nature portfolio

Peer Review File

Orthogonalization of spontaneous and stimulus-driven activity by hierarchical neocortical areal network in primates

Corresponding Author: Professor Kenichi Ohki

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)
This paper proposes the very interesting idea that, in contrast to mice, in marmosets spontaneous and evoked activity in different brain areas become more orthogonal as one proceeds higher in the visual system. If true I think this is an important finding for the field. However I think significant improvements to the work are needed before the conclusions can be accepted.

What exactly the authors mean by "spontaneous activity" is not clear. Was there a single period of spontaneous activity before stimuli were presented? Or were they interleaved? Is the activity during the "blank period of uniform gray (4s)" what is meant? This is not explained.

The paper relies heavily on PCA. However no information regarding the variance explained against the number of PCs included is provided. For each unique case of using PCs please provide a graph of cumulative variance explained vs number of PCs.

Figures 1 and 2:

These figures consist solely of individual examples, with litte attempt at quantification or statistical testing of the claims. The Methods state that 12 marmosets were used, however there is no indication of the variability of the results between individuals. If any animals were excluded from any analyses this should be explained and justified.

A comparison is made with mouse data, but no methods or n values relating to mice are mentioned in the manuscript. Please can the authors explain why they nevertheless ticked the box in the Reporting Summary that the sample size was provided for each experimental group.

The authors say they have revealed columnar maps; however since their imaging is entirely 2-dimensional they can make no claims about any 3d structure (e.g. columns), only 2d maps.

Line 111 claims wave-like activity, however the support for this seems to be one supplementary movie. To support this claim quantification and statistical analysis across animals should be performed to demonstrate these wave-like properties.

Line 113 claims patchy activity, supported only by example pictures. Again quantification and statistical analysis is needed to demonstrate that there is indeed consistently patchy activity across areas and animals. In fact, the authors appear to have strongly biased the analysis they do perform towards finding their desired patchy result: they state they (i) applied a lowpass filter "to accentuate patchy structures", they only analyzed cases "with clearly visible patchy activation patterns", and isolated patches were then "selected manually to avoid.... the fusion of multiple patches". The authors should perform an unbiased quantitative analysis of all their data, rather than selecting only the data that support their hypothesis.

In Fig 2g it is claimed that the correlations became "successively smaller in DM and MT". However the statistical test used (Kolmogorov-Smirnov) tests only for differences in distributions, not differences in means. In order to make any claims about "bigger" or smaller" the authors need to test for differences in means using e.g. a t-test or Wilcoxon rank-sum. This error is sadly common in the neuroscience literature, so I'm sure the authors can point to many published examples where as similar

error has been made, however it's still an error.

Extended data 1a,b are not, as stated in the caption, snapshots of spontaneous activity.

It seems odd to say "it is likely that the proportion of neurons co-active in the spontaneous patches is smaller in DM and MT than in V1 and V2". Either demonstrate this with an appropriate statistical test or drop the claim.

Moving on to Figures 3 and 4:

I think these are 2-photon data, but this is not explicitly stated.

As far as I can tell the authors are not in fact using the methods of ref 3 here. The methods of that paper are strongly bound up with those authors' goal of accounting for behavioral variance, which is not relevant in the current paper. As in ref 3 please provide equations to clarify the exact methods used. The recent paper https://pubmed.ncbi.nlm.nih.gov/33871351 is also a relevant point of comparison here.

Why project out only top 50 spontaneous PCs rather the whole spontaneous space? And why then use all the trial-averaged visual responses rather than just the top 50 of those?

Fig 3b: what is meant by a "representative" PC? PCs are strictly ordered. Or do the authors mean an example from one animal of the 1st PC? In general in the biological literature it is often the case that when authors say "representative data" they actually mean "the best data we have": please demonstrate that the data shown here is actually representative in that it lies in the middle of the distribution.

Fig 3c: Please explain what is the n value that was used to calculate SEMs. It should be 12, the number of animals. In this case please show all 12 points on each bar as well as the means and SEMs. Also please confirm that the Bonferroni correction factor was 4x3x2x1 = 24, the total number of possible pairwise comparisons for the 5 cases.

Reviewer #2

(Remarks to the Author)
Matsui, Hashimoto and colleagues explored the relationship between stimulus-evoked and spontaneous activity patterns across visual areas in the marmoset and in primary visual cortex (V1) of mice. This study is the first of its kind in marmosets. The authors found that stimulus-evoked and spontaneous activity patterns were more similar in marmoset V1 compared to higher-order visual areas in the marmoset and V1 in mice. The study is rigorous experimentally and clear to read, but there are a few computational analyses that would make the results of the paper more robust. There is a discrepancy with a previous study that may be resolved by the suggested additional computational analyses.
1. The authors in Figure 2, as far as I understand, define evoked responses as trial-averaged responses, and compare these trial-averaged responses to spontaneous activity patterns. This analysis, done as just described, answers the question of how similar evoked responses are to spontaneous patterns in the overall sense. Can the authors clarify that the trial-averaged responses were used in Fig 2g? Otherwise, the authors are quantifying the fraction of spontaneous activity that continues during the stimulus-evoked periods: suppose that spontaneous activity and stimulus-evoked activity add on top of each and spontaneous activity variance is some multiple of stimulus-evoked variance, then on a single trial the correlation of a stimulus response to a spontaneous frame would be larger depending on this multiple. This multiple could change across brain areas, rather than the similarity itself. Depending on the question, the stimulus-related variance computation is the crucial analysis, rather than single trial quantification like in Figure 4c.
2. What a previous study (Stringer et al, citation [4]) found was that there is similarity between spontaneous and stimulus-evoked patterns, but the similar subspace (the shared subspace) only had one dimension which had substantial stimulus variance (the fraction of the response that was repeatable across presentations to the same visual stimuli). Once this 1-dimensional subspace was projected out of spontaneous and stimulus-evoked subspaces, the variance of the stim-only directions during spontaneous activity were low, and the variance of the spont-only dimensions during stimulus-evoked activity were low. Can the authors quantify the stimulus variance of each dimension of the 50-dimensional shared subspace in the same way as [4]? This computation consists of splitting the projection's responses into two halves, averaging over repeats in each half to obtain a vector the size of the number of unique stimuli, then correlating the two halves. This seems necessary given the results in Figure 4B are different from the results in paper [4]. Also, can the authors then perform their analyses with a one-dimensional shared subspace? I think the subsequent analyses will still hold. Just to note that the number of stimulus-related dimensions of the shared subspace is a critical question. If the dimensionality is low, then it calls into question the hypotheses of the function of spontaneous activity: a Bayesian prior or predictive coding framework embedded in spontaneous activity requires multi-dimensional interactions between stimulus patterns.
3. The authors performed the experiments in anesthetized marmosets, compared to awake mice – this distinction should be discussed as a caveat in the discussion. In fact, in anesthetized mice, spontaneous and stimulus-driven patterns in V1 have been found to be similar and low-dimensional (Okun et al, 2012, Jneurosci).

Reviewer #3

(Remarks to the Author)
General comments

Matsui et al. have investigated the neural geometry of stimulus-evoked and spontaneous activity. They compared the orthogonalization of stimulus versus spontaneous activity across multiple areas along the cortical hierarchy in marmosets, and compared this to mouse V1. This topic is highly relevant for many researchers working in systems neuroscience, especially considering that multi-species comparisons are rare, but necessary if neuroscientists wish to make claims about generalizability of findings that are often only reported in mice. The manuscript is well written and the experiments seem to be performed competently. I believe this is an interesting study with great potential, but in its current form the analyses are not described clearly enough to interpret the results with sufficient confidence. I also have a more conceptual issue about the authors' explanations of their results, in the sense that I am not convinced by their logic that the differences they find between marmosets and mice must be caused by the presence/absence of orientation columns. I believe this issue would perhaps best be resolved by adding a simple computational model to support (or perhaps provide an alternative to) their conceptual interpretations. In conclusion, this is an interesting study with merit, but it requires some revisions and additions.

Critical issues

Comment on analysis (lines 464-474): I do not fully understand the analysis the authors have performed. According to their methods, the stimulus-only and shared activity are the trial-averaged activity projected either into the spontaneous subspace to create the shared activity, or into its nullspace (by regressing out the spontaneous subspace activity) to create the stimulus-only activity. The total variance of these two projections must therefore add up to 100% variance of the original trial-averaged activity. However, when I look at figure 4C, "Stim-only" and "Shared" do not add up to 100%, and there is a non-zero component of "Spont-only" in the vis stim block. The authors state: "Spontaneous only activity was obtained by regressing out trial-averaged visual responses from each frame of spontaneous activity." (lines 470-471)
If this is the case, then I do not understand how one can project the stimulus activity into the spontaneous subspace, since this was already regressed out, and should therefore be an empty null space?
Conversely, the same holds true for projecting the spontaneous activity into the stim-only space. The analyses the authors have performed require further methodological details, because it is not clear what they are doing. They should add relevant mathematical equations in the method section that describe their geometrical analyses, because I cannot currently check whether what the authors have done makes any sense.

On the orthogonality of stimulus- and spontaneous activity patterns:
"The species-related difference may reflect the presence and
222 absence of columnar cortical circuits in marmosets and mice, respectively 35,48. Synaptic
223 inputs to a single neuron are more likely to share similar function 49 in marmoset V1 with
224 functional columns than in mouse V1 without functional columns 50. Because neurons
225 with similar functional properties tend to have high noise correlation 51 as well as
226 spontaneous correlation 52, marmoset V1 neurons are more likely to be activated by
227 correlated inputs from presynaptic neurons that share similar functional properties than
228 are mouse neurons. (Supporting Discussion 1)."
I can see how this organizational difference could more tightly constrain the possible space of neural activity in marmosets than in mice: each mouse V1 neuron receives more disordered input, so the total subspace that could potentially be occupied is larger than in marmosets, where only specific population patterns are possible due to the more ordered synaptic connectivity pattern. This line of thought makes sense to me. However, I do not see the connection between the more disordered structure in mice and why this would lead to reduced orthogonalization of spontaneous and stimulus-driven activity. The synaptic connectivity patterns do not change between stimulus and spontaneous periods, so why would they have a differential effect on neural activity orthogonalization? This is possible, but certainly not necessary. I expect rodents to have a larger traversable neural space than marmosets, but I do not understand the logic that would explain the observed difference in orthogonalization. Perhaps adding a simple computational model with rate neurons that either has orientation columns or not might help to explain how the presence of orientation columns could lead to reduced orthogonalization.

Major comments
Suggestions for further analyses:
1) Comparison of absolute dimensionality:
I already alluded to this point in my comment above, but I believe it would be very helpful to add a comparison of the intrinsic dimensionality of neural activity between the different areas, especially between mouse and marmoset V1. The analysis the authors have done now splits the space into three parts that add up to 100%, which allows for an easy comparison of the different contributions within a block. However, an equally important analysis, especially considering my above comment, is whether the intrinsic dimensionality of neural activity in mouse V1 differs from that of marmoset V1 (for a fixed number of neurons and trials).

