

Issues that should be addressed in revision:

1- Because different regions of the brain have different SNR it isn't really appropriate to use one brain area to try to estimate the noise ceiling or noise floor of another area. So as far as I'm concerned it would be less misleading to simply eliminate supplementary figure 1a.

– We agree that all brain areas have different SNR, and did not mean to imply that the control region (precentral gyrus) has the same SNR as any particular visual field map. In fact, the figure plots the distribution of variance explained for 6 different visual field maps, all of which differ from one another in SNR. And yet the distributions for these 6 maps are far more similar to one

another than any of them are to the distribution of the control region. Our point was to provide an approximation of what kind of variance explained one might expect from the pRF model from voxels that are not visually responsive, *and* to show that averaging data across conditions does not tend to increase pRF variance explained for non-visually driven responses. We perhaps misled the reader by referring to this calculation as “a noise floor” in the text. In fact, we did not use the value computed from the control region distribution for any further calculations in the manuscript.

Action: To clarify our purpose, we removed the phrase “noise floor” (**Line:** 189) and modified the description of Supplementary Figure 1 (**Line:** 9-10).

2- I also have to quibble with the authors' claim that enforcing a hard pRF model is equivalent to regularization. Regularization is a method for separating signal from noise, by enforcing a prior noise model. The pRF model does not model the noise, it provides a prior form for modeling the signal. The model of the noise and the model of the signal are two different things. Others have written extensively about this issue in the regression literature…

– We removed reference to regularization (**Line:** 10078).

Other comments about issues that should **not** impede publication.

1- As the authors note the pRF model is quite complicated, and involves lots of prior assumptions. In such circumstances, the best route forward is to compare several versions of a model that differ in the aspects of the model form that are most likely to lead to confounds, artifacts, or incorrect conclusions. If all of these multiple variants lead to the same conclusion, then one can be certain that the prior assumptions did not confound the results. It would be straightforward to test several competing pRF models, and I've always been confused about why this approach isn't common. Given that there isn't really anything in this report that conflicts with prior studies I am not going to request that the authors do that this time. But I would recommend that they do it in the future.

– Thanks for the recommendation.

2- Many of the authors' responses to reviewer's complaints about experimental parameters (e.g., 1 or 2 s jitter) or model fitting (e.g., using a canonical HRF, aggregating vertices across ROIs) were based on an argument that it was done this way because it was easiest. That isn't a great argument, and it doesn't inspire confidence. There isn't really anything in this report that conflicts with prior studies, and I like this group and have a high opinion of their work in general, so I am not going to litigate all the points. But my advice is that in the future the authors should consider trying to improve their computational methods to avoid these problems.

– We agree with this notion. Perhaps we made an error in how we explained our choices. We mentioned in some places that alternative choices would have entailed large computational problems or other issues, but we also indicated the scientific problem. For example, in the case

of duration jitter, had we varied the stimulus duration continuously rather than discretely, we would have had no test-retest measure and could not have cross-validated the model. Implementing a model to try to cross-validate without repeated stimuli would have entailed making assumptions about how stimulus duration affects neural responses, a question that has no general solution even after a century of studying adaptation. Please note that we never mentioned in the text of the manuscript that choices were made for convenience; thus, this issue is limited to the response to reviewers.

3- I would advice the authors to avoid the use of MB6 in the future. MB6 is too high to provide optimal BOLD SNR, and it is even worse if combined with other forms of acceleration. This is widely known in the field.

– Thanks for the recommendation.

Reviewer #3 (Remarks to the Author):

The authors have responded to my comments. I am still unconvinced by their novelty argument, but I appreciate the work that has been put into clarifying the other points raised.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

THIRD REVIEW

Reviewer #1 (Remarks to the Author):

I appreciate the authors' careful response to my concerns. They are completely right that the d' difference I commented on reflects a very very small difference in behavior, and I should have caught that – my mistake. I have no further concerns, and look forward to citing this work in the future.

– We thank the reviewer for their constructive feedback on our work.

Reviewer #6 (Remarks to the Author):

The current study investigates how covert spatial attention cues alter population receptive field properties across visual cortex, measured via fMRI and psychophysics. Tünçok and colleagues mapped pRFs during anticipation, the moment between the cue and target stimulus onset, revealing spatially-tuned push-pull dynamics in pRFs along the visual hierarchy. More specifically, relative to distributed attention, focal attention elicited similar levels of enhancement of the task-relevant location across visual areas, with increased suppression of unattended regions along the visual hierarchy. This is an important study for the on-going quest to elucidate the neural circuits that control attention-related enhancement and suppression of visual representations.

- We thank the reviewer on their feedback and comments.

Past reviewers had several concerns, a couple of them on the authors' definition of "suppression" and "distractor". In my opinion, the authors have appropriately addressed these concerns by improving the consistency of word choice and placement; but (and if possible) would the authors be opposed to incorporating Supplementary Figure 8 as a subpanel in one of the main figures, maybe Figure 1 or Figure 5? Having that visualization in the main text could facilitate an understanding in more readers that the attend-all condition serves as the more intuitive baseline for quantifying suppression than an attend-fixation condition, where resources are already withdrawn.

- We now moved the Supplementary Figure 8 as Figure 8 to our main manuscript.

Additional comments include requests to clarify behavioral and pRF results, methodology and statistical analyses. The authors have refined and supplemented all of the above. I was brought on to evaluate whether Reviewer 2's concerns were addressed specifically. Major issues from Reviewer 2 were (1) that the noise floor/ceiling of one brain area shouldn't be used to estimate the noise floor/ceiling in another, so comparing the variance explained between brain areas in Supplementary Figure 1 might be misleading; and (2) enforcing a hard pRF model is not equivalent to regularization.

The authors state that “noise floor” was a mischaracterization on their part, as they agree that each brain area has varying levels of SNR to begin with. If anything, this point strengthens their findings in Supplementary Figure 1 that the visual areas have more similar distributions of variance explained than they do with the pre-central gyrus. Additionally, the authors omit regularization, avoiding the previous false equivalency altogether. The authors acknowledge the remainder of Reviewer 2’s recommendations for the future.

Overall, I believe the authors have appropriately addressed all Reviewer concerns and I enjoyed reading their manuscript.

Minor comments:
- Is ref. 26 missing? It does not appear in the References section on my end.
- Figure 8 has a very low resolution on my end as well.
- In Figure 1A, along the time dimension, I believe there is a small Adobe Illustrator artifact: a red square with a plus, which usually means a text box that is too small.
- Potential typo on Lines 661–662; my correction suggestion in brackets: “The task-relevant target (the Gabor) was low contrast, small and briefly presented to minimize its effect on the BOLD [signal].”
- Do all supplementary figures need to be referenced in text? If so, Supplementary Figures 5 and 6 are not.
    - We fixed all the minor formatting issues pointed out by the reviewer.