

Temporal coding carries more stable cortical visual representations than firing rate over time

Corresponding Author: Professor Chong Xie

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

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

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)

The stability of the temporal dynamics of sensory codes has been relatively understudied due to the difficulty of chronically measuring neuronal activity with electrophysiological techniques compared to calcium imaging. Using the newly developed Ultraflexible Nanoelectroic Threads (NETS), Zhu et al measured the activity of 1204 single units across the visual cortices of 5 mice in response to a variety of visual stimuli. Approximately 600 of these units were continuously tracked for 15 days. Like others, they show that conventional stimulus decoding accuracy decreases as a function of time when the decoder is trained only on data from the first day. They then use a variety of analysis techniques to quantify the stability of the temporal dynamics, including LDA embedding trajectory correlation, LDA based stimulus decoding, and UMAP embedding distance. They show that taking temporal dynamics into account increases the stimulus decoding accuracy overall and increases decoding stability. Finally, they quantify network connectivity and find that changes in network connectivity correlate with changes in temporal component based tuning curve correlation but not firing rate based tuning curve correlation.

Main comments

1. Implied causality: in my opinion the wording of the title and throughout the paper implies causality between temporal coding and representational drift. E.g. "Temporal coding stabilizes cortical visual representations". This is unusual phrasing and somewhat misleading language because what is actually being described is the stability of the temporal code, not the effect of temporal coding on representational drift. A more appropriate title might be "Temporal coding is more stable than rate coding in visual cortex". This language throughout the paper should be corrected in a similar way.
2. Unclear throughput of NETs: Because NETs are fairly novel it is important to accurately convey the experimental throughput. In parts it is unclear how many neurons were tracked. E.g. in the introduction "We tracked the same neuronal populations (1204 single units) over the course of 15 consecutive days..." while Fig1 indicates only ~600 neurons were tracked consecutively across 5 mice (so about ~100 neurons per mouse). The number of continuously tracked neurons per mouse should be clearly indicated. Also in the Discussion and Limitations section when-ever referring to the 1204 tracked single units. It should also be acknowledged in the limitations section that calcium imaging still offers substantially higher throughput in terms of chronically trackable neurons per mouse.
3. Xia et al., 2021 is not cited: Xia et al., 2021 (10.1038/s41467-021-25437-2) look at temporal dynamics in response to movie presentations, and needs to be cited in introduction and discussed in the discussion. Although they use calcium imaging and therefore look at a different timescale of temporal codes the paper is too close in topic to not discuss. For example, they also find that drift is higher in "non-coding" dimensions which should also be mentioned with respect to point 4 in paragraph 10 of the discussion.
4. PSTH correlation is missing: I think there is a step missing when going from rate code based tuning curve correlation to LDA component based tuning curve correlation. I think it would make sense to also do peristimulus time histogram (PSTH) correlation (primarily in Fig4) as an intermediate step between the two approaches i.e. for a given neuron take the average

PSTH for each stimulus and correlated across days, then take the average of this stimulus wise PSTH correlation across stimuli to get a temporal code correlation which is invariant to rate based tuning curve correlation. This would give an idea of the temporal code stability without the denoising effect of LDA.

5. Network connectivity and temporal dynamics: Fig6j-k shows that the correlation between network connectivity and tuning curve correlation is higher when based on temporal components compared to rate code. However, it is not clear that this is a time dependent effect as only one interval length (7 days) is shown. To make the argument that stability of temporal codes is associated with network connectivity, the authors should repeated the analysis for multiple time intervals and plot the correlation as a function of time interval.

Minor points and suggestions

1. Shuffle in Fig1j is unclear. I think the description is inconsistent between legend and results. It should be clear from the legend that this is not a complete shuffle of neurons but a progressive 10% per day shuffle.
2. It is not clear to me why the difference between temporal code stability and rate code stability is larger for neurons with higher response reliability. Is it because neurons with large amplitude variability across repeats still have very stable temporal profiles? Maybe doing additional analysis comparing rate, component and PSTH correlation comparisons across units would help here.
3. Tuning reliability inherently affects the measured drift magnitude. But this is not “true” drift. Response reliability is rather something that needs to be accounted for than (e.g. normalization Marks et al., 2021 or by focusing on statistically significant changes Bauer et al., 2024). I think this is not put into sufficient context and is wrongly framed as an effector rather than a confounder.
4. I am not sure if the indication of time is accurate or approximate in Fig3b-c. I would suggest presenting these plots with multiple alternative color schemes in an extended figure, e.g. Color coded by time, day, ori, dir, sf. A video/GIF in which these plots are rotating might also be helpful for visualization.
5. Why did the authors use distance metrics in UMAP embeddings space in Fig 5. While UMAP is useful for visualization in 2 dimensions, it should generally not be used as a quantitative embedding space. This is because the distance between clusters in UMAP embedding space is not meaningful.
6. In Fig3f-g the term “50% max separation time” is unclear in my opinion. Maybe “feature separation time” would be more easily understandable.
7. In the field of representational drift it is largely up to the authors if they frame their data as showing more drift or less drift than expected. Considering most analysis does show some time dependent changes I think the message is leaning a bit too heavily towards there is very little drift. I think the wording could be more balanced in the introduction and discussion.
8. I find it strange that the legends contain no n numbers. They are listed in a supplementary table, but it is quite tedious to refer to a separate supplementary document for n numbers, units and statistics.
9. Pooling neurons across animals assumes uncoordinated drift within the population. I think it is fine to assume this but it needs to be stated and discussed. E.g. in Fig3 and 5.
10. Reference 8 is a thesis, but the corresponding paper is peer reviewed, and should be cited in addition or instead.
11. I'm not sure it makes sense to have Fig 3 (population level analysis) before fig4 (single cell analysis), when Fig 5 again looks at population analysis. It might make the paper more readable to switch Fig 3 and 4, or to have Fig 5 as a supplement.
12. It would help to summarize the results in extended Fig8-9 using correlation heat maps.
13. A recent paper by Bauer et al., (10.1038/s41467-024-53326-x) models the effect of spontaneous and experience dependent connectivity changes on rate code stability in mouse V1. This paper should be mentioned and discussed in relation to the findings in Fig6.
14. In future, it would help to have line numbers to make it easier to refer to specific sentences during the review process.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The paper by Zhu, He et al. addresses factors determining stability and representational drift in mouse visual cortex. Representational drift is a topic that is currently intensely discussed in the field, with recent publications differing quite strongly in the degree of stability found, depending on the brain area studied and the methods employed. The authors take a fresh look at this matter, leveraging a very recently introduced novel type of highly flexible, multi-site electrical recording probe (“NET”). These electrodes allow for recording multiple single units over weeks, with many individual neurons followed over time, as the authors convincingly show. When simply using average spike rate as a readout, as all previous studies on drift have done, they observe a certain degree of drift, depending on the visual stimulus used, in line with existing data. Crucially, they found drift to be substantially lower, when, instead of spike rate, using the temporal dynamics of the response in the range of tens of milliseconds. Taking response timing into account also improved stimulus discriminability and decoding accuracy. Finally, they show that the stability of temporal codes, as compared to a rate code, is better linked with functional connectivity, as assessed by single spike based cross-correlations.

Overall, these are very exciting results, which provide crucial insight into factors underlying the stability of neuronal representations in the neocortex. The study is carried out at very high technical standards and employing elaborated data analysis tools. I would like to point out that the longitudinal cross-correlation data (Fig. 6) alone are extremely rich, adding

crucially to our understanding of how connectivity in the visual cortex changes over time.

The paper is very nicely written, with very instructive figures. I only have a few, mostly minor comments.

1. The recent Nature Comm paper by Bauer et al. on drift in visual cortex should be cited
2. Check legend for Fig. 2c
3. Line 176: Point to respective Figure panel
4. Line 182: I don't see any indication of a discrete group of putative anchor cells
5. Figure 3a: How were these example neurons selected? Would the plot look similar for randomly selected neurons?
6. Regarding the temporally resolved responses: Was stimulus grating spatial phase randomized?
7. Out of interest: How long can individual cells be reliably tracked using their electrodes?
8. Line 560: please check this sentence, not sure it is clear like this
9. Line 570: should read 'drifting'
10. Ext. Fig. 1b: it would be nice to see such data from an electrode that actually did move
11. Ext. Fig. 1g: p values are not shown in the Figure, but are referred to in the legend
12. Line 814: 'with' missing
13. References: several of these are incomplete

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

In this manuscript, the authors record neuronal responses in areas V1 and LM of the mouse for several days (up to 15) and test the stability of their tuning for various stimulus categories at both the single unit and population level. They found that, in general, the tuning is quite stable across days, although to a different degree for different units. Interestingly, the extent to which a neuron is able to maintain its tuning for visual stimuli over time correlates with its trial-by-trial reproducibility. Additionally, they show that, taking into account the dynamics of the response during stimulus presentation, as opposed to just computing the overall firing rate, provides more discriminatory information about the visual stimuli and such information is more stable across days of recordings.