2) Quantification of total variance in neural activity due to either trial-by-trial variability or stimulus-driven average activity
In many analyses, the authors use only trial-averaged activity, but the interpretation of these analyses hinges on the contribution of these averaged signals to the total variance in neural activity. If stimulus-averaged activity only contributes 5% to the total variance in neural activity this warrants a different interpretation of the results than if it contributes 95%. The authors could for example perform a cross-validated regression of neural activity from stimulus identities (i.e., activity prediction) to quantify the level of stimulus-driven contribution to neural activity.

Question regarding methods of mouse data:
It is somewhat ambiguous what the recording and stimulus parameters for mouse V1 were. The authors state the recordings

were done in a FoV of 800 μm × 800 μm in 512 × 512 pixels, but it is not clear whether this was for marmosets only, or also for mice. For a mouse recording this seems rather large, but not impossibly so. Could the authors state more clearly what the methods for the mouse recordings were? If different settings were used, this can affect SNR; and measurement noise components might show up as shared components between spontaneous/stim driven activity. If settings differed between rodents and primates, this can therefore affect the orthogonalization estimates, so the authors should rule out this possibility. If not, the authors should just state more clearly that the parameters applied to both marmosets and mice.

Comment on DM data: the DM data are recorded in only 1 animal with 3 FOVs (line 448), which leads to rather wide error bars in various analyses. It might be better to either record more DM data, or remove the DM data from the manuscript, as V1 => V2 => MT is already sufficient to make some statements about the progression of dimensionality and orthogonalization across the cortical visual hierarchy.

Comment on data pooling: it seems the authors take each FoV from each animal as an independent data point, regardless of whether those FoVs came from the same or different animals. It would be good if the authors add a supplementary figure where they average the results over FoVs for each animal first, and then show the results as mean +/- sem over animals, where n=number of animals. As the data are presented now, some animals may strongly dominate the results if the number of FoVs per animal is not uniformly distributed.

Minor comments
Line 38, “higher mammals”: this term is not a meaningful biological descriptor to refer to primates versus rodents as both orders are placental mammals. It appears from the rest of the paper the authors use this description to imply a distinction in animals with and without orientation columns, but even then “higher mammals” is not a useful way of describing this difference, as it obfuscates the relevant difference between the two species (i.e., the existence of orientation columns). Moreover, the evolution of orientation columns is still unclear: it might have either evolved multiple times in for example both Carnivora and Primates independently, or alternatively, the ancestral state is in fact the existence of orientation columns, which may have been lost by Rodentia as space-saving adaptation. In this latter case, mice would therefore represent the “more evolved” species. Either way, the authors should avoid the term “higher mammals” in this context.

Line 82, “activitis”: misspelt, should probably be activity.

Line 223, “orothogonalization”: misspelt, should probably be orthogonalization.

Figure 3A: “Project OUT Trial. Averaged Vis. Resp” should probably be “Project OUT Trial Averaged Vis. Resp PC”, as the visual response space without restricting it to a limited set of principal components encompasses the full neural space, unless true collinearities exist between pairs of neurons (which probably is not the case).

References: I understand it is impossible to cite all prior work on population coding, geometric analyses and computations, but I feel the authors might have overlooked some important (recent) papers in the field, such as the ones below. Please note these are merely reading suggestions: it is entirely up to the authors whether they wish to cite any of these papers, and this is certainly not a requirement:
https://www.nature.com/articles/s41593-019-0477-1
https://www.cell.com/neuron/fulltext/S0896-6273(19)30053-4
https://www.cell.com/cell-reports/fulltext/S2211-1247(16)30996-2
https://www.nature.com/articles/nn.3707
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006446
https://www.nature.com/articles/s41593-022-01088-4

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Thanks for thoroughly addressing my feedback with new analyses, the claims in the paper are now well-supported.

Also, I would recommend that the authors share at least one dataset from each brain area (e.g. one marmoset's recordings) -- this data is impressive and unique and will greatly benefit the community.

In fact, Nature Communications requires the sharing of a minimum dataset: "The data availability statement must make the conditions of access to the “minimum dataset” that are necessary to interpret, verify and extend the research in the article, transparent to readers. This minimum dataset may be provided through deposition in public community/discipline-specific repositories, custom proprietary repositories for certain types of datasets, or general repositories like Figshare, Zenodo and Dryad."

Reviewer #3

(Remarks to the Author)
General comments

The revised version of the paper by Matsui et al. has improved considerably. I especially appreciate the expanded method section that now provides much more information on their analyses. There seem to be some inconsistencies in the description of their methods, however, so the authors should double-check if everything is correct.
I also like the addition of the simple computational model that aims to show that the anatomical organization in orientation columns can explain the difference in orthogonalization of spontaneous and stimulus-evoked activity between mouse and marmoset V1 in terms of 1) the correlation between spontaneous and stimulus-evoked activity patterns and 2) the fraction of shared neural space between spontaneous and stimulus-evoked neural activity. While I like the model, its limits should be presented in a more nuanced discussion. As I will explain in more detail below, I do not believe the interpretation of their model's result is as clear-cut as the authors currently make it out to be.
Finally, while this is not a critical issue, there are some awkward phrasings throughout the manuscript (e.g., line 166: "we next conducted geometrical analysis introduced in the recent mouse studies"). I therefore suggest the authors give their manuscript another proofread before submitting the final production version.
In summary, the manuscript is clearly improved, but some outstanding issues remain.

Remaining issues
Model: In the authors' new model, the implementation of the difference between the marmoset and mouse is not a columnar structure versus salt-and-pepper, as they imply with panel A in Extended Data Figure 15. Instead, the difference between species in their model arises only from a difference in the width of a circular convolution. The absence of orientation columns in mice will indeed cause an effectively wider integration of inputs, but the same would happen in species where orientation columns exist, but the integration window is wider. In fact, this is exactly how the authors implemented it in their model. The conclusion of the model in my opinion is therefore that the defining factor is the synaptic integration width over the space of orientation-tuning rather than the existence of columns. The absence of columns is merely one way of widening this integration. That said, the neural architecture of the mouse cortex would lead to a wider integration than in marmosets, so it's not that their model is wrong or inappropriate; their model is just more general than the specific difference the authors wish to address here (columns vs no columns). The authors should clarify in the manuscript that there are limits to the biophysiological plausibility of their model, and that it is not in fact simulating columns vs no columns, but simulates the secondary effect that the presence of orientation columns would have on the synaptic integration width.

PCA: I appreciate the extended method section, which now includes more details than in the previous version. Most steps are now clear to me, but I do not understand why the sponta-PC and sponta-only PC analyses, as shown in Extended Data Figure 7, do not have same total number of components. See for example the V2 data: in the sponta-PC case, there is a data set (FOV) that has <200 components, while in the sponta-only PC plot all datasets have >200 components. Since the analyses were performed on the same datasets, how is this possible? The total number of components (i.e., neurons) should be identical, regardless of the type of analysis performed on it.

Method section inconsistencies:
While the methods are improved, they are rather sloppy, as I identified some inconsistencies:
Line 344 says "three Thy1-GCaMP6 mice (4 males; GP 4.3, JAX#024275 (…)". Is it three or four? It cannot be both.
Line 447 states the FOV is "800 μm × 800 μm", while the authors' response to my question about the FOV in the rebuttal was that it was 500 x 500. Again, it cannot both be true.
While these are most likely typos, this does not fill me with confidence that the rest of the method section is reliable, as there are many things that I cannot check and have to take the authors' word for. I suggest the authors carefully check their method section again to make sure what they write is in fact how they performed the experiments and analyses, and no errors (typographic or otherwise) remain in their manuscript.

Minor comments:
1) I found the following sentence in the abstract rather difficult to parse:
"Stimulus-evoked and spontaneous activities show orthogonal
(dissimilar) patterns in the primary visual cortex (V1) of mice (4-6), which is likely to be
beneficial for separating sensory signals from internally generated noise (5, 7-12);
however, those in V1 of carnivores and primates show highly similar patterns (3, 13-17)."
It might be better to remove the semicolon here. On a related note, I would also remove the reference to carnivores, as the current study is only investigating rodents and primates.

2) Line 130: the comparison is missing statistical testing. Based on the figures, I believe the difference in patch width is highly significant, but a t-test's p-value (or something similar) should be added to the manuscript.

3) Line 137 (and elsewhere): The Bonferroni correction is very conservative, so the authors might want to consider instead using a Bonferroni-Holm correction, which leads to fewer Type II errors. However, if they feel this is not worth the effort, a Bonferroni correction is fine too.

4) Line 628-629: "PCs that explained less than 3% of variance were discarded."
How many PCs remained for the various data sets, and more importantly, what was the total amount of explained variance of the remaining components? If 50 PCs were used initially, this step likely reduced that number by a considerable amount.

5) Line 640: "shard" should be "shared".

6) Extended Data Figure 2 c,d, parentheses on y-axis labels are inverted

7) Extended Data Figure 4 could be added to figure 1, but I leave that up to the authors

8) Extended Data Figure 5: I would add the median values corresponding to the data in Figure 2g to this comparison, as figure 2g shows cumulative plots but not the median correlation values

9) Extended Data Figure 8-9: a (short) description of the results in the figure legend is missing.

10) Extended Data Figure 11: I believe lines 1042-1047 were erroneously copied here, as they do not seem to relate to what is shown in the figure

11) Line 1070, “factions” should be “fractions”.

12) Extended Data Figure 15: the legend has several language errors, so I suggest the authors have another look at this. E.g.: “Each unit has preferred orientation.”, “are spatially structure by”, etc.

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
The latest round of revisions of the manuscript by Matsui et al. address all my previous comments. I would like to congratulate the authors with an excellent study. The inter-species comparison must have been a difficult and laborious project, but it adds something unique to the literature. Bold studies like these are what push the frontiers of science, so the authors should be proud of what they have achieved, and I hope they will continue their rigorous work in the future. I have no further comments.

Reviewer #1 (Remarks to the Author):

This paper proposes the very interesting idea that, in contrast to mice, in marmosets spontaneous and evoked activity in different brain areas become more orthogonal as one proceeds higher in the visual system. If true I think this is an important finding for the field. However I think significant improvements to the work are needed before the conclusions can be accepted.

We thank the reviewer for positive evaluation of our study and many valuable comments. We have extensively revised the manuscript and added new analyses/figures in response to the reviewer's comment.

What exactly the authors mean by "spontaneous activity" is not clear. Was there a single period of spontaneous activity before stimuli were presented? Or were they interleaved? Is the activity during the "blank period of uniform gray (4s)" what is meant? This is not explained.

Spontaneous activity was recorded separately from visually evoked responses in dedicated scans. During scans for spontaneous activity recordings, the LCD monitor used for visual stimulation was turned off. For each FOV, we conducted a spontaneous activity scan either before or after the visual stimulus scan(s). The interval between the two scans was no longer than 30 min. We have described this point in the Methods section (page 17, lines 470-475).

The paper relies heavily on PCA. However no information regarding the variance explained against the number of PCs included is provided. For each unique case of using PCs please provide a graph of cumulative variance explained vs number of PCs.