Overall, the focus of the study and the conclusions reached by the authors are interesting and relevant. Not many studies have been able to track neuronal tuning across days. As such the results about the day-by-day stability of the visual representation are novel and important for the community, especially the relationship with the trial-by-trial reproducibility of the response. The results concerning the additional information provided by the dynamics of the responses (what the authors refer to as temporal code) are less surprising (by definition, a higher dimensional representation contains richer information than a lower dimensional one). Still, it is interesting how the authors investigate the various temporal profiles of the responses and the way they allow a better discrimination of the visual stimuli.

In summary, in terms of content, I think that the study is valuable and does provides some new knowledge to the vision science community. The problem, however, is the way the manuscript is written, the analyses are described, and the results are presented. Sentences are often very unclear and convoluted or syntactically wrong. The details of the analyses are often neglected. In the Results section, the single unit and population metrics used to assess stimulus discriminability and neuronal tuning are almost never explained intuitively. In the Methods section, if such metrics are explained, the text is often so short and confusing to be of little help (equations are rarely used). Same goes for the legends. This is a problem, especially for those analyses that are not very conventional and because the figures are extremely rich and complex (with tens of subplots and panels) and the authors jump from one analysis to the next without providing much explanation. This is a serious problem because, in many instances, I could only marginally grasp the meaning and the solidity of the results presented by the authors.

To sum up, the manuscript, as it is currently written, it is not suitable for publication, which is a pity, given that, as I wrote above, several of the findings presented here seem interesting. So, I urge the authors to carry out a deep revision of the way the write, present, explain and discuss their analyses and results. Below, I provided some more detailed feedback.

MAJOR CONCERNS

1. Clarity of the writing. As explained above, the lack of clarity in the writing is so pervasive that it is hard to point a few specific instances. Still, I provide some comments below for what concerns the title, abstract and introduction.

Title.

The title, "Temporal coding stabilizes visual cortical representations", lacks clarity and impact for two main reasons. First, the expression "temporal coding" is quite vague: it does not refer to a specific attribute of neural responses, but rather encompasses a broad set of features related to the temporal dimension of neural coding. For instance, (Panzeri et al., 2010) distinguish between "latency code", "interspike intervals", and "phase of firing", each pinpointing to different aspects of the temporal dynamics of neural activity. The fact that the title does not specify which exact dimension of temporal coding is being discussed introduces ambiguity and potentially undermines the paper's accessibility and comprehensibility. The

second issue concerns the use of the term “stabilizes”, which is somewhat ambiguous without further specification. Does it refer to stabilization over time or in space? Since the paper focuses on temporal stabilization, adding “over time” to the end of the title would have helped clarify the intended meaning. Also, stabilization over time can simply mean over the course of a few hundreds of ms or seconds - see for instance (Piasini et al., 2021; Rudelt et al., 2024). What is peculiar of this work is that stability is assessed across days. So it would be worth stating this upfront in the title.

Abstract.

The abstract assumes some specific prior knowledge and is somewhat convoluted. Expressions such as “representational drift” are quite specialized and may not be universally recognized. Expressions such as “temporal coding” on the other hand are too vague and broad (see the criticism above). It would be better if the authors simply tried to explain what they did rather than use jargon of this kind, which may be either too specialized or too vague for many readers. Additionally, some sentences are cryptic, such as: “Across various types of stimuli, we found that neurons exhibited varying degrees of day-to-day stability in their slower, second time-scale rate code tuning, with this stability being explained by the reliability of their tuning”. The meaning of such sentences can only be grasped after reading them several times.

Introduction.

The Introduction fails to provide the necessary background for the reader to clearly understand the research question posed by the authors, its relevance within the field, and the logical progression that led to the experimental design. The concepts underpinning the entire work, such as rate coding and temporal coding, are never precisely defined, nor are they clearly contrasted as differential aspects of neural coding. Furthermore, this section does not outline the predictions regarding potential results and their interpretations, making it challenging for the reader to follow the underlying rationale of the study. Moreover, the summary of the results is way too dense, squeezed in poorly readable sentences such as: “Neurons showed different day-to-day rate code tuning stability, with the degree of stability best explained by tuning reliability” (lines 61-62), and “Accordingly, adding temporal dynamic information stably improved the longitudinal causal decoding accuracy. Lastly, we found that the reliability of temporal code was associated with static and dynamic properties of network functional connectivity.” (lines 67-69).

Several syntactical imprecisions in sentence construction undermine readability. For example, the sentence “These snapshot studies, however, have not elucidated whether the temporal dynamics are consistent chronically or their role in supporting the stability of visual representations.” (lines 49-51) is particularly unclear.

Discussion.

This section of the manuscript also contains sentences whose meaning is difficult to grasp. Examples include:

Lines 568-570: “A breakdown characterization for each individual component (Supplementary Fig. 8a) showed that the stabilization effect was not due to individual components themselves drifted less than firing rates.”. This sentence is grammatically awkward and unclear, making it challenging to understand the intended point about the stabilization effect and the individual components.

Lines 571-574: “We showed that (Supplementary Fig. 8c) the reduction in representational drift (Figure 4f) offered by temporal code was not trivially due to temporal codes increasing reliability (Supplementary Fig. 4) and hence increasing the denominator (CC_{ws}) of the representational drift index.”. The explanation here is highly technical and convoluted, with key elements left unexplained. The connection between temporal code reliability and the denominator of the representational drift index requires clearer elaboration.

Lines 588-589: “Therefore, network functional connectivity metrics better explained temporal components based single neuron tuning stability than that of firing rate based.”. This sentence is syntactically flawed and lacks clarity, making it hard to discern how functional connectivity metrics relate to the stability of single-neuron tuning.

These examples reflect the broader need for improved clarity and precision throughout the Discussion section.

2. Lines 119-121. “We used a published method, an actively-learned ensemble of hierarchical clustering trees to match neurons ...”. Too little info is provided about how neurons were matched across sessions. Fine to refer to a published method but some qualitative explanation of how the method works should be provided. Especially if a figure is shown with the outcome of the matching procedure. The figure by itself is very poorly explained. The legend is basically useless. The authors should explain what they are plotting and why in a way that is detailed enough to be useful and understandable.

3. Lines 170-171. “The tuning similarity across all four stimulus types ...”. What similarity metric was used here? A Pearson corr. coeff? If so, it should be stated somewhere more clearly. Also, in the histograms, a given cell how many data points contributed? ?? It should be explained. Also, in panels of Fig. 2 >=b the tuning similarity metric seems different from a, it is not bounded anymore between 0 and 1, so it cannot be a simple Pearson corr coeff. What is it then? This is a typical example of how poorly the metrics and the analyses are explained.

4. Lines 224-226. “We fitted a decoder with a 90 ms window length at 160 ms after stimulus onset (from 70 ms to 160 ms) to ensure that different stimulus categories were readily decodable.” The writing of the manuscript is in general quite poor, sentences are too long and convoluted. This makes it hard to understand some of the analyses. This is an example. I guess what the authors mean is that they built a linear decoder based on a rate codes, i.e., firing rates computed between 70 and 160 ms. If so the sentence “We fitted a decoder with a 90 ms window length at 160 ms after stimulus onset” is very misleading, because it seems that the integration window started at 160 ms and not at 70.

Leaving this aside, it is not so clear what happened next. The authors somehow project the pop response vectors (measured over time in small 10 ms time bins) over a space obtained by the LDA. But how is this space defined? LDA should return a single weight vector that defines a direction in the representational space. If data points are projected along this vector/direction, they are maximally discriminable. But this is a single direction (a single dimension), while the authors seem to extract multiple directions/axes. From the main text, it is impossible to understand how these further directions were extracted. The legend, the methods and the figure itself provide some hints. There appear to be 3 dimensions here. But while I can understand how the authors ended up having two separate axes for orientation and direction or for orientation and SF (they could have built separate LDAs to discriminate orientation and direction or orientation and SF), I really don't understand how they can get two different axes for the same property (i.e., two axes for orientation). They need to explain this. It looks like they did something equivalent to a PCA but using LDA, but how this was achieved is really unclear. I would suggest using some equations in the Methods section to explain it in details and then provide an intuitive explanation in the Results, perhaps by adding a figure with a cartoon to illustrate the procedure.

5. Lines 265-266. "To effectively extract temporal codes, we utilized cross-validated, regularized linear discriminant analysis (LDA) to determine the temporal components that best separated the different stimulus conditions." Again, this section is not understandable because it is not explained how this LDA based decomposition works.

6. Lines 288-290. "Across all stimuli, unit tuning curves were more similar to themselves after 7 days (Fig. 4e) when derived jointly from temporal components 1-3 than those derived from firing rates". Again, it is not explained what the similarity metrics are on the axes here, whether they are bounded, etc.