According to the reviewer's comment, we have added a new figure to show graphs of cumulative variance explained vs. the number of PCs for all cases where we used PCA (New Extended Data Figure 7a). We have also added figures to show the singular values for cases in which we used SVD (New Extended Data Figure 7b) (page 7, line 176).

Figures 1 and 2:

These figures consist solely of individual examples, with litte attempt at quantification or statistical testing of the claims. The Methods state that 12 marmosets were used, however there is no indication of the variability of the results between individuals. If any animals were excluded from any analyses this should be explained and justified.

In response to the reviewer's comment, we have added new figures to show the variability in the results between individual animals (New Extended Data Figure 5). In the revised manuscript, we used data

obtained from 12 animals. Of the 12 animals, 10 were used to obtain widefield imaging data shown in the figures [Large FOV covering the occipital-parietal cortex, 2 animals; V1/V2, 3 animals; MT, 4 animals; frontal cortex, 1 animal]. We also collected two-photon imaging data from nine animals (V1, 12 FOVs in 3 animals; V2, 23 FOVs in 4 animals [2 were not used in the widefield imaging data); MT, 11 FOVs in 4 animals]. We have added these points to the Animals section of the Methods section (page 13, lines 350-355). As we have described in the original manuscript (page 17, lines 447-449 in the original manuscript), for two-photon imaging, we selected FOVs (1) with stable recordings of both visual stimulation scan and spontaneous activity scan and (2) with a sufficiently large number of active cells (> 100). FOVs that did not meet these criteria were excluded from the analyses.

A comparison is made with mouse data, but no methods or n values relating to mice are mentioned in the manuscript. Please can the authors explain why they nevertheless ticked the box in the Reporting Summary that the sample size was provided for each experimental group.

We apologize for the unclear description of the mouse data in our original manuscript. In the original manuscript, we stated in the "Two-photon Data Analysis" section of the Methods that the data of seven FOVs from three mice were used, and we ticked the box in the Reporting Summary because of this statement. In the revised manuscript, for better visibility, we have also described the number of mice in the Methods section. During the revision, we included additional data from 1 mouse. Thus, the mouse data in the revised manuscript include data on 12 FOVs from four mice. We have added these points to the Methods section (page 13, lines 354-355).

The authors say they have revealed columnar maps; however since their imaging is entirely 2-dimensional they can make no claims about any 3d structure (e.g. columns), only 2d maps.

We agree with the reviewer that 3-dimensional data are necessary to strictly determine the columnar structure. In the revised manuscript, we have rephrased the term "columnar" to "patchy" wherever the term appeared.

Line 111 claims wave-like activity, however the support for this seems to be one supplementary movie. To support this claim quantification and statistical analysis across animals should be performed to demonstrate these wave-like properties.

According to the reviewer's comment, we have made two revisions 1) we have added an additional example and 2) we have removed the original statement that the wave-like activity was propagating across the cortex. Regarding the additional data, we have added a new Supplementary Movie of widefield spontaneous activity in the large FOV covering occipital-parietal cortex of another marmoset (Supplementary Video 4). The followings are rationale for the second point of our revision: According to the reviewer's comment, we conducted the new analysis which analyzed wave propagations in

calcium imaging data (Linden et al., 2021). Most pixels had velocities close to zero (Figure Ad-e), and the upper bound of the overall velocity of the GBA was 17.2 ± 7.7 μm/s (mean ± s.d.) (Figure Af), which was considerably slower than the speed of spontaneous activity propagation reported in mice (16 mm/sec; Han et al., 2008) and humans (0.2 m/sec; Mitra et al., 2014). Based on the extremely slow speed of GBA, we conclude that the present data lacks evidence for the propagation of spontaneous activity. Therefore, in the revised manuscript, we have deleted the description related to the dynamics of the widefield spontaneous activity. We thank the reviewer for this valuable comment.

The followings are the details of the analysis of the wave propagations in calcium imaging data (Linden et al., 2021). In this analysis, calcium imaging data were preprocessed using temporal high-pass filtering (>0.1Hz), spatial lowpass filtering (a Gaussian filter with σ = 365 μm), spatial down-sampling to 128 by 128 pixels (from 256 by 256 or 512 by 512 pixels) and pixel-wise z-normalization. Because we observed that the positions of the patchy structures were relatively stable while active cortical regions appeared to traverse across the cortex, we used spatial low-pass filtering to reveal large-scale activity patterns while removing the patchy texture on top. Spatial down sampling was performed to satisfy the computational power requirements. After preprocessing, we first detected the periods in which wave-like activity was likely to be present. Similar to the study by Linden et al., 2021 and to our previous study (Matsui et al., PNAS, 2016), we calculated the time course of the fraction of active pixels (pixels exceeding $z > 1.5$) within the FOV (see Figure Aa attached below). We then extracted peaks of the time course whose height exceeded 0.4 (red dots in Figure Aa). Periods of global activity activities (GBA) were defined as a 10-s time window centered at these peaks. For each period of the global activity waves, similar to Linden et al., we calculated the optical flow using the Horn-Shunk method (as implemented in the Matlab function opticalflowHS) (Figure Ab-Ac). Most pixels had velocities close to zero (Figure Ad-e). To estimate the upper bound of the overall velocity of the GBA, we quantified the mean velocity of the top 10 percentile pixels. Across all 28 GBAs, the mean of the GBA velocity was 17.2 ± 7.7 μm/s (mean ± s.d.) (Figure Af), which was considerably slower than the speed of spontaneous activity propagation reported in mice (16 mm/sec; Han et al., 2008) and humans (0.2 m/sec; Mitra et al., 2014). Based on the extremely slow speed of GBA, we conclude that the present data lacks evidence for the propagation of spontaneous activity.

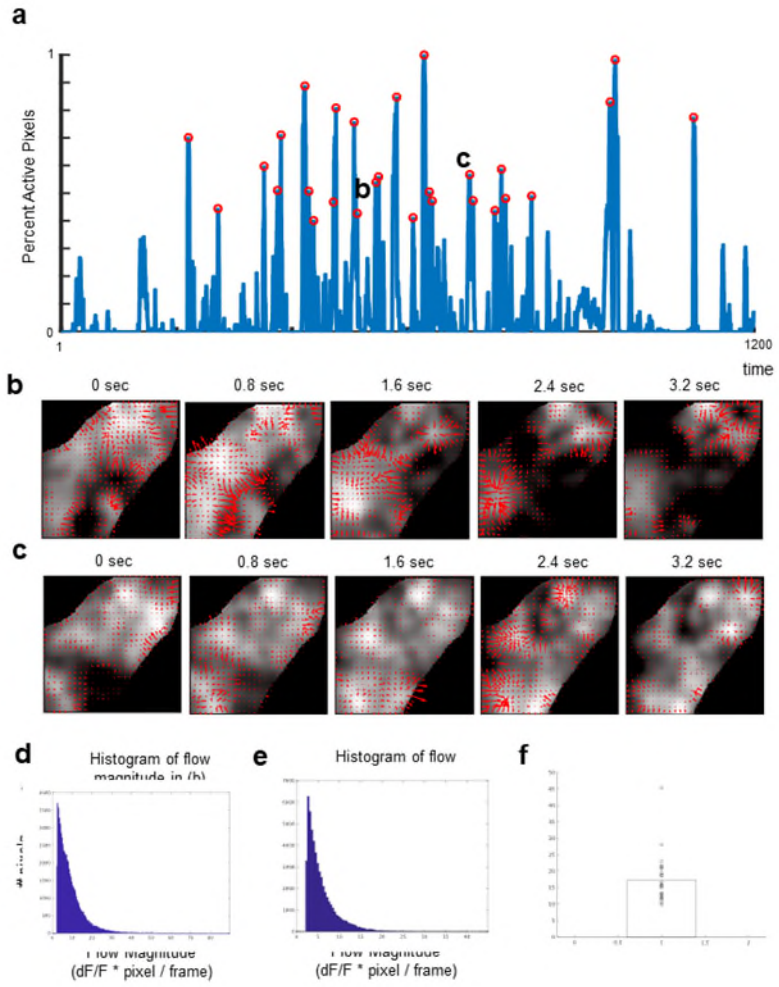

Figure A

Reference

Han F, Caporale N, Dan Y. 2008 Reverberation of recent visual experience in spontaneous cortical waves. Neuron 60(2): 321-327.

Linden NJ, Tabuena DR, Steinmetz NA, Moody WJ, Brunton SL, Brunton BW. 2021 Go with the FLOW: visualizing spatiotemporal dynamics in optical widefield calcium imaging. J. R. Soc. Interface18:20210523.

Mitra A, Snyder AZ, Hacker CD, Raichle ME. 2014 Lag structure in resting-state fMRI. J Neurophysiol. 111(11): 2374-2391.

Line 113 claims patchy activity, supported only by example pictures. Again quantification and statistical analysis is needed to demonstrate that there is indeed consistently patchy activity across areas and animals. In fact, the authors appear to have strongly biased the analysis they do perform towards finding their desired patchy result: they state they (i) applied a lowpass filter "to accentuate patchy structures", they only analyzed cases "with clearly visible patchy activation patterns", and isolated patches were then "selected manually to avoid.... the fusion of multiple patches". The authors should perform an unbiased quantitative analysis of all their data, rather than selecting only the data that support their hypothesis.

We apologize for haven misspelled the low-”cut” filter as a lowpass in the original manuscript (page 16, line 719 in the original manuscript). A high-pass (low-cut) filter is used instead of a low-pass filter. We have corrected this error accordingly (page 33, line 919). We apologize for this mistake and thank the reviewer for noting it. We have also deleted the phrase “to accentuate patchy structures” in the revised manuscript because the phrase might have caused a false impression that the preprocessing of the data artificially created patchy spatial patterns. In the followings, we first explain the reason why we believe that the highpass filtering is less likely to bias the data toward those matching with particular hypothesis. Then, according to the reviewer’s comment (“The authors should perform an unbiased qualitative analysis”), we explain strategy of the new quantitative analysis added in this revision, and explain results obtained with the new analysis. The results of the new analysis quantitatively support the notion that patchy spatial patterns of the spontaneous activity reflect biological structure of the marmoset neocortex, rather than mere artifacts derived from data processing.

The reviewer has raised a valid concern that the use of spatial filters may have artificially created patchy spatial patterns in the data. However, we believe that this is not the case for two reasons. First, a high-pass filter was used instead of a bandpass filter. Whereas bandpass filters may artificially bias patchy patterns by selectively suppressing both low- and high-spatial frequency components, high-pass filters are less biased towards a particular frequency band. Second, patchy spatial patterns that resemble the spatial patterns of the orientation columns are readily visible in the data without high-pass filtering (Supplementary Video 2-3; Please note that, as indicated in the figure legends, these supplementary videos are not spatially filtered). Therefore, we believe that the patchy spatial pattern of spontaneous activity was not an artifact created by data preprocessing. We have added these points to the Methods section of the revised manuscript (page 18, lines 504-511).

Nevertheless, we agree that we were somewhat biased toward analyzing patchy spontaneous activity. This is because many previous studies, particularly the seminal work by Grinvald and colleagues, have reported that spontaneous cortical activity in the primary visual areas of carnivores and primates shows patchy spatial patterns (Kenet et al., 2003; O’Hashi et al. 2018; Smith et al., 2018). Hence, we consider it important to check the generalizability of the previous observations beyond the

level of V1. We have clarified this point in the revised manuscript (page 9, lines 240-244).

In response to the reviewer's comment ("The authors should perform an unbiased qualitative analysis"), we also added a new quantitative analysis of the spatial patterns of widefield spontaneous activity. For this analysis, we followed procedures commonly employed in previous studies, namely spatial filtering and PCA (Smith et al., 2018; O'Hashi et al., 2018). In the first step of the analysis, because the spatial patterns of spontaneous activity observed in widefield imaging have highly heterogeneous spatial structures (e.g. large activity patterns and small patchy patterns; see Supplementary Movies), spatial filtering is commonly applied to remove spontaneous activities with very low spatial frequencies. Decomposition methods such as PCA are commonly used to obtain the underlying spatial patterns from the data in an unbiased manner (Kenet et al., 2003; O'Hashi et al., 2018). We added a new Extended Data Figure 13 to show the top 20 PCs for the two of the representative data. We added these points to the revised manuscript (page 18, lines 500-502, 511-514).

We understand the reviewer's concern that the above procedure (spatial filtering + PCA) tends to bias the analysis towards the detection of patchy spatial patterns. It is important to show that patchy spatial patterns obtained using decomposition methods are biologically meaningful. Previous studies have shown that some patchy spatial patterns obtained from spontaneous activity are overlaps with the columnar patterns of visual cortical responses (Kenet et al., 2003; O'Hashi et al., 2018). Like the previous studies, we conducted an additional analysis and found that, in V1 and V2, some of the PCs obtained from spontaneous activity resembled (patchy) visual responses. As shown in the new Extended Data Figure 6, we confirmed this result in both widefield imaging covering only the V1/V2 areas and widefield imaging covering a large field of view spanning the occipital to parietal areas (original Figure 1c-d). Unlike V1/V2, in area MT, no PC resembled the spatial patterns of visual responses. The difference in similarity between PCs and visual response patterns was statistically significant across animals ($p < 0.02$, two-sample t-test; new Extended Data Figure 6c), which was consistent with the results of cellular-scale imaging (original Figures 2). We have added these points to the revised manuscript (page 6, lines 157-160).

Regarding the across areal difference in the spontaneous patches, in response to the reviewer's concern, we have deleted our original analysis of the position dependence of the patch size along the anterior-posterior axis (original Extended Data Figure 2) and instead added a new analysis comparing spontaneous patches in V1 and V2 in a more data-driven manner (new Extended Data Figure 2). We have deleted the original analysis for the following reasons: 1) this analysis did not contribute to the main conclusion of this study, 2) the data used in this analysis were small (n=2 animals), and 3) the analysis was performed by manually selecting spontaneous patches (which was necessary to select patches that were isolated from the other patches and/or blood vessels). In the new analysis, similar to previous studies (e.g., O'Hashi et al., 2018), we first performed PCA on widefield

spontaneous activity and then obtained a PC that best correlated with the visually evoked activity. We chose to analyze widefield imaging data covering V1 and V2 in the same FOV because these data allowed us to compare the two cortical areas without being affected by differences in animals and/or other experimental conditions. In this analysis, we performed autocorrelation analysis separately for V1 and V2 and compared the sizes of the two autocorrelation profiles. In all three examined data, we found that the half-width-half-maximum of the autocorrelation profile was larger for V2 than for V1 (mean difference, 19.9 %). Because this analysis used widefield imaging data obtained from one FOV per animal, we did not have sufficient number of data to perform statistical testing. We have added these points to the revised text (page 5 line 124 – page 6, line 133; page 18, lines 514-522).

Reference

Smith, G.B., Hein, B., Whitney, D.E. et al. Distributed network interactions and their emergence in developing neocortex. Nat Neurosci 21, 1600–1608 (2018).

O'Hashi, K., Fekete, T., Deneux, T., Hildesheim, R., van Leeuwen, C., Grinvald, A. Interhemispheric Synchrony of Spontaneous Cortical States at the Cortical Column Level, Cerebral Cortex, 28(5), 1794–1807 (2018).

In Fig 2g it is claimed that the correlations became "successively smaller in DM and MT". However the statistical test used (Kolmogorov-Smirnov) tests only for differences in distributions, not differences in means. In order to make any claims about "bigger" or smaller" the authors need to test for differences in means using e.g. a t-test or Wilcoxon rank-sum. This error is sadly common in the neuroscience literature, so I'm sure the authors can point to many published examples where as similar error has been made, however it's still an error.

According to the reviewer's comment, we conducted the Wilcoxon rank-sum test for statistical testing, as shown in Figure 2g.

Extended data 1a,b are not, as stated in the caption, snapshots of spontaneous activity.

We thank the reviewer for pointing this out. We have corrected the figure captions accordingly.

It seems odd to say "it is likely that the proportion of neurons co-active in the spontaneous patches is smaller in DM and MT than in V1 and V2". Either demonstrate this with an appropriate statistical test or drop the claim.

We thank the reviewer for carefully reading the manuscript. We apologize for the insufficient explanation of this point in the original manuscript. In widefield imaging, we observed spontaneous activity of the marmoset visual cortex showing patchy patterns in all areas tested (V1, V2 and MT; Please note that we have removed the DM results in response to Reviewer #3's comment). However,

at the cellular level (i.e. two-photon imaging), we observed some differences in spontaneous activity patterns across areas; spontaneously active neurons appeared less clustered in the MT than in the V1 and V2. Extended Figure 4 in the original manuscript quantitatively tests this by comparing the spatial clustering of spontaneously co-active neurons. We found that the proportion of coactive neurons in spatial proximity was significantly lower in the MT than in the V1 and V2 ($p < 0.0001$, two-sample t test corrected by Bonferroni's method), consistent with the notion that spontaneously co-active neurons were less clustered in the MT than in the V1 and V2. We have clarified these points in the revised manuscript (page 6, lines 133-141).

Moving on to Figures 3 and 4:

I think these are 2-photon data, but this is not explicitly stated.

Yes. We used cellular level data obtained by two-photon imaging. We have clarified this point in the revised figure captions.

As far as I can tell the authors are not in fact using the methods of ref 3 here. The methods of that paper are strongly bound up with those authors' goal of accounting for behavioral variance, which is not relevant in the current paper. As in ref 3 please provide equations to clarify the exact methods used. The recent paper https://pubmed.ncbi.nlm.nih.gov/33871351 is also a relevant point of comparison here.

According to the reviewer's comment, we have extensively revised the description of the methods using mathematical equations (page 20, lines 570- page 23, line 641). We would like to note that we indeed followed the analysis of Stringer et al. The reviewer is correct in that the focus of the study by Stringer et al. was to account for behavioral variance, a goal that is not related to the current study. Nevertheless, Stringer et al. analyze "spontaneous activity", as in this study, without considering the behavior. In the third paragraph of the section entitled "Stimulus-evoked and ongoing activity overlap along one dimension", Stringer et al. describe the result as follows "Similarly, the space of ongoing activity, defined by the top 128 principal components of spontaneous firing, contained 23% of the total stimulus-related variance, 86% of which was contained in one dimension (85% positive weights)". Also in the method section, Stringer et al. write that "An identical analysis was used to define the stimulus-spontaneous shared dimension and spontaneous-only subspace by replacing the subspace of the top 32 behavioral components with the subspace of the top 128 principal components of activity computed in one-half of the spontaneous period". We used the same procedure except that we used top 50 principal components of activity instead of 128. We have noted this point in the revised manuscript (page 21, lines 587-597).

We also thank the reviewer for the highly relevant study by Avitan and colleagues (Avitan et al., Elife, 2021). Avitan et al. reported that in the zebrafish optic tectum, spontaneous and visually

evoked activity patterns become less similar and geometrically diverse throughout development. It is of great interest for future studies to examine whether the relationship between the spontaneous and evoked activity in the mouse V1 follows a similar developmental pattern. We have added these points to the Discussion section of the revised manuscript (page 10, lines 269-273).

Reference

Avitan, L., Pujic, Z., Mölter, J., Zhu, S., Sun, B., & Goodhill, G. J. (2021). Spontaneous and evoked activity patterns diverge over development. Elife, 10, e61942.

Why project out only top 50 spontaneous PCs rather the whole spontaneous space? And why then use all the trial-averaged visual responses rather than just the top 50 of those?

In our spontaneous activity scans, the number of frames (>1000) exceeded the number of cells (<1000). Therefore, the full dimensions of the spontaneous space are equal to the number of cells. Because the visual responses were recorded in the same set of cells, if we used all spontaneous spaces, the entire variance of the visual responses would be projected onto the spontaneous space. Thus, we limited the spontaneous space to smaller dimensions. To set the number of PCs to use, we referred to a previous study by Stringer et al., where they used the first 128 dimensions of spontaneous PCs for the recoding of ~10000 cells. Because the number of cells recorded in our study was approximately 1000 at most (~500 on average), we reasoned that 50 PCs would be sufficient to construct a shared space. To confirm that the present results did not depend on the exact number of PCs used, we conducted the same analyses with 100 spontaneous PCs (new Extended Data Figures 14). The results were similar with the original results obtained by using 50 spontaneous PCs.

For trial-averaged visual responses, the number of frames of trial-averaged visual responses (64 frames) was smaller than that of the cells. Thus, to obtain a spontaneous-only space, we projected out all trial-averaged visual responses from the spontaneous activity (to avoid overestimating the spontaneous-only space). We have added these points to the Methods section (page 21, lines 587-599 and page 22, lines 610-614).

Fig 3b: what is meant by a "representative" PC? PCs are strictly ordered. Or do the authors mean an example from one animal of the 1st PC? In general in the biological literature it is often the case that when authors say "representative data" they actually mean "the best data we have": please demonstrate that the data shown here is actually representative in that it lies in the middle of the distribution.

In the original Fig. 3b, we have selected “representative PCs” based on its similarity with the (trial-averaged) stimulus-evoked responses. This was because the 1st PC of Shared PC and Evoked-Only PC did not always have the highest spatial correlation with the stimulus-evoked responses. Nevertheless, we understand the reviewer’s concern that this selection of PC, and the use of the term “representative”,

is somewhat arbitrary. Therefore, we have used 1st PCs in the revised Fig. 3b. We have replaced the term “representative PC” by “1st PC” in the revised manuscript (page 34, line 951).

Fig 3c: Please explain what is the n value that was used to calculate SEMs. It should be 12, the number of animals. In this case please show all 12 points on each bar as well as the means and SEMs. Also please confirm that the Bonferroni correction factor was 4x3x2x1 = 24, the total number of possible pairwise comparisons for the 5 cases.

The n value in Fig.3c indicates the number of FOVs in two photon imaging. As we have explained in the response to the reviewer’s comment above (The comment starting with “Figures 1 and 2: These figures consist solely of…”), we used data obtained from a total of 12 animals. Of the 12 animals, 9 were used to collect two-photon imaging data in 9 of animals [V1, 12 FOVs in 3 animals; V2, 23 FOVs in 4 animals (2 were not used in widefield imaging data); MT, 11 FOVs in 4 animals]. We have clarified these points in the revised manuscript (page 13, lines 350-354). In the revised figures, we have shown all points as well as the mean and SEM. To confirm that the overall trend was similar across animals, we have also added new figures with each point representing each animal (new Extended Data Figure 5).

Regarding the Bonferroni correction factor, we used ${}_4C_2 = 6$ (number of pairwise comparisons). We have clarified this point in the figure legend (page 34, lines 959-960). Please note that the data for the DM have been removed according to Reviewer #3’s comment.

We thank the reviewer for the helpful comments.

Reviewer #2 (Remarks to the Author):

Matsui, Hashimoto and colleagues explored the relationship between stimulus-evoked and spontaneous activity patterns across visual areas in the marmoset and in primary visual cortex (V1) of mice. This study is the first of its kind in marmosets. The authors found that stimulus-evoked and spontaneous activity patterns were more similar in marmoset V1 compared to higher-order visual areas in the marmoset and V1 in mice. The study is rigorous experimentally and clear to read, but there are a few computational analyses that would make the results of the paper more robust. There is a discrepancy with a previous study that may be resolved by the suggested additional computational analyses.

We thank the reviewer for positive evaluation of our study (*"This study is the first of its kind in marmosets… The study is rigorous experimentally and clear to read, but there are a few computational analyses that would make the results of the paper more robust."*). We hope that added analyses and revisions satisfactorily address the reviewer's remaining concerns.

1. The authors in Figure 2, as far as I understand, define evoked responses as trial-averaged responses, and compare these trial-averaged responses to spontaneous activity patterns. This analysis, done as just described, answers the question of how similar evoked responses are to spontaneous patterns in the overall sense. Can the authors clarify that the trial-averaged responses were used in Fig 2g? Otherwise, the authors are quantifying the fraction of spontaneous activity that continues during the stimulus-evoked periods: suppose that spontaneous activity and stimulus-evoked activity add on top of each and spontaneous activity variance is some multiple of stimulus-evoked variance, then on a single trial the correlation of a stimulus response to a spontaneous frame would be larger depending on this multiple. This multiple could change across brain areas, rather than the similarity itself. Depending on the question, the stimulus-related variance computation is the crucial
analysis, rather than single trial quantification like in Figure 4c.

Yes, the question we asked in Figure 2 (and most of Figures 3-4) is how similar the evoked responses are to spontaneous patterns in the overall sense. As the reviewer correctly pointed out, we used trial-averaged visual responses to answer this question. Figure 4c is the only one in which we used single-trial evoked responses. We have clarified these points in the revised manuscript (page 6, lines 143-144; page 7, line 189). Regarding the trial-to-trial variance of the stimulus-evoked responses, in response to the comment of Reviewer#3, we have added a new analysis showing that 30-50% of the variance of the single-trial activity was explained by the trial-averaged evoked activity (Extended Data Figure 8; page 7, lines 183-188).

2. What a previous study (Stringer et al, citation [4]) found was that there is similarity between spontaneous and stimulus-evoked patterns, but the similar subspace (the shared subspace) only had one dimension which had substantial stimulus variance (the fraction of the response that was repeatable across presentations to the same visual stimuli). Once this 1-dimensional subspace was projected out of spontaneous and stimulus-evoked subspaces, the variance of the stim-only directions during spontaneous activity were low, and the variance of the spont-only dimensions during stimulus-evoked activity were low. Can the authors quantify the stimulus variance of each dimension of the 50-dimensional shared subspace in the same way as [4]? This computation consists of splitting the projection's responses into two halves, averaging over repeats in each half to obtain a vector the size of the number of unique stimuli, then correlating the two halves. This seems necessary given the results in Figure 4B are different from the results in paper [4].

In accordance with the reviewer's suggestion, we quantified the stimulus variance of each dimension of the 50-dimensional shared subspace in the same manner as that described by Stringer et al. Specifically, we used half of the trials (even trials) to calculate the Shared PCs and the other half (odd trials) to calculate the projected variance. Overall, PCs in marmoset V1 and V2 had larger projected variances compared to those in marmoset MT and mouse V1, consistent with the relatively large, shared subspace in marmoset V1 and V2 than in the other two areas. We have added a new Extended Data Figure 9 and described the results in the revised Results section (page 8, lines 199-202).

We agree with the reviewer's opinion that the present results of the stimulus-related variance in the shared space in mouse V1 seems different from those of Stringer et al. In our mouse V1 data, the variance projected onto the first PC was approximately 8% and comparable to the 2nd and 3rd PCs. In Stringer et al., the variance projected onto the first PC was approximately 15% and larger than the other PCs. One potential reason for this discrepancy is that mice in the present study were lightly anesthetized, whereas those in the Stringer study were awake. Stringer et al. reported that the one-dimensional behavior-stimulus shared subspace can be interpreted as a multiplicative gain modulation of cortical stimulus responses (see their Fig. 4L). Such multiplicative gain modulation has been linked to attentional modulation of cortical activity (McAdams and Maunsell, 1999; Reynolds and Heeger, 2009). It is likely that the anesthetic condition used in the present study strongly attenuated this component of neuronal activity (in fact the variance projected to the 1st PC in our data was 8% whereas that in Stringer et al. was 15%).

We chose to use anesthesia in the mouse experiments because the marmoset experiments were performed under anesthesia. Thus, we believe that the difference in the number of dimensions of the shared space between the marmoset and mouse is unlikely to be attributable to differences in the experimental conditions. We clarified the rationale for the use of anesthesia in mouse experiments in the Discussion section (page 10, lines 279-283).

Also, can the authors then perform their analyses with a one-dimensional shared subspace? I think the subsequent analyses will still hold. Just to note that the number of stimulus-related dimensions of the shared subspace is a critical question. If the dimensionality is low, then it calls into question the hypotheses of the function of spontaneous activity: a Bayesian prior or predictive coding framework embedded in spontaneous activity requires multi-dimensional interactions between stimulus patterns.

In accordance with the reviewer's suggestion, we performed the analysis using one-dimensional shared subspaces. We found that the stimulus-related variance projected to the one-dimensional shared subspace systematically decreased along the cortical hierarchy, as expected. We have added a new Extended Data Figure 10 and described the results in the revised Results section (page 8, lines 202-203).

To estimate the number of dimensions of the shared subspace, in the new Extended Data Figure 12a, we counted the number of PCs with a large projected variance. In marmoset V1, the estimated number of dimensions in the shared subspace was three, which was larger than the dimensionality of mouse V1. These results suggest that, even if spontaneous activity in marmoset V1 serves as a Bayesian prior, it may only provide rudimentary information about the visual scene (e.g., orientation or color). We have added these points to the revised manuscript (page 8, lines 211-215 and page 11, lines 314-316).

3. The authors performed the experiments in anesthetized marmosets, compared to awake mice – this distinction should be discussed as a caveat in the discussion. In fact, in anesthetized mice, spontaneous and stimulus-driven patterns in V1 have been found to be similar and low-dimensional (Okun et al, 2012, Jneurosci).

We thank the reviewer for highlighting this important point. We would like to kindly note that we used anesthetized mice in the present study (as opposed to Stringer et al. who used awake mice). We agree with the reviewer that anesthesia could affect spontaneous and stimulus-driven activity patterns. Okun and colleagues recorded spontaneous and evoked activity in the primary auditory cortex (A1) of anesthetized rats and, unlike the Stringer et al. study conducted in awake V1, showed considerable similarity between spontaneous and evoked neuronal activity patterns (Okun et al., J Neurosci, 2012). Thus, the use of anesthesia could affect the degree of similarity between spontaneous and evoked neuronal activity. In the present study, to account for this possibility and minimize the effect of anesthesia on the comparison between mouse V1 and marmoset V1, we conducted mouse experiments under anesthetized conditions comparable to those used for marmosets. Hence, the difference between marmoset V1 and mouse V1 found in the present study is unlikely to be explained solely by anesthesia. However, some differences in the results of Stringer et al. and the present study in mouse V1 (e.g.

larger dimensionality of the shared space) may be attributable to the use of anesthesia. We have added these points to the revised Discussion section (page 10, line 274 - page 11, line 286).

Reference

Okun, M., Yger, P., Marguet, S. L., Gerard-Mercier, F., Benucci, A., Katzner, S., ... & Harris, K. D. (2012). Population rate dynamics and multineuron firing patterns in sensory cortex. Journal of Neuroscience, 32(48), 17108-17119.

We thank the reviewer for the insightful feedback.

Reviewer #3 (Remarks to the Author):

General comments

Matsui et al. have investigated the neural geometry of stimulus-evoked and spontaneous activity. They compared the orthogonalization of stimulus versus spontaneous activity across multiple areas along the cortical hierarchy in marmosets, and compared this to mouse V1. This topic is highly relevant for many researchers working in systems neuroscience, especially considering that multi-species comparisons are rare, but necessary if neuroscientists wish to make claims about generalizability of findings that are often only reported in mice. The manuscript is well written and the experiments seem to be performed competently. I believe this is an interesting study with great potential, but in its current form the analyses are not described clearly enough to interpret the results with sufficient confidence. I also have a more conceptual issue about the authors' explanations of their results, in the sense that I am not convinced by their logic that the differences they find between

marmosets and mice must be caused by the presence/absence of orientation columns. I believe this issue would perhaps best be resolved by adding a simple computational model to support (or perhaps provide an alternative to) their conceptual interpretations. In conclusion, this is an interesting study with merit, but it requires some revisions and additions.

We thank the reviewer for positive evaluation of the study (*"I believe this is an interesting study with great potential… In conclusion, this is an interesting study with merit, but it requires some revisions and additions."*). According to the reviewer's comments, we have with extensively revised the manuscript with additional analyses and figures. Particularly, we have added a new computational simulation that support our conceptual interpretation that the presence/absence of orientation columns (at least partially) explain orthogonality of spontaneous and evoked activities.

Critical issues

Comment on analysis (lines 464-474): I do not fully understand the analysis the authors have performed. According to their methods, the stimulus-only and shared activity are the trial-averaged activity projected either into the spontaneous subspace to create the shared activity, or into its nullspace (by regressing out the spontaneous subspace activity) to create the stimulus-only activity. The total variance of these two projections must therefore add up to 100% variance of the original trial-averaged activity. However, when I look at figure 4C, "Stim-only" and "Shared" do not add up to 100%, and there is a non-zero component of "Spont-only" in the vis stim block.The authors state: "Spontaneous only activity was obtained by regressing out trial-averaged visual responses from each frame of spontaneous activity." (lines 470-471)

If this is the case, then I do not understand how one can project the stimulus activity into the spontaneous subspace, since this was already regressed out, and should therefore be an empty null

space?

Conversely, the same holds true for projecting the spontaneous activity into the stim-only space. The analyses the authors have performed require further methodological details, because it is not clear what they are doing. They should add relevant mathematical equations in the method section that describe their geometrical analyses, because I cannot currently check whether what the authors have done makes any sense.

We apologize for the insufficient description of this method in the original manuscript. We have extensively revised the Methods section related to this part of the analysis including the relevant mathematical equations (page 20, lines 557- page 23, line 661).

Regarding the reviewer's comment *"when I look at figure 4C, "Stim-only" and "Shared" do not add up to 100%, and there is a non-zero component of "Spont-only" in the vis stim block … If this is the case, then I do not understand how one can project the stimulus activity into the spontaneous subspace, since this was already regressed out, and should therefore be an empty null space?"*, we believe that this is because "Sponta-only" and "Shared" spaces were determined by "trial-averaged visual response", but single-trial visual response not only contained "trial-averaged visual response" components but also "trial-to-trial variability" components. Spatial patterns of the latter components in our data likely contained spatial patterns of both "Shared" and "Sponta-only" PCs. It has been reported that non-visual ("spontaneous") components have an additive effect on the visual responses (Shimaoka et al., eLife, 2019; Stringer, et al., Science, 2019; Musall et al., Nature Neuroscience, 2019), and this additive interaction likely contributed to the "trial-to-trial variability" projected to "Sponta-only" space. We have noted these points in the revised manuscript (page 23, lines 650-656).

Regarding the second part of the reviewer's question *"Conversely, the same holds true for projecting the spontaneous activity into the stim-only space."*, we believe that this was because we projected out top 50 PCs of spontaneous activity, instead of the full PCs, to define "Stim-only" space. The residual PCs of spontaneous activity that were not used to define "Stim-only" space could have contributed to the projection to the "Stim-only" space. We have mentioned these points in the revised manuscript (page 23, lines 657-661). We would like to note that a similar situation may be found in the previous study by Stringer and colleagues (Stringer et al., Science, 2019). In their Fig.4G and 4H, we can see small but non-zero projected variance on stim-only subspace during the spontaneous activity period (sponta period).

We did not use the full PCs of spontaneous activity because of the following reason. In our spontaneous activity scans, the number of frames (>1000) exceeded the number of cells (<1000). Therefore, the full dimensions of the spontaneous space are equal to the number of cells. Because the visual responses were recorded in the same set of cells, if we used all spontaneous spaces, the entire variance of the visual responses would be projected onto the spontaneous space. Thus, we limited the spontaneous space to smaller dimensions. To set the number of PCs to use, we referred to a previous

study by Stringer et al., where they used the first 128 dimensions of spontaneous PCs for the recoding of ~10000 cells. Because the number of cells recorded in our study was approximately 1000 at most (~500 on average), we decided to use (=project out) top 50 PCs of the spontaneous activity to construct stim-only space.

On the orthogonality of stimulus- and spontaneous activity patterns:
"The species-related difference may reflect the presence and
222 absence of columnar cortical circuits in marmosets and mice, respectively 35,48. Synaptic
223 inputs to a single neuron are more likely to share similar function 49 in marmoset V1 with
224 functional columns than in mouse V1 without functional columns 50. Because neurons
225 with similar functional properties tend to have high noise correlation 51 as well as
226 spontaneous correlation 52, marmoset V1 neurons are more likely to be activated by
227 correlated inputs from presynaptic neurons that share similar functional properties than
228 are mouse neurons. (Supporting Discussion 1)."
I can see how this organizational difference could more tightly constrain the possible space of neural activity in marmosets than in mice: each mouse V1 neuron receives more disordered input, so the total subspace that could potentially be occupied is larger than in marmosets, where only specific population patterns are possible due to the more ordered synaptic connectivity pattern. This line of thought makes sense to me. However, I do not see the connection between the more disordered structure in mice and why this would lead to reduced orthogonalization of spontaneous and stimulus-driven activity. The synaptic connectivity patterns do not change between stimulus and spontaneous periods, so why would they have a differential effect on neural activity orthogonalization? This is possible, but certainly not necessary. I expect rodents to have a larger traversable neural space than marmosets, but I do not understand the logic that would explain the observed difference in
orthogonalization. Perhaps adding a simple computational model with rate neurons that either has orientation columns or not might help to explain how the presence of orientation columns could lead to reduced orthogonalization.

In accordance with the reviewer's comment, we created simple two-layer neural network models that mimicked the marmoset or mouse V1 to test whether a more disordered structure in mice leads to the orthogonalization of spontaneous and stimulus-driven activity (new Extended Data Figure 15). A two-layer fully connected network was used as the base architecture for both the marmoset and the mouse V1 models. In this model, the inputs to the network were orientation-selective neural activity (such as the neural activity in layer 4 of mouse V1) (L1 in Extended Data Figure 15a). The neurons in the input layer are connected to the top layer (L2) in an all-to-all fashion but with different weights. Ordered vs. disordered architectures (i.e. columnar vs. salt-and-pepper architectures) of the marmoset and mouse V1, respectively, are modeled by two distinct profiles of connection weights as

functions of difference in preferred orientations of the connected neurons (right panels in the new Extended Data Figure 15a). In the marmoset V1 model, a weight profile is narrowly tuned for difference in the preferred orientation of the connected neurons. Thus, a neuron in L2 preferentially samples inputs from L1 neurons sharing similar preferred orientation. In contrast, the mouse V1 model has a broadly tuned weight profile with respect to the difference in preferred orientations, such that a neuron in L2 samples inputs from L1 neurons with diverse preferred orientations. The weight profile of the mouse V1 model was chosen based on previous physiological studies reporting that mouse L2/3 neurons receive broadly orientation selective inputs (Ko et al., 2011).

We ran simulations of visually evoked activity and spontaneous activity using these networks (bottom panels in the new Extended Data Figure 15a): For the simulation of spontaneous activity, we injected a structured noise to L1 neurons. The structured noise is modeled based on previous reports showing similarity of signal correlation and noise correlation. In both primates and rodents, previous studies (Kohn & Smith, J. Neurosci., 2005; Ko et al., Nature, 2011) showed that V1 neurons with high signal correlation, hence similar preferred orientations, also showed high noise correlation and spontaneous activity correlation. To simulate these previous observations, we convolved a white noise by using a Von Mises function to introduce noise correlation between L1 units sharing similar preferred orientations. The new Extended Data Figure 15b shows examples of simulated spontaneous activity in the mouse V1 and marmoset V1 models. Although the structured noise inputs to L1 neurons are the same for the marmoset V1 model and the mouse V1 model, correlated activities in L1 neurons are weighted more by each L2 neuron in the marmoset V1 model than the mouse V1 model because of a narrower tuning of the weight profile in the marmoset V1 model. Thus, L2 neurons sharing similar preferred orientations in the marmoset V1 model should show more correlated spontaneous activity. On the other hand, L2 neurons sharing similar preferred orientation in the mouse V1 model should show less correlated spontaneous activities, because each L2 neuron put more weights on diverse (less correlated) inputs from L1 neurons. Consistently, the activity correlation between L2 neurons revealed a more structured correlation matrix for the marmoset V1 model than for the mouse V1 model (right panels in the new Extended Data Figure 15b). To simulate visually evoked activity, in addition to the structured noise, we injected activity tuned at a selected preferred orientation into L1 neurons, mimicking the presentation of oriented gratings. The new Extended Data Figure 15c shows examples of simulated visually evoked activity in the mouse V1 and marmoset V1 models. Activity correlation between L2 neurons (“signal correlation”) was similar for the two models (right panels in new Extended Data Figure 15c). The orientation tuning curve of each L2 neuron was narrower in the marmoset V1 model than in the mouse V1 model, mimicking previous animal studies (Niell and Stryker, 2008; Yu and Rosa, 2014). These correlation structures suggest that spontaneous activity patterns and visually evoked activity patterns were similar, at least in the correlation structure, in the marmoset V1 model but not in the mouse V1 model.

To confirm that the simulation reproduced essential patterns of the animal results, using the simulated data, we conducted the same analyses that we used for the animal data. As shown in the Main Figure 2g, the maximum correlation value between the spontaneous and evoked-activity frames were significantly higher in the marmoset V1 model than in the mouse V1 model ($p < 10^{-42}$, rank-sum test; new Extended Data Figure 15e). Similarly, as in shown in Main Figure 4a, the fraction of the variance of visually evoked activity projected to the shared space was larger for the marmoset V1 model than for the mouse V1 model ($p < 10^{-33}$, rank-sum test across 100 instances of model pairs; new Extended Data Figure 15f). These simulation results are consistent with animal data and suggest that spontaneous and visually evoked activity are more orthogonalized in the mouse V1 model than in the marmoset V1 model. Taken together, these results suggest that a more disordered connection leads to greater orthogonalization between spontaneous and evoked activities. We have added these results to the revised Supporting Discussion 1 (page1, line 16- page 6, line 155 of Supplementary Discussion).

We thank the reviewer for thoughtful suggestions.

Ko H, Hofer SB, Pichler B, Buchanan KA, Sjostrom PJ, Mrsic-Flogel TD. Functional specificity of local synaptic connections in neocortical networks. Nature 473(7345):87-91.

Niell CM, Stryker MP. 2008 Highly selective receptive fields in the mouse visual cortex. J Neurosci. 28(30): 7520-7536.

Yu HH, Rosa MGP. 2014 Uniformity and diversity of response properties of neurons in the primary visual cortex: Selectivity for orientation, direction of motion, and stimulus size from center to periphery. Visual Neurosci. 31(1): 85-98.

Major comments

Suggestions for further analyses:

1) Comparison of absolute dimensionality:

I already alluded to this point in my comment above, but I believe it would be very helpful to add a comparison of the intrinsic dimensionality of neural activity between the different areas, especially between mouse and marmoset V1. The analysis the authors have done now splits the space into three parts that add up to 100%, which allows for an easy comparison of the different contributions within a block. However, an equally important analysis, especially considering my above comment, is whether the intrinsic dimensionality of neural activity in mouse V1 differs from that of marmoset V1 (for a fixed number of neurons and trials).

According to the reviewer's comment, we estimated the intrinsic dimensionality of neural activity in the shared subspace by counting the number of PCs with large explained variance. We have added a new figure (new Extended Data Figure 12a) to show the results. In marmoset V1, the estimated number of dimensions in the shared subspace was three, which was larger than the dimensionality of

mouse V1. These results suggest that spontaneous activity in the marmoset V1 may provide rudimentary information about the visual scene (e.g., orientation or color). We have added these points to the revised Results (page 8, lines 211-215) and Discussion (page 11, lines 314-318).

We also conducted a control analysis using fixed numbers of cells and trials (number of cells = 400 and number of trials = 20). Consistent with the original results, we found that the stimulus-related variance projected onto the shared subspace was significantly larger for marmoset V1 than for mouse V1. Similarly, the dimensionality of the shared subspace was larger for marmoset V1 than for mouse V1. We added a new figure to show these results (new Extended Data Figure 12b-c)(page 8, lines 205-208; page 20, lines 567-569).

2) Quantification of total variance in neural activity due to either trial-by-trial variability or stimulus-driven average activity

In many analyses, the authors use only trial-averaged activity, but the interpretation of these analyses hinges on the contribution of these averaged signals to the total variance in neural activity. If stimulus-averaged activity only contributes 5% to the total variance in neural activity this warrants a different interpretation of the results than if it contributes 95%. The authors could for example perform a cross-validated regression of neural activity from stimulus identities (i.e., activity prediction) to quantify the level of stimulus-driven contribution to neural activity.