7. Lines 392-394. "adding temporal components in addition to firing rate in the fixed decoder consistently reduced decoding error for all four stimuli across 8 future testing days when trained on the first 7 days ($P < 0.0089$, Extended Fig6e ...". It is very unclear what this figure shows. The y axis labels do not help. They also seem inconsistent across classes of stimuli. For some classification, apparently a decoding performance is reported, while for others, an error rate is reported. In the case of the accuracy, it is very odd that accuracy is higher for NI classification than for SG. So, I guess that, whatever this accuracy is, it is not the percentage of correct choices of the classifier. So, I wonder why the authors did make it so hard to understand this decoding analysis and so difficult to compare it across stimulus sets. Cannot they simply report (as customary in this type of analyses) the classification accuracy of the classifier (with both firing rates and temporal components) as the fraction (or percentage) of correct choices? Putting a dashed line indicating the chance level? And also reporting the performance of the classifier with shuffled labels as a reference? I urge them to do so, otherwise the reader cannot really understand and judge the solidity of their claims.

8. Fig. 6b. These plots are esthetically nice but hard to parse, if one has to judge the stability of the strength of the functional correlations across days. Also, they do not lean themselves to an easy way to quantify how stable these patterns of functional correlations are. The quantification done in Fig. 6c and d is very rough. Simply checking that the number of significantly correlated pairs is stable or that the magnitude of the correlation is stable does not quantify if it is the same pairs that remain connected over time. So here is my suggestion.

Instead of the plots shown in Fig. 6a,b, the authors should simply report a squared correlation matrix, where the intensity of the color codes the intensity of the correlation between each pair of units. This matrix can be made across days, making it easier to look for similarities of the connection patterns over time. More importantly, the authors can then compute the correlation between these correlation matrixes as a measure of the stability of the connection patterns across days. E.g., they can compute the correlation between day 6 and 7, day 6 and 8, day 6 and 9, etc, and produce a line plot like the one in c, d.

9. Fig. 6f. It is hard to understand what is being plotted here. What does "attendance" mean exactly in the figure? How can the range between the first and last seen session also be expressed as a percentage/fraction? More importantly, it is not clear if all these analyses refer to pairs that were first detected on day 1, as it should be done, or if neurons first detected in later days were also included. Fig. 6g is equally hard to understand. What is showing exactly. The fraction of pairs lost in a given day, compared to the previous day? Or compared to the first day? This should be stated clearly.

10. The authors wrote that the electrodes were implanted in two different visual areas, V1 and LM, which are known to be in a hierarchical relationship – e.g. see (Wang and Burkhalter, 2007; Wang et al., 2012). However, these areas are never directly compared in any of the reported results. This raises the question of why both areas were considered in the first place, as why their respective roles and differences are not addressed or discussed.

MINOR CONCERNS

1. legend of Fig. 1j. "Neuron identity shuffled decoder approached chance ...". This sentence is misleading because it seems that all neurons have been shuffled, while, based on the main text, only 10% per session were shuffled. Please specify here that only 10% of cells were shuffled.

2. Lines 143-144. "... we observed visually consistent day-to-day most preferred. Conversely, we also detected longitudinal drifts in units' tuning to less preferred images (Fig. 1i, zoomed-in view)". This does not seem a fair description of what shown in the zoomed-in views. It seems to me that what is changing over time in several neurons is the response to the originally preferred stimulus, which ceases to be such after 7 days. This does not happen for all the units, but some of them seem to actually loose their tuning.

3. Lines 151-152. "We observed similar decoding error indexed stability and drift for all other stimulus types (Supplementary Fig2)." Not clear which stimuli the authors are referring to here.

4. Lines 199-201. "Neurons exhibited diverse temporal response profiles to an example stimulus, showing 0° drifting grating evoked daily averaged, smoothed firing time courses from example neurons pooled from 5 mice over 7 different experiment days." The writing of this sentence is truly terrible! The result is that it is hard to understand what is it that the figure actually shows. Is it the response to a single, specific stimulus (the 0° DG?) of a population of neurons across consecutive days?

5. Lines 234-236. "Similarly, the four spatial frequencies of static gratings stimuli were established in four parallel planes, where the six orientations evolved radially outward forming a hexagonal formation (Fig. 3c, Supplementary Fig3b, Supplementary Movie 2)." The figure does not really allow to distinguish easily the traces referring to different SFs. The dashed lines connecting the traces to the images of the gratings do not help much. The authors could perhaps try to encircle those traces referring to the same SF and connecting the circle to the image of the grating with the corresponding SF.

6. Lines 296-297. Another very convoluted and odd sentence "Consistent with the finding in firing rate-based tuning (Fig. 2b), stability in temporal components defined tuning was also associated with component tuning reliability, strength, and selectivity." This is very hard to interpret. Please revise it.

7. Lines 310-312. "Components increased stimuli separation than firing rates, showing an example neuron's daily trial-averaged firing rates (Left) and first 3 components (Right) distribution (n=15 days) under 16 different stimuli." The writing here is terrible ... please revise it making sure a subject precedes a verb and then a complement or whatever follows. But please make things clearer.

8. Lines 316-318. "Across all stimuli, temporal components-based unit tuning curves were more similar to themselves after 7 days than that of firing rate-based. Paired t- test." Be more specific about what is plotted. Do the dots refer to the neurons in the population? I guess so, but it should be explicitly reported along with the number of units being shown.

REFERENCES CITED IN THE REVIEW

Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33:111–120.

Piasini E, Soltuzu L, Muratore P, Caramellino R, Vinken K, Op de Beeck H, Balasubramanian V, Zoccolan D (2021) Temporal stability of stimulus representation increases along rodent visual cortical hierarchies. *Nat Commun* 12:4448.

Rudelt L, Marx DG, Spitzner FP, Cramer B, Zierenberg J, Priesemann V (2024) Signatures of hierarchical temporal processing in the mouse visual system. *PLOS Comput Biol* 20:e1012355.

Wang Q, Burkhalter A (2007) Area map of mouse visual cortex. *J Comp Neurol* 502:339–357.

Wang Q, Sporns O, Burkhalter A (2012) Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *J Neurosci* 32:4386–4399.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

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

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have done a great job addressing all the points raised.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have addressed all my comments.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

I am satisfied with the revision of the manuscript and I thank the authors for having added many of the descriptions and clarifications I felt were missing in the previous version of their manuscript.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

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

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

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

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license,

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

The stability of the temporal dynamics of sensory codes has been relatively understudied due to the difficulty of chronically measuring neuronal activity with electrophysiological techniques compared to calcium imaging. Using the newly developed Ultraflexible Nanoelectroic Threads (NETS), Zhu et al measured the activity of 1204 single units across the visual cortices of 5 mice in response to a variety of visual stimuli. Approximately 600 of these units were continuously tracked for 15 days. Like others, they show that conventional stimulus decoding accuracy decreases as a function of time when the decoder is trained only on data from the first day. They then use a variety of analysis techniques to quantify the stability of the temporal dynamics, including LDA embedding trajectory correlation, LDA based stimulus decoding, and UMAP embedding distance. They show that taking temporal dynamics into account increases the stimulus decoding accuracy overall and increases decoding stability. Finally, they quantify network connectivity and find that changes in network connectivity correlate with changes in temporal component based tuning curve correlation but not firing rate based tuning curve correlation.

Main comments

1. Implied causality: in my opinion the wording of the title and throughout the paper implies causality between temporal coding and representational drift. E.g. "Temporal coding stabilizes cortical visual representations". This is unusual phrasing and somewhat misleading language because what is actually being described is the stability of the temporal code, not the effect of temporal coding on representational drift. A more appropriate title might be "Temporal coding is more stable than rate coding in visual cortex". This language throughout the paper should be corrected in a similar way.

A: We have now changed all mentions of "stabilize" into "temporal coding carries more stable visual representations than rate coding" and alike. The title has also been changed to "Temporal coding carries more stable cortical visual representations than firing rate over time". The phrase "over time" was suggested by Reviewer 3.

We have changed all mentions of "stabilization". E.g. in discussion:

"The mechanisms underlying temporal codes' stabilization effect" has been changed to "The mechanisms underlying temporal codes' stability effect" has been changed to "The mechanisms underlying the enhanced stability of temporal codes over rate codes".

2. Unclear throughput of NETs: Because NETs are fairly novel it is important to accurately convey the experimental throughput. In parts it is unclear how many neurons were tracked. E.g. in the introduction "We tracked the same neuronal populations (1204 single units) over the course of 15 consecutive days..." while Fig1 indicates only ~600 neurons were tracked consecutively across 5 mice (so about ~100 neurons per mouse). The number of continuously tracked neurons per mouse should be clearly indicated. Also in the Discussion and Limitations section when-ever referring to the 1204 tracked single units. It should also be acknowledged in the limitations section that calcium imaging still offers substantially higher throughput in terms of chronically trackable neurons per mouse.

A: We have changed this description into "... 1204 tracked neurons (Fig. 1d), they appeared for an average of 10.96 ± 0.15 days (mean + s.e.m.). Among which 118.80 ± 26.97 neurons per mouse showed up on all 15 days (mean $\pm$ s.e.m.).". We removed "1204 single units" in the introduction. We have acknowledged in the limitations section that calcium imaging still offers substantially higher throughput in terms of chronically trackable neurons per mouse

"... We also acknowledge limitations common to all electrical recordings, e.g., only neurons/cell types that fire sufficiently frequently are reliably sortable[41]. In addition, Calcium imaging still offers substantially higher throughput in terms of chronically trackable neurons per mouse."

3. Xia et al., 2021 is not cited: Xia et al., 2021 (10.1038/s41467-021-25437-2) look at temporal dynamics in response to movie presentations, and needs to be cited in introduction and discussed in the discussion. Although they use calcium imaging and therefore look at a different timescale of temporal codes the paper is too close in topic to not discuss. For example, they also find that drift is higher in "non-coding" dimensions which should also be mentioned with respect to point 4 in paragraph 10 of the discussion.

A: In the introduction, we modified the following (in underline) “...naturalistic stimuli evoke responses that drift considerably over the course of minutes to weeks[10], occurring in a large number of visual sub-regions and neuron types[10, 11]. In recent literature[ref], it is observed that while responses to 30 seconds-long natural movie drifted across weeks at the level of individual neurons, a populational neural code denoised by unsupervised machine learning provided a more stable representation of time within the movie clip. However, it remains unclear what neural mechanisms generally anchor consistent visual perception at sub-second timescale and are relevant to both individual neurons and neural populations.”

In the concerned paragraph of the discussion, we modified the following (in underline) “...Yet, such potential non-coding excessive activation was only partly projected to the 1st component while minimally affecting the 2nd and 3rd components, thus still preserving sufficient stability. Similarly, it is observed[ref] that drift is higher in “non-coding” dimensions, despite visual responses being studied on a different timescale and using a different neural code extraction mechanism. In the same vein, temporal codes were less explained by non-stimulus variables (Supplementary Fig8e). These finding might also be consistent with a recent report[46] that fast spiking dynamics could de-mix responses to multiple sensory inputs.”

[ref] Xia, J., Marks, T.D., Goard, M.J. et al. Stable representation of a naturalistic movie emerges from episodic activity with gain variability. Nat Commun 12, 5170 (2021). <https://doi.org/10.1038/s41467-021-25437-2>

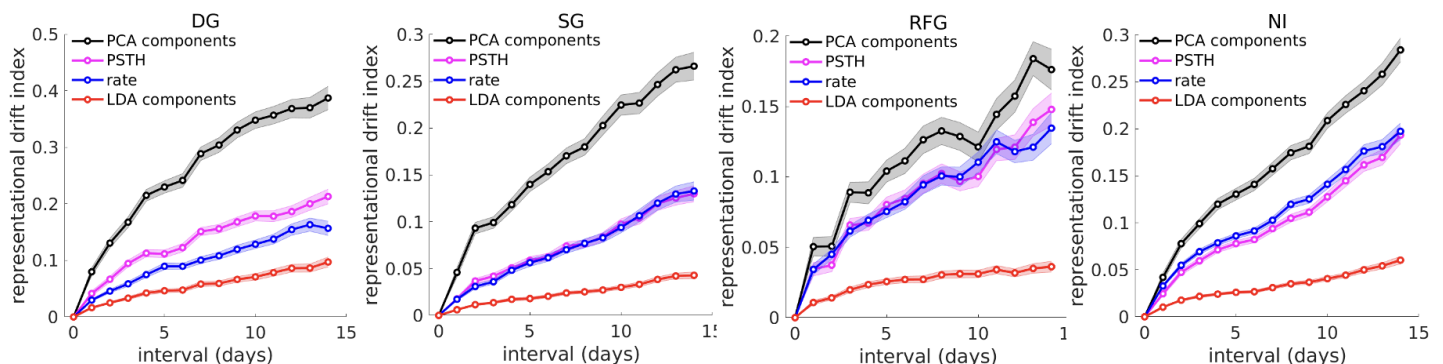
4. PSTH correlation is missing: I think there is a step missing when going from rate code based tuning curve correlation to LDA component based tuning curve correlation. I think it would make sense to also do peristimulus time histogram (PSTH) correlation (primarily in Fig4) as an intermediate step between the two approaches i.e. for a given neuron take the average PSTH for each stimulus and correlated across days, then take the average of this stimulus wise PSTH correlation across stimuli to get a temporal code correlation which is invariant to rate based tuning curve correlation. This would give an idea of the temporal code stability without the denoising effect of LDA.

A: Thanks for the suggestion to make our argument more complete, we have now adapted an algorithm similar to those used in Kim et al., (their Extended Data Fig. 3) to establish PSTH based tuning curves (right before our LDA step) across days and derived the corresponding representational drift index. We have overlayed its trajectory across days in our Ext.Fig.4d. This showed that temporal dynamics by itself was not necessarily more stable than rate code. This further implies that LDA’s ability to find stimulus-specific subspaces from the temporal dynamics contributed to the reduced drift.

We modified the results section in the main text:

“...Importantly, merely increasing the dimension of firing rate-based tuning did not lead to decreased drift across time (Extended Fig4d). This is further strengthened by the fact that top three components extracted from a simpler method (PCA) failed to reduce drift (Extended Fig4d). Finally, high dimensional PSTH (before LDA or PCA dimension reduction) failed to meaningfully reduce drift (Extended Fig4d).”

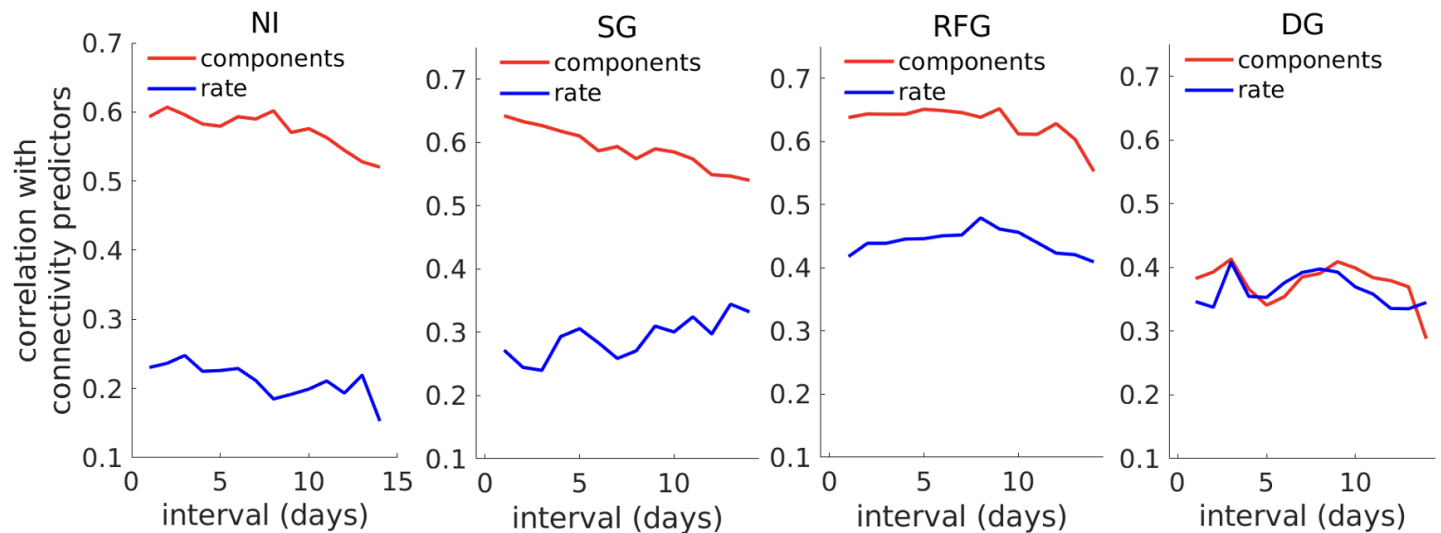
Kim, JH., Daie, K. & Li, N. A combinatorial neural code for long-term motor memory. Nature 637, 663–672 (2025). <https://doi.org/10.1038/s41586-024-08193-3>



Legend: Merely increasing the dimension of the tuning curve did not necessarily lead to reduced representational drift across time for all four stimulus types (DG: drifting gratings; SG: static gratings; RFG: receptive field Gabors; NI: natural images), as exemplified by 1. Using top 3 PCA components extracted from the PSTH or the PSTH themselves to form the tuning curves did not meaningfully reduced drift than rate coding based tuning curves.

5. Network connectivity and temporal dynamics: Fig6j-k shows that the correlation between network connectivity and tuning curve correlation is higher when based on temporal components compared to rate code. However, it is not clear that this is a time dependent effect as only one interval length (7 days) is shown. To make the argument that stability of temporal codes is associated with network connectivity, the authors should repeated the analysis for multiple time intervals and plot the correlation as a function of time interval.

A: Thanks for the suggestion to make our argument more solid. We have now extended the same analysis in Fig6j-k to all 14 day-lag intervals. The new Extended Fig8b figure panel below shows that for 3 out of the 4 stimuli, correlation between network connectivity and tuning curve correlation is higher when based on temporal components compared to rate code. The effect held true for the intervals tested. Please also note that Fig6j-k and corresponding figures for the other 3 stimuli has been slightly if visibly changed since we modified the calculation for 4 independent variables ($\Delta \text{popCoup}$, ΔstPA , $\Delta \text{CCG \#}$, $\Delta \text{CCG Amp}$) so that they are now calculated from the corresponding day interval to adapt to the new analysis. Previously, they were derived from averaging the calculation from all intervals.



Legend: The plot shows correlation coefficient (square root of R^2 of linear regression) between tuning curve similarity and assorted connectivity metrics over all time intervals. For 3 out of the 4 stimuli, correlation between network connectivity and tuning curve correlation is higher when based on temporal components compared to rate code.

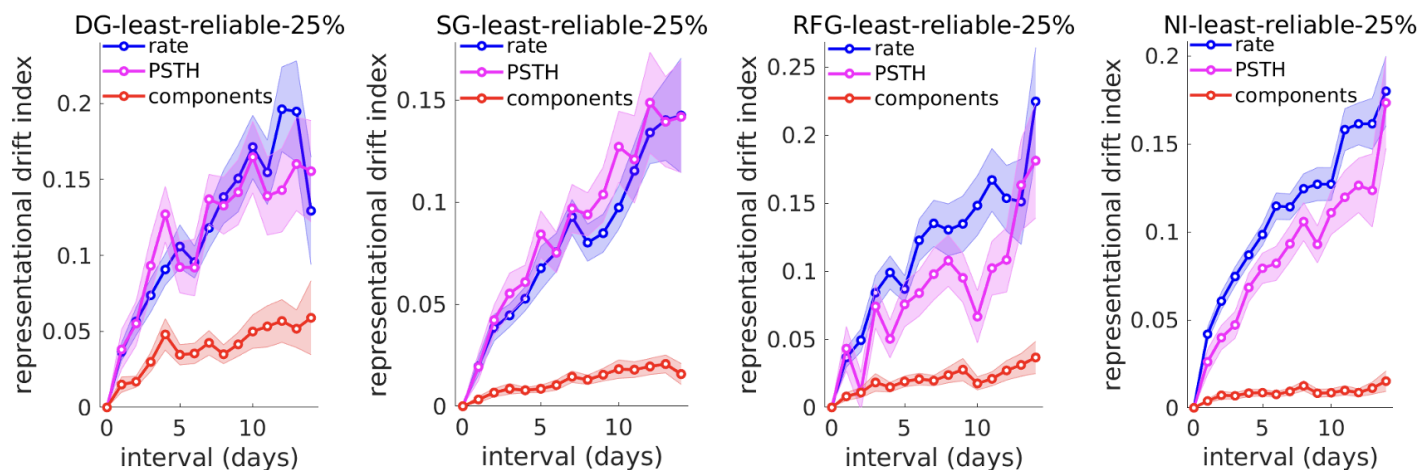
Minor points and suggestions

1. Shuffle in Fig1j is unclear. I think the description is inconsistent between legend and results. It should be clear from the legend that this is not a complete shuffle of neurons but a progressive 10% per day shuffle.

A: Thanks. We have changed the figure legend, noting that "10% of cells were shuffled per session, resulting an expected $(0.9)^{\text{day}}$ fraction of unshuffled neurons at any given day." to be consistent with the result.

2. It is not clear to me why the difference between temporal code stability and rate code stability is larger for neurons with higher response reliability. Is it because neurons with large amplitude variability across repeats still have very stable temporal profiles? Maybe doing additional analysis comparing rate, component and PSTH correlation comparisons across units would help here.

A: We assume the reviewer meant "difference between temporal code stability and rate code stability is larger for neurons with **lower** response reliability", since that was what we reported in Fig4g. We plotted below PSTH correlation based representational drift index for the least reliable 25% of the neurons as in Fig4g. Per the reviewer's hypothesis, it may be argued the temporal profiles for the least reliable neurons were more stable than rate coding for RFG and NI stimuli but not for DG and SG stimuli. Given these results, we speculate that LDA's denoising effect is more prominent for those least reliable neurons.



Reviewer Figure Legend: Same as in Fig4g but with the addition of PSTH based tuning curve representational drift index for the least reliable 25% of the neurons.

3. Tuning reliability inherently affects the measured drift magnitude. But this is not “true” drift. Response reliability is rather something that needs to be accounted for than (e.g. normalization Marks et al., 2021 or by focusing on statistically significant changes Bauer et al., 2024). I think this is not put into sufficient context and is wrongly framed as an effector rather than a confounder.

A: Thanks for the comment, we have now modified our statement in reliability, suggesting it as an indicator of tuning curve similarity across days without emphasizing it as an effector. It might be conceivable that tuning reliability inherently affects drift magnitude, yet past studies did not quantitatively characterize this relationship, E.g. Xia et al., which the reviewer cited simply reported within-session and across-session stability without explicitly studying their relationship. Similarly, while Marks et al., as the reviewer pointed out, implied that reliability inherently affects stability through normalization. However, the normalization was performed by dividing the sum of reliability and stability, not simply dividing by reliability. Thus, the exact reliability – stability relationship was also unclear from their report. We thus find it of some value to at least report this relationship. This was also echoed in reviewer 3’s comment “Interestingly, the extent to which a neuron is able to maintain its tuning for visual stimuli over time correlates with its trial-by-trial reproducibility.” Below we detailed changes from abstract to discussion on how claims regarding reliability have been modified.

Abstract: “... The across day stability correlated with tuning reliability.”

Introduction: “Neurons displayed varying levels of day-to-day stability in their rate code tuning. This stability correlated with their tuning reliability.”

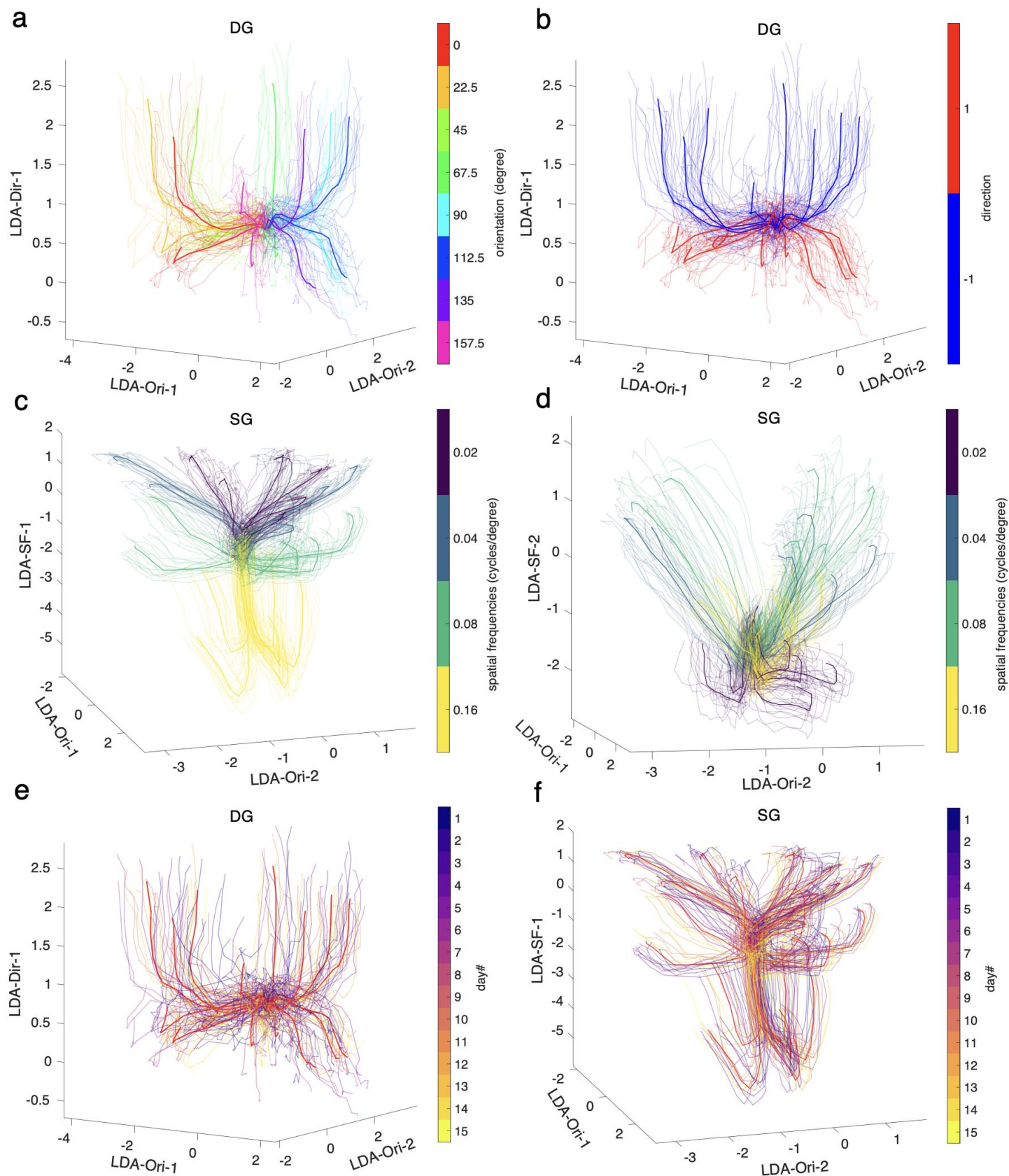
Result: “~~Critically, when considering all these contributors together, the correlation coefficient marginally (change in $r < 0.03$) or non-significantly increased from that of the tuning reliability across the four stimuli (Fig. 2c, e, Extended Fig3b, d), indicating that the tuning reliability was the dominant predictor of the tuning stability. Finally, these contributors could also jointly explain tuning curve similarity across days (Fig. 2c, e, Extended Fig3b, d).~~”

Discussion: “... with across-session stability strongly associated with tuning reliability. ~~The degree of this association appeared to be stronger than that observed in past imaging studies[9, 10], which might be due to differences in the experimental procedures used.~~ In line with reports that tuning reliability may underlie the functional specialization of different brain regions[73] and functional evolution of developmental stages[ref], our results further showed that response reliability could be an indicator of chronic rate code representation stability (Fig2).”

[ref]: Trägenap, S., Whitney, D.E., Fitzpatrick, D. et al. The developmental emergence of reliable cortical representations. *Nat Neurosci* **28**, 394–405 (2025). <https://doi.org/10.1038/s41593-024-01857-3>

4. I am not sure if the indication of time is accurate or approximate in Fig3b-c. I would suggest presenting these plots with multiple alternative color schemes in an extended figure, e.g. Color coded by time, day, ori, dir, sf. A video/GIF in which these plots are rotating might also be helpful for visualization.

A: Thanks for the suggestion, we added new plots to our supplementary figures which color-code Fig3b-c by orientation, direction, spatial frequency (LDA1 and LDA2) , and by day. We previously included one Supplementary Figure 3 to plot the detailed temporal evolution of Fig3b-c, thus here we did not include a plot that is color-coded by time. Corresponding supplementary videos1-7 that show the temporal evolution and rotation of these newly added plots have also been added.



Legend. Supplementary Fig3: Additional visualization on temporally resolved populational representation: **a-b**. Disentangled population representation dynamics during drifting gratings from an example animal. The plot shows daily trial-averaged (lighter lines) and 15-day-averaged (darker lines) traces, projected onto the first two linear discriminant orientation coding dimensions and the first direction coding axis. Orientations (a) or directions (b) were used to color-code the traces. **c-d**. Same as in (a) but for static grating stimuli, projected onto the first two orientation coding dimensions and the first (c) or second (d) spatial frequency coding axis. Stimulus spatial frequencies were coded in color. **e-f**. same as in (a) and (c) but color-coded by day number.

5. Why did the authors use distance metrics in UMAP embeddings space in Fig 5. While UMAP is useful for visualization in 2 dimensions, it should generally not be used as a quantitative embedding space. This is because the distance between clusters in UMAP embedding space is not meaningful.

A: We have added the below paragraph to our Methods section:

"UMAP generates low-dimensional embeddings from the original high-dimensional space, preserving the probability of connections (transformed from distance measures) between data points and their nearest neighbors. While a single distance measurement in the UMAP space is unitless and less meaningful on its own, we believe the distance ratio between the same stimulus across days and nearby competing stimuli is more informative. Conceptually, this ratio resembles an odds ratio in statistics. We posit that when representational drift is low, stimulus representations from one day are more likely to be connected to the same stimulus from another day rather than to competing stimuli."

Our quantitative use of UMAP is consistent with the following literature: Testard et al., quantified distance between clusters in UMAP embedding space (their Fig5f). UMAP has also been quantitatively used for decoding analysis as in Vollan et al., (their Ext.Fig.4).

Testard, C., Tremblay, S., Parodi, F. et al. Neural signatures of natural behaviour in socializing macaques. Nature 628, 381–390 (2024). <https://doi.org/10.1038/s41586-024-07178-6>

Vollan, A.Z., Gardner, R.J., Moser, MB. et al. Left–right-alternating theta sweeps in entorhinal–hippocampal maps of space. Nature (2025). <https://doi.org/10.1038/s41586-024-08527-1>

6. In Fig3f-g the term "50% max separation time" is unclear in my opinion. Maybe "feature separation time" would be more easily understandable.

A: We have changed all mentions of "50% max separation time" to "feature separation time".

7. In the field of representational drift it is largely up to the authors if they frame their data as showing more drift or less drift than expected. Considering most analysis does show some time dependent changes I think the message is leaning a bit too heavily towards there is very little drift. I think the wording could be more balanced in the introduction and discussion.

A: We have made the following changes in the Introduction:

"... temporal codes with fast dynamics (tens of milliseconds), compared with rate codes (hundreds of milliseconds), ~~significantly~~ improved the stability of single neuron tuning, especially for less reliable neurons."

"We found that visually evoked spiking time courses could ~~consistently~~ largely represent stimulus features over time consistently when the readout time bins were fixed."

And in the Discussion:

"... along stimulus-coding dimensions at 10 ms resolution across 15 days and found overall stable, distinct dynamical representations."

8. I find it strange that the legends contain no n numbers. They are listed in a supplementary table, but it is quite tedious to refer to a separate supplementary document for n numbers, units and statistics.

A: We have added back N numbers.

Given the number of statistical tests and detailed statistical test entries we have (>250), we previously thought it might be distracting to include all n numbers, units and statistical entries in the main manuscript file. Most of our figure legend is close to or has surpassed the 350-word limit (<https://www.nature.com/documents/ncomms-submission-guide.pdf>). Our decision to move them elsewhere might be consistent with past literature facing the similar dilemma, e.g. de Vries et al., where n numbers of major figures (e.g. their Fig2-5) are reported outside main text, e.g. in their extended figures 3-9). We did not report in extended figures (as a table) because the table was too large and should be included as a data file per manuscript submission instructions https://mts-ncomms.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions "Tables larger than one page (in general, >50 rows or >10 columns) should be provided as tabular data files ..."

de Vries, S.E.J., Lecoq, J.A., Buice, M.A. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat Neurosci* 23, 138–151 (2020). <https://doi.org/10.1038/s41593-019-0550-9>

9. Pooling neurons across animals assumes uncoordinated drift within the population. I think it is fine to assume this but it needs to be stated and discussed. E.g. in Fig3 and 5.

A: This is a great point. Yet, other than the qualitative presentation in Fig3a, we did not pool neurons (if you meant creating pseudo-animals with more neurons per animal but fewer animals) across animals in Fig3 or Fig5 before identifying low dimensional populational stimulus representations. We were explicit about not pooling neurons in figure captions. E.g. in the legend of Fig3b "Disentangled population representation dynamics during drifting gratings from *an example animal*". Fig3f "linear mixed effect, time, stimulus category as fixed effect, *individual animals* as random effect." Fig5a-b "The original high-dimensional representation was constructed with all tuned neurons that appeared for 15 days in *1 example mouse*" Fig5e-h "Linear mixed effect: time, representation definition method as fixed effect, *animal-stimulus pairs* as random effect".

We apologize for any misunderstanding. We have now made it more clear in the Results for Fig3 and Fig5 that neurons were not pooled across animals after Fig3a. Fig3: "... To quantify the stability of the population dynamics that encode individual stimulus categories (e.g., 24 static gratings can be grouped into either 6 orientations or 4 spatial frequencies), we projected the population firing time course of each animal into a low-dimensional stimulus coding space with cross-validated LDA decoders...".

Fig5: "... We next extended the study from single neuron tuning to the longitudinal stability of visual representations formed by neuronal populations within each animal. We constructed populational representations on "super sessions," which involved pooling tuning curves from neuron populations from the same animals across all longitudinal sessions for joint dimension reduction"

10. Reference 8 is a thesis, but the corresponding paper is peer reviewed, and should be cited in addition or instead.

A: We have improved this citation item, citing the corresponding peer-reviewed paper.

Jeon, B.B., Swain, A.D., Good, J.T., Chase, S.M. & Kuhlman, S.J. Feature selectivity is stable in primary visual cortex across a range of spatial frequencies. *Scientific Reports* 8, 15288 (2018).

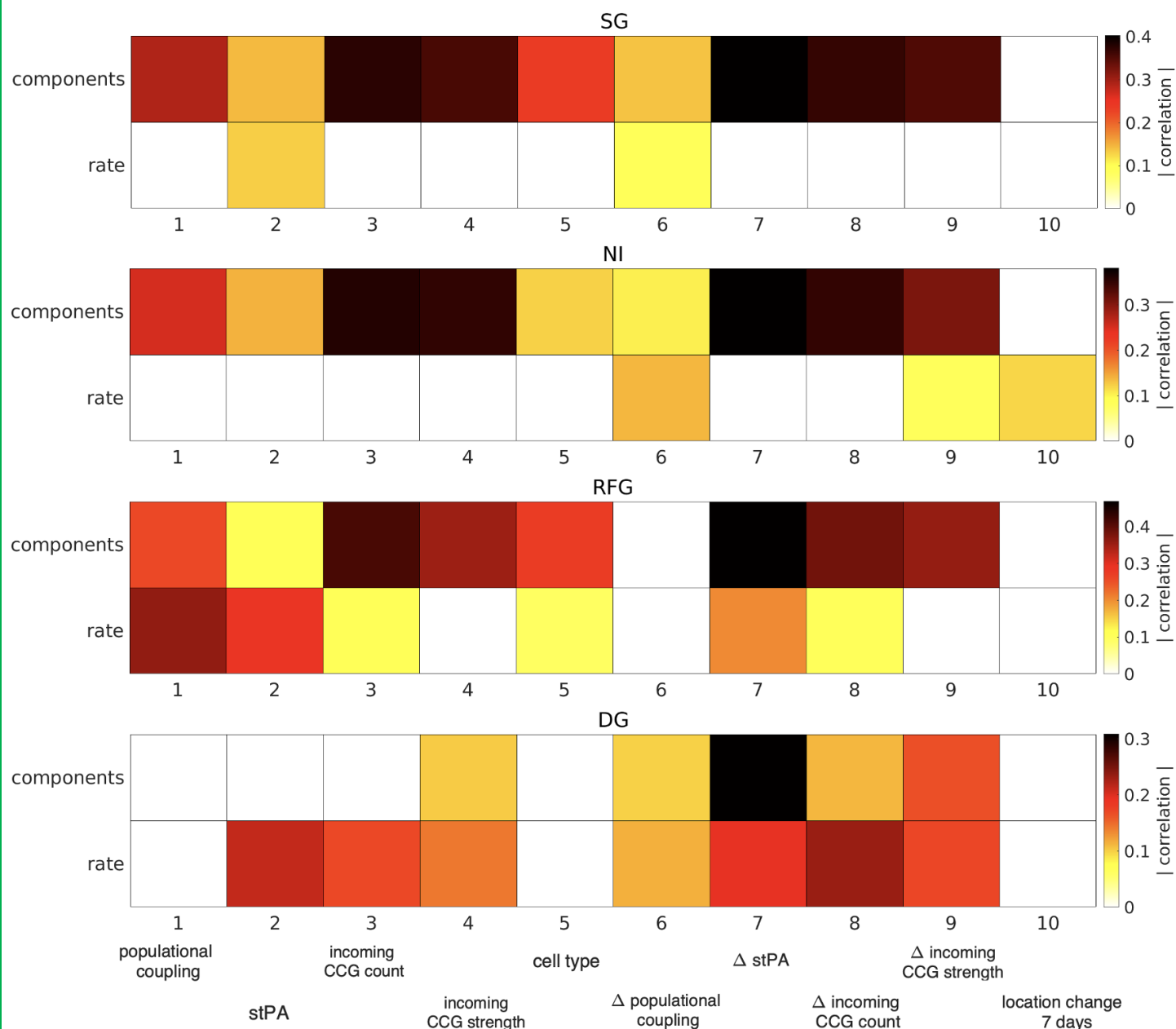
11. I'm not sure it makes sense to have Fig 3 (population level analysis) before fig4 (single cell analysis), when Fig 5 again looks at population analysis. It might make the paper more readable to switch Fig 3 and 4, or to have Fig 5 as a supplement.

A: The order of Fig3 and Fig4 was explained previously in the paragraph just before Fig3 "As a result, the stability of the neural code representing sensory information at higher temporal resolution (e.g., tens of milliseconds) remains unknown. Therefore, we studied the fine temporal resolution of electrophysiological recordings to explore the chronic stability of the temporal patterns of neural firing (Fig. 3) and how these patterns contributed to enhancing stimulus coding reliability and stability (Fig. 4)." Despite being a populational analysis, at its core Fig3 still uses rate code, just

at a high temporal resolution. We believe the result in Fig3 brings a natural transition into the “true” temporal code at single neuron level (Fig4), and later populational analysis built from single neuron temporal code in Fig5.

12. It would help to summarize the results in extended Fig8-9 using correlation heat maps.

A: We have now summarized the results in correlation heat maps using an additional extended figure.



Legend Extended Fig 10: A summary of the correlation strength between tuning curve similarity after 7 days and different variables tested in Extended Fig8-9. Insignificant correlations were set to 0. Temporal components based tuning curve similarity were better associated with many connectivity metrics than firing rate based for 3 out of 4 types of stimuli tested (visual inspection, statistics not performed).

13. A recent paper by Bauer et al., (10.1038/s41467-024-53326-x) models the effect of spontaneous and experience dependent connectivity changes on rate code stability in mouse V1. This paper should be mentioned and discussed in relation to the findings in Fig6.

A: We have now discussed Bauer et al., in relation to the findings in Fig6 in our Discussion. (third paragraph from the last).

"...Therefore, network functional connectivity metrics explained tuning stability more effectively when stability was defined in terms of stimulus-coding temporal components, compared to when it was defined by firing rate. Recently, it is proposed[ref] that drift can result from both experience dependent Hebbian plasticity and experience-independent synaptic volatility. Since the novelty of the visual experience was not explicitly manipulated in this study, our findings suggest that temporal codes may better capture synaptic volatility compared to rate codes. The role of temporal code in Hebbian plasticity awaits further studies."

[ref] Bauer, J., Lewin, U., Herbert, E. et al. Sensory experience steers representational drift in mouse visual cortex. Nat Commun 15, 9153 (2024). <https://doi.org/10.1038/s41467-024-53326-x>

14. In future, it would help to have line numbers to make it easier to refer to specific sentences during the review process.

A: We did not number the Word file due to these instructions for submission "If you upload a Word document the system will number the lines for you" https://mts-ncomms.nature.com/cgi-bin/main.plex?form_type=display_auth_instructions. The PDF version contained line number which the other two reviewers utilized; Yes, we will also try to number the Word file during revision submissions if it doesn't seem to conflict with the line number in the resulting system-converted PDF file.

Reviewer #2 (Remarks to the Author):

The paper by Zhu, He et al. addresses factors determining stability and representational drift in mouse visual cortex. Representational drift is a topic that is currently intensely discussed in the field, with recent publications differing quite strongly in the degree of stability found, depending on the brain area studied and the methods employed. The authors take a fresh look at this matter, leveraging a very recently introduced novel type of highly flexible, multi-site electrical recording probe ("NET"). These electrodes allow for recording multiple single units over weeks, with many individual neurons followed over time, as the authors convincingly show. When simply using average spike rate as a readout, as all previous studies on drift have done, they observe a certain degree of drift, depending on the visual stimulus used, in line with existing data. Crucially, they found drift to be substantially lower, when, instead of spike rate, using the temporal dynamics of the response in the range of tens of milliseconds. Taking response timing into account also improved stimulus discriminability and decoding accuracy. Finally, they show that the stability of temporal codes, as compared to a rate code, is better linked with functional connectivity, as assessed by single spike based cross-correlations.

Overall, these are very exciting results, which provide crucial insight into factors underlying the stability of neuronal representations in the neocortex. The study is carried out at very high technical standards and employing elaborated data analysis tools. I would like to point out that the longitudinal cross-correlation data (Fig. 6) alone are extremely rich, adding crucially to our understanding of how connectivity in the visual cortex changes over time.

The paper is very nicely written, with very instructive figures. I only have a few, mostly minor comments.

1. The recent Nature Comm paper by Bauer et al. on drift in visual cortex should be cited

A: Thanks for the comprehensive summary of our work. We have now discussed Bauer et al., in relation to the findings in Fig6 in our Discussion

"...Therefore, network functional connectivity metrics explained tuning stability more effectively when stability was defined in terms of stimulus-coding temporal components, compared to when it was defined by firing rate. Recently, it is proposed[ref] that drift can result from both experience dependent Hebbian plasticity and experience-independent synaptic volatility. Since the novelty of the visual experience was not explicitly manipulated in this study, our findings suggest that temporal codes may better capture synaptic volatility compared to rate codes. The role of temporal code in Hebbian plasticity awaits further studies."

[ref] Bauer, J., Lewin, U., Herbert, E. et al. Sensory experience steers representational drift in mouse visual cortex. Nat Commun 15, 9153 (2024). <https://doi.org/10.1038/s41467-024-53326-x>

2. Check legend for Fig. 2c

A: We have changed the legend into "Reliability, tuning strength, selectivity, and cell types could jointly explain tuning similarity after 7 days"

3. Line 176: Point to respective Figure panel

A: We have now pointed to the figure panel "Fig. 2b, d, 4th panel. Extended Fig3a, c"

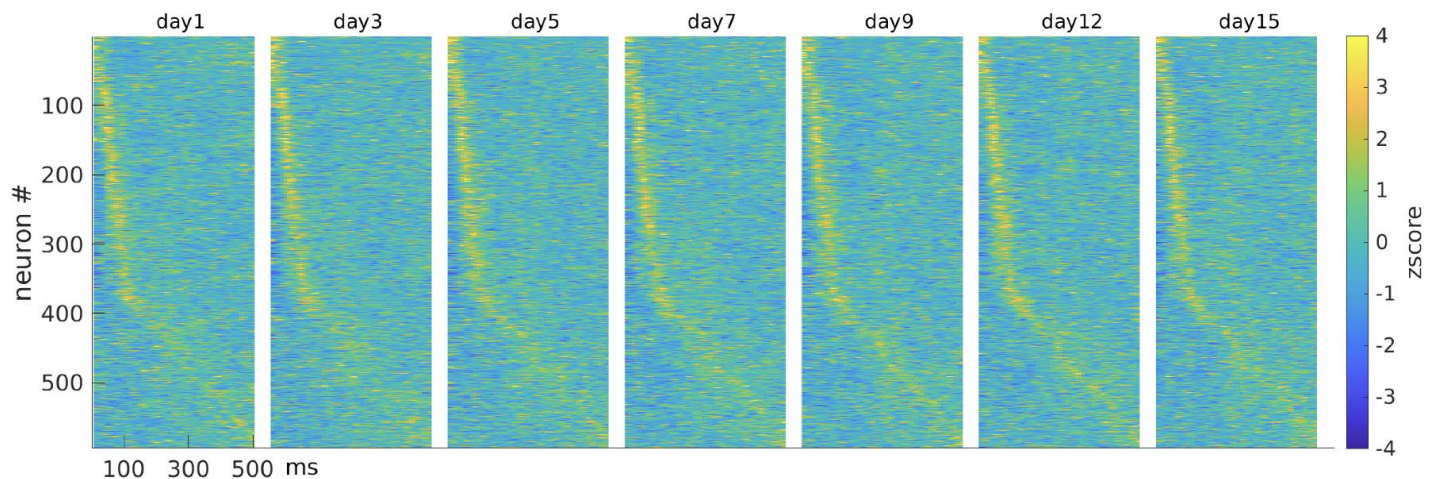
4. Line 182: I don't see any indication of a discrete group of putative anchor cells

A: We indeed don't see two distinct clusters of units where one cluster is stable where the other is not. Thus, we have revised the sentence into:

"These results indicate that neurons exhibit varying stability within-session (reliability) and across daily sessions. A small portion of neurons responded highly reliably and stably to visual stimuli despite noise[39] and mixed coding[40] in the cerebral cortex. These units might serve as anchors[41] for stable visual representations."

5. Figure 3a: How were these example neurons selected? Would the plot look similar for randomly selected neurons?

A: We selected neurons with neurons that appeared for all 15 days and tuned to drifting gratings. This description has been added to the figure legend. We have now included a reviewer figure below (this journal publishes all reviewer report after 2022) that shows the same information as Fig. 3a but simply using all neurons that appeared for all 15 days without the additional tuning criterion.



Reviewer Figure Legend: The stimulus-evoked spiking time course to a single, specific stimulus (i.e. 0° drifting grating) of a population of neurons across multiple days. Neurons were pooled from 5 mice over 7 different experiment days. Neurons were sorted by maximum activation time and were not resorted between days.

6. Regarding the temporally resolved responses: Was stimulus grating spatial phase randomized?

A: For drifting gratings stimuli, we did not modulate the stimulus initial spatial phase at the starting time of each stimulus. We have now added this description in to the Methods section. For static gratings stimuli, only two opposite phases were used. They were presented in a randomized order, but as described in the Methods, the responses from the two phases were not studied as we focused on the orientations and spatial frequencies of the static gratings.

7. Out of interest: How long can individual cells be reliably tracked using their electrodes?

A: Thanks for your interest, in our group's previous publication Luan et al., 2017 (their Fig3D FigS6 FigS7), we have made proof of principle demonstration of month-long tracking of same neurons with their waveform from single electrodes. We expect to separately publish more systematic months-long tracking of same neurons with over-sampling electrodes as reported in this manuscript in the future.

Luan, L., Wei, X., Zhao, Z., Siegel, J. J., Potnis, O., Tuppen, C. A., ... & Xie, C. (2017). Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Science advances*, 3(2), e1601966.
<https://www.science.org/doi/10.1126/sciadv.1601966>.

8. Line 560: please check this sentence, not sure it is clear like this

A: Thanks. We have changed the sentence into:

"Here, we further showed that temporal components exhibited reduced drift in single neuron stimulus tuning compared to firing rates-based tuning. The reduction was particularly prominent for neurons with low firing rate-based reliability".

9. Line 570: should read 'drifting'

A: Thank you for kindly pointing this out. We have now changed the word into 'drifting'.

10. Ext. Fig. 1b: it would be nice to see such data from an electrode that actually did move

A: These plots in Ext.Fig.1b are quite representative of the typical movements of our electrodes. In our group's previous publication Luan et al., 2017 (their Fig3D FigS6 FigS7), we also performed two photon imaging (their Fig4J) around the implanted electrodes and observed ~25 microns of relative movement between neurons and our probe across 1 month.

Luan, L., Wei, X., Zhao, Z., Siegel, J. J., Potnis, O., Tuppen, C. A., ... & Xie, C. (2017). Ultraflexible nanoelectronic probes form reliable, glial scar-free neural integration. *Science advances*, 3(2), e1601966.
<https://www.science.org/doi/10.1126/sciadv.1601966>.

11. Ext. Fig. 1g: p values are not shown in the Figure, but are referred to in the legend

A: Thanks a lot for the reminder, we have corrected the sentence in the legend to "p<0.0001 with values reported in supplementary table 1."

12. Line 814: 'with' missing

A: we have now added the missing word 'with'. We appreciate your detailed review of our manuscript.

13. References: several of these are incomplete

A: Thanks for pointing this out. We have improved the following reference in terms of missing volume/page number/journal and misplaced author surname vs given name: #1, #3, #7, #8, #29, #32, #33, #43, #46, #55, #57, #61, #73, #83, #86, #88, #89, #92, (as numbered in the original submission.) Please let us know if there are remaining incomplete references that you have concerns.

1.Hengen, K.B., Torrado Pacheco, A., McGregor, J.N., Van Hooser, S.D. & Turrigiano, G.G. Neuronal Firing Rate Homeostasis Is Inhibited by Sleep and Promoted by Wake. *Cell* 165, 180-191 (2016).

3.Mankin, E.A., Diehl, G.W., Sparks, F.T., Leutgeb, S. & Leutgeb, J.K. Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* 85, 190-201 (2015).

7.Montijn, J.S., Meijer, G.T., Lansink, C.S. & Pennartz, C.M. Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective. *Cell Reports* 16, 2486-2498 (2016).

8.Jeon, B.B., Swain, A.D., Good, J.T., Chase, S.M. & Kuhlman, S.J. Feature selectivity is stable in primary visual cortex across a range of spatial frequencies. *Scientific Reports* 8, 15288 (2018).

29.Zhao, Z., et al. Ultraflexible electrode arrays for months-long high-density electrophysiological mapping of thousands of neurons in rodents. *Nat Biomed Eng* 7, 520-532 (2023).

32.Woloszyn, L. & Sheinberg, D.L. Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron* 74, 193-205 (2012).

33.Zhu, H., et al. Clustering with fast, automated and reproducible assessment applied to longitudinal neural tracking. *arXiv preprint arXiv:2003.08533* (2020).

43.Yiling, Y., Klon-Lipok, J. & Singer, W. Joint encoding of stimulus and decision in monkey primary visual cortex. *Cerebral Cortex* 34, bh420 (2024).

- 46.Oude Lohuis, M.N., Marchesi, P., Olcese, U. & Pennartz, C.M.A. Triple dissociation of visual, auditory and motor processing in mouse primary visual cortex. *Nat