According to the reviewer's comment, we have performed a cross-validated regression analysis of neural activity from activity prediction. Cross-validation was performed using half of the trials to make the predictor (i.e., trial-averaged neural activity) and the other half of the trials to estimate the trial-to-trial variance. The new Extended Data Figure 8 shows the fraction of the variance explained by stimulus-driven neural activity. In marmoset V1, V2, MT, and mouse V1 ranged between 30-50%. Thus, although large trial-to-trial variability still existed, the trial-averaged neural activity explained a substantial fraction of the single-trial stimulus-evoked responses in all tested visual areas. We have added these points to the revised manuscript (page 7, lines 183-188; page 20, lines 563-567).

Question regarding methods of mouse data:

It is somewhat ambiguous what the recording and stimulus parameters for mouse V1 were. The authors state the recordings were done in a FoV of 800 μm × 800 μm in 512 × 512 pixels, but it is not clear whether this was for marmosets only, or also for mice. For a mouse recording this seems rather large, but not impossibly so. Could the authors state more clearly what the methods for the mouse recordings were? If different settings were used, this can affect SNR; and measurement noise components might show up as shared components between spontaneous/stim driven activity. If settings differed between rodents and primates, this can therefore affect the orthogonalization estimates, so the authors should rule out this possibility. If not, the authors should just state more clearly that the parameters applied to

both marmosets and mice.

We apologize for the lack of clarity in the original manuscript. The 800 um × 800 um in 512 × 512 pixels with a ×16 objective were used for the marmoset. For the mouse, we used 500 um × 500 um in 512 × 512 pixels with a ×25 objective. We have added this point and other details of the mouse recordings to the revised Methods section (page 16, lines 446-448).

The reviewer raised a valid concern that different settings might cause differences in the SNR between mice and marmosets. In particular, because of the difference in FOV size, the number of pixels allocated to one neuron was larger for the mouse than for the marmosets, which could yield a better SNR for the mouse. To rule out the possibility that this difference affected the overall results, we re-analyzed the mouse data, with the number of pixels allocated to each cell being comparable to that of the marmoset. The median numbers of pixels allocated to each cell in the marmoset and mouse V1 were 110 and 208, respectively. To match the number of pixels allocated to each cell in mouse V1, we randomly selected 100 pixels and used them to obtain new cell time courses. These new time courses were then used to calculate the stimulus-related variance projected onto a shared space. The overall results of this control analysis did not differ from the original results; the value of the projected variance obtained with the new time courses of mouse V1 was smaller than that of the marmoset visual cortex (new Extended Data Figure 11). Moreover, the projected variance of the mouse V1 calculated using the new and original time courses did not show a statistically significant difference (p>0.1, Wilcoxon signed-rank test). We have added these points to the revised manuscript (pages 8, lines 203-205; page 19, lines 535-538).

Comment on DM data: the DM data are recorded in only 1 animal with 3 FOVs (line 448), which leads to rather wide error bars in various analyses. It might be better to either record more DM data, or remove the DM data from the manuscript, as V1 => V2 => MT is already sufficient to make some statements about the progression of dimensionality and orthogonalization across the cortical visual hierarchy.

According to the reviewer's suggestion, we have deleted the DM data.

Comment on data pooling: it seems the authors take each FoV from each animal as an independent data point, regardless of whether those FoVs came from the same or different animals. It would be good if the authors add a supplementary figure where they average the results over FoVs for each animal first, and then show the results as mean +/- sem over animals, where n=number of animals. As the data are presented now, some animals may strongly dominate the results if the number of FoVs per animal is not uniformly distributed.

According to the reviewer's comment, we have added new figures showing the results across animals (i.e., n=number of animals) (new Extended Data Figures 5). These new results show that the

overall results were the same even when the data were summarized for individual animals.

Minor comments

Line 38, “higher mammals”: this term is not a meaningful biological descriptor to refer to primates versus rodents as both orders are placental mammals. It appears from the rest of the paper the authors use this description to imply a distinction in animals with and without orientation columns, but even then “higher mammals” is not a useful way of describing this difference, as it obfuscates the relevant difference between the two species (i.e., the existence of orientation columns). Moreover, the evolution of orientation columns is still unclear: it might have either evolved multiple times in for example both Carnivora and Primates independently, or alternatively, the ancestral state is in fact the existence of orientation columns, which may have been lost by Rodentia as space-saving adaptation. In this latter case, mice would therefore represent the “more evolved” species. Either way, the authors should avoid the term “higher mammals” in this context.

We agree with the reviewer that the terms “higher/lower mammals” are not thoughtful biological descriptors. In the revised manuscript, we have rephrased these terms as primates/carnivores and rodents, respectively. When appropriate, we have also used the terms “mammals with/without functional columns” to replace “higher/lower mammals”.

Line 82, “activitis”: misspelt, should probably be activity.

Line 223, “orothogonalization”: misspelt, should probably be orthogonalization.

We thank the reviewer for pointing out misspellings. We have corrected the mistakes.

Figure 3A: “Project OUT Trial. Averaged Vis. Resp” should probably be “Project OUT Trial Averaged Vis. Resp PC”, as the visual response space without restricting it to a limited set of principal components encompasses the full neural space, unless true collinearities exist between pairs of neurons (which probably is not the case).

In this analysis, we treated each frame of the trial-averaged visual evoked responses as an individual visual response pattern (frames corresponding to stimulus-off periods were discarded). Depending on the number of orientations/directions tested, the total number of trial-averaged visual responses for each set of stimuli ranged from 32 to 64. Because the number of trial-averaged visual responses was smaller than the number of recorded cells in each FOV (>100), visual response space did not encompass full neural activity space which has number of dimensions equal to the number of recorded neurons. Thus, we projected out all the trial-averaged visual responses instead of projecting out some PCs of the trial-averaged responses. We have clarified this point in the revised Methods section (page 20, lines 571-574 and page 22, 610-614).

References: I understand it is impossible to cite all prior work on population coding, geometric analyses and computations, but I feel the authors might have overlooked some important (recent) papers in the field, such as the ones below. Please note these are merely reading suggestions: it is entirely up to the authors whether they wish to cite any of these papers, and this is certainly not a requirement:

https://www.nature.com/articles/s41593-019-0477-1

https://www.cell.com/neuron/fulltext/S0896-6273(19)30053-4

https://www.cell.com/cell-reports/fulltext/S2211-1247(16)30996-2

https://www.nature.com/articles/nn.3707

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006446

https://www.nature.com/articles/s41593-022-01088-4

We thank the reviewer for providing a list of relevant papers. As there is no restriction on the number of references in this journal, we have cited and discussed all the papers in the revised manuscript.

Semedo and colleagues reported that the macaque V1 preferentially communicated specific activity patterns, not necessarily the largest ones, to V2 (Semedo et al., Neuron, 2019). We have mentioned this point in the revised manuscript (page 11, lines 305-306). It is of great interest to investigate whether such specific inter-areal communication enables the progressive orthogonalization of spontaneous and evoked activity in a hierarchical visual network.

Similar to the findings of Rumyantsev, Schnitzer et al. (Rumyantsev et al., Nature, 2020), Montijn and colleagues reported that in the high-dimensional neural activity space, trial-to-trial variability of visually evoked activity occurs in a direction orthogonal to the direction encoding the stimulus orientation (Montijn et al., Cell Reports, 2016). We have mentioned this in the revised manuscript (page 11, line 300).

Progressive orthogonalization of spontaneous and evoked activity parallels the hierarchical processing of visual information in the dorsal pathway. A recent preprint by Manning et al. suggested that neural coding in V1 and V2 is optimized for sensory representation whereas that in the MT is optimized for visual discrimination (Manning et al., 2023). Ruff and Cohen consistently reported that visual attention further optimizes stimulus representation by MT neurons by aligning it to the axis of the activity space, which is important for guiding behavior. We have mentioned and discussed these points in the revised manuscript (page 12, lines 326-330).

Recent studies on recurrent neural networks have suggested that the dimensionality of spontaneous and evoked neural network activity is related to network connectivity. Across areas and across species, differences in the dimensionality of the shared space may also be attributable to differences in local connectivity patterns (Recanatesi et al., Plos Comp Bio, 2019; Dubreuil et al., Nature Neurosci, 2022). We discussed these points in the revised manuscript (page 10, lines 266-269).

Reference

Manning et al., Transformations of sensory information in the brain reflect a changing definition of optimality. bioRxiv, 2023 Mar 25;2023.03.24.534044. doi: 10.1101/2023.03.24.534044.

We thank the reviewer for thoughtful comments and suggestions.

Reviewer #2 (Remarks to the Author):

Thanks for thoroughly addressing my feedback with new analyses, the claims in the paper are now well-supported.

Also, I would recommend that the authors share at least one dataset from each brain area (e.g. one marmoset's recordings) -- this data is impressive and unique and will greatly benefit the community.

In fact, Nature Communications requires the sharing of a minimum dataset: "The data availability statement must make the conditions of access to the “minimum dataset” that are necessary to interpret, verify and extend the research in the article, transparent to readers. This minimum dataset may be provided through deposition in public community/discipline-specific repositories, custom proprietary repositories for certain types of datasets, or general repositories like Figshare, Zenodo and Dryad."

Reviewer #2 (Remarks on code availability):

The shared code does not include the code for performing analyses, so it does not enable the reproduction of all the figures. However, I do not think improving the code is as important as the authors sharing the data needed to replicate some of the findings of the paper.

We thank the reviewer for a positive evaluation of the revised manuscript.
According to the reviewer’s suggestion, we shared example marmoset data (V1, V2, MT and mouse V1) in Figshare (doi:10.6084/m9.figshare.25448167).

Reviewer #3 (Remarks to the Author):

General comments

The revised version of the paper by Matsui et al. has improved considerably. I especially appreciate the expanded method section that now provides much more information on their analyses. There seem to be some inconsistencies in the description of their methods, however, so the authors should double-check if everything is correct.

We thank the reviewer for a positive evaluation of the revised manuscript. We sincerely apologize for typos and other mistakes in the manuscript. In this revision, we have thoroughly checked the manuscript, and also used a professional English editing service to check the manuscript.

I also like the addition of the simple computational model that aims to show that the anatomical organization in orientation columns can explain the difference in orthogonalization of spontaneous and stimulus-evoked activity between mouse and marmoset V1 in terms of 1) the correlation between spontaneous and stimulus-evoked activity patterns and 2) the fraction of shared neural space between spontaneous and stimulus-evoked neural activity. While I like the model, its limits should be presented in a more nuanced discussion. As I will explain in more detail below, I do not believe the interpretation of their model's result is as clear-cut as the authors currently make it out to be.

We appreciate the reviewer's positive comment on our computational model. Regarding the interpretation of the modeling results, please see our response to the reviewer's comment related to this point (1st comment in "Remaining issues").

Finally, while this is not a critical issue, there are some awkward phrasings throughout the manuscript (e.g., line 166: "we next conducted geometrical analysis introduced in the recent mouse studies"). I therefore suggest the authors give their manuscript another proofread before submitting the final production version.

In summary, the manuscript is clearly improved, but some outstanding issues remain.

We apologize again for typos and other mistakes in the manuscript. In this revision, we have thoroughly proofread the manuscript, and also used a professional English editing service to check the manuscript.

Remaining issues

Model: In the authors' new model, the implementation of the difference between the marmoset and mouse is not a columnar structure versus salt-and-pepper, as they imply with panel A in Extended Data Figure 15. Instead, the difference between species in their model arises only from a difference in the width of a circular convolution. The absence of orientation columns in mice will indeed cause an effectively wider integration of inputs, but the same would happen in species where orientation columns exist, but the integration window is wider. In fact, this is exactly how the authors implemented it in their model. The conclusion of the model in my opinion is therefore that the defining factor is the synaptic integration width over the space of orientation-tuning rather than the existence of columns. The absence of columns is merely one way of widening this integration. That said, the neural architecture of the mouse cortex would lead to a wider integration than in marmosets, so it's not that their model is wrong or inappropriate; their model is just more general than the specific difference the authors wish to address here (columns vs no columns). The authors should clarify in the manuscript that there are limits to the biophysiological plausibility of their model, and that it is not in fact simulating columns vs no columns, but simulates the secondary effect that the presence of orientation columns would have on the synaptic integration width.

In this model, we assumed that neurons in the columnar cortical circuit preferentially receive inputs from functionally similar neurons. Although similar bias toward inputs from functionally similar neurons exists in the salt-and-pepper type cortical circuit (Chen et al., Nature, 2013), the bias is likely weaker than that in the columnar cortical circuit (Wilson et al., Nature Neuroscience, 2016). As the reviewer correctly pointed out, we implemented this difference as the difference in the width of circular convolution. We noted this point in the revised Supporting Discussion (page 2, lines 33-36). We have also revised Extended Data Figure 15a. We removed the part of the figure showing columnar vs. non-columnar appearance of mouse vs. marmoset. We have also revised the figure legend accordingly and indicated that the weight parameter (i.e. width of the circular convolution) is the only difference between mouse and marmoset V1 models (page 39, lines 1110-1120).

We agree with the reviewer that columnar arrangements of neuronal cell bodies do not necessarily define connectivity among those neurons, because sharp integration of inputs is possible in the absence of orientation columns. In the present study, based on the empirical observations mentioned above (Chen et al., Nature, 2013; Wilson et al., Nature Neuroscience,2016), we modeled the effects of the presence and absence of orientation columns by sharp and broad integrations of inputs, respectively. We have clarified this point in the revised manuscript (page 10, lines 258-262). We fully agree with the reviewer that this model "simulates the secondary effect that the presence of orientation columns would have on the synaptic integration width". We have clarified these points in

the revised Supporting Discussion (page 2, lines 36-43). We have also clarified that there are certain limitations in the biophysiological plausibility of the model (page 2, lines 43-47 in Supporting Discussion).

We thank the reviewer for their thoughtful comments.

PCA: I appreciate the extended method section, which now includes more details than in the previous version. Most steps are now clear to me, but I do not understand why the sponta-PC and sponta-only PC analyses, as shown in Extended Data Figure 7, do not have same total number of components. See for example the V2 data: in the sponta-PC case, there is a data set (FOV) that has <200 components, while in the sponta-only PC plot all datasets have >200 components. Since the analyses were performed on the same datasets, how is this possible? The total number of components (i.e., neurons) should be identical, regardless of the type of analysis performed on it.

We thank the reviewer for carefully inspecting the figure and noticing our mistake. Sponta-PC of V2 data shown in Extended Data Figure 7 erroneously included some data from some animals that were excluded from the analysis. As we have described in the original manuscript (page 17, lines 447-449 in the original manuscript; same as page 19, lines 550-552 of the revised manuscript), for two-photon imaging, we selected FOVs (1) with stable recordings of both visual stimulation scan and spontaneous activity scan and (2) with a sufficiently large number of active cells (> 100). FOVs that did not meet these criteria were excluded from the analyses. We have removed these data and corrected Extended Data Figure 7 accordingly. We apologize for the mistake and thank again the reviewer for noticing the mistake.

Method section inconsistencies:
While the methods are improved, they are rather sloppy, as I identified some inconsistencies:
Line 344 says “three Thy1-GCaMP6 mice (4 males; GP 4.3, JAX#024275 (…)”. Is it three or four? It cannot be both.
Line 447 states the FOV is “800 µm × 800 µm”, while the authors’ response to my question about the FOV in the rebuttal was that it was 500 x 500. Again, it cannot both be true.
While these are most likely typos, this does not fill me with confidence that the rest of the method section is reliable, as there are many things that I cannot check and have to take the authors’ word for. I suggest the authors carefully check their method section again to make sure what they write is in fact how they performed the experiments and analyses, and no errors (typographic or otherwise) remain in their manuscript.

We thank the reviewer for noticing our mistakes. We used four Thy1-GCaMP6 mice, and 500 μm × 500 μm FOV. We apologize again for typos and other mistakes in the manuscript. In this revision, we have thoroughly proofread and edited the manuscript to make sure that the methodological procedures we used are correctly stated (changes made were indicated by blue fonts in the revised manuscript). We have also used a professional English editing service to check the manuscript.

Minor comments:

1) I found the following sentence in the abstract rather difficult to parse:
"Stimulus-evoked and spontaneous activities show orthogonal
(dissimilar) patterns in the primary visual cortex (V1) of mice (4-6), which is likely to be
beneficial for separating sensory signals from internally generated noise (5, 7-12);
however, those in V1 of carnivores and primates show highly similar patterns (3, 13-17)."
It might be better to remove the semicolon here. On a related note, I would also remove the reference to carnivores, as the current study is only investigating rodents and primates.

According to the reviewer's comment, we have removed the semicolon and edited the sentences (page 2, lines 29-31). We thank the reviewer for the suggestion. Regarding the reference to carnivores, we would like to keep the reference as is, because the present work is inspired by previous studies in carnivores, particularly the seminal works by Grinvald and colleagues which were done in the cat V1. An important point is that both carnivores and primates have visual cortex with functional columns. In the revised manuscript, we replaced the term "carnivores and primates" with "mammals with functional columns" to clarify this point (page 2, lines 32-35).

2) Line 130: the comparison is missing statistical testing. Based on the figures, I believe the difference in patch width is highly significant, but a t-test's p-value (or something similar) should be added to the manuscript.

We thank the reviewer for giving us a chance to explain why we did not perform statistical testing on these figures. Individual data in these figures are 15 PCs of widefield spontaneous activity obtained from three FOVs (1 FOV per animal). Because we sampled 5 PCs from each FOV, these data points were not independent. This is the reason why we avoided using statistical testing in these figures. We have noted this point in the revised figure legend (page 35, lines 1008-1009). When we applied the binomial test, we found that both the longer axis and shorter axis were significantly larger for V2 patches than V1 patches ($p < 10^{-4}$ for both Extended Data Figure 2b & 2c).

3) Line 137 (and elsewhere): The Bonferroni correction is very conservative, so the authors might want to consider instead using a Bonferroni-Holm correction, which leads to fewer Type II errors. However, if they feel this is not worth the effort, a Bonferroni correction is fine too.

We thank the reviewer for this suggestion. We agree that Bonferroni's correction is conservative. Indeed, we chose to use the Bonferroni correction because we wanted to use conservative statistical criteria. Because key statistical differences were significant using this strict criterion, we prefer to use Bonferroni correction as is.

4) Line 628-629: "PCs that explained less than 3% of variance were discarded."
How many PCs remained for the various data sets, and more importantly, what was the total amount of explained variance of the remaining components? If 50 PCs were used initially, this step likely reduced that number by a considerable amount.

We used this threshold when estimating the dimensionality of the Shared Space (Extended Data Figure 12a). The mean number of remaining PCs and the amount of projected variance explained by the remaining components were as follows: V1: 3.0, 34.1%; V2 3.0, 30.8%; MT, 1.9, 14.6%; Mouse V1: 1.9, 12.1% [Area Name: mean number of remaining shared PCs, mean of total explained variance projected onto the remaining components]. We have clarified this point in the revised manuscript (page 22, lines 632-634) and added these results to the figure legend of Extended Data Figure 12a related to this topic (page 38, lines 1085-1091). We also used the same threshold to select PCs used for projecting the single-trial stimulus-evoked activities to Evoked-Only and Sponta-Only subspaces (Fig. 4c). We imposed this thresholding because full dimensionality of Sponta-Only subspace is equal to the number of neurons in the FOV and hence makes it impossible to divide activity into three subspaces. Because this thresholding necessarily discarded considerable amount of the activity variance, in order to compare relative contribution of the three subspaces across FOVs, we normalized the projected variance by total activity variance accounted for by the remaining components. Mean number of remaining PCs were as follows: V1: 1.1, 2.4; V2 1.6, 2.4; MT, 3.2, 2.1; Mouse V1: 4.7, 2.5 [Area Name: mean number of remaining Evoked-Only PCs, mean number of remaining Sponta-Only PCs]. Because the explained variance shown in Fig. 4c is normalized by the total variance accounted for by all remaining PCs of all spaces, we only listed number of components. We have added these points to the revised manuscript (page 22, lines 634-637 and lines 652-655).

5) Line 640: “shard” should be “shared”.

We have corrected the typo.

6) Extended Data Figure 2 c,d, parentheses on y-axis labels are inverted

We have corrected the mistake.

7) Extended Data Figure 4 could be added to figure 1, but I leave that up to the authors

We appreciate the reviewer’s suggestion. We would like to keep the Extended Data Figure 4 as it is because adding the Extended Data Figure to figure 1 (or figure 2) would make the panels of these figures, which are already tiny, even smaller.

8) Extended Data Figure 5: I would add the median values corresponding to the data in Figure 2g to this comparison, as figure 2g shows cumulative plots but not the median correlation values

According to the reviewer’s suggestion, we have added the median correlation value in Extended Data Figure 5a. In the course of the revision, we noticed that the y axis of Extended Data Figure 5a was the “mean” max correlation value and not the “median”. We corrected the y label in the Extended Data Figure 5a and the corresponding figure legend (page 36, lines 1022-1025). We kept Extended Data Figure 5a as it is because this shows that the result remains the same regardless of the use of mean or median.

9) Extended Data Figure 8-9: a (short) description of the results in the figure legend is missing.

According to the reviewer’s comment, we have added descriptions of the results in the figure legend (page 37, lines 1049-1052 and lines 1055-1057).

10) Extended Data Figure 11: I believe lines 1042-1047 were erroneously copied here, as they do not seem to relate to what is shown in the figure

We have corrected the mistake.

11) Line 1070, “factions” should be “fractions”.

We have corrected the mistake.

12) Extended Data Figure 15: the legend has several language errors, so I suggest the authors have another look at this. E.g.: “Each unit has preferred orientation.”, “are spatially structure by”, etc.

We have corrected typos. We have also used a professional English editing service to check language error.

Reviewer #3 (Remarks on code availability):

The code ran without problems and produced the relevant figures of the model output.

We thank the reviewer for checking the code.

We thank again the reviewer for their thoughtful comments and for thoroughly checking the manuscript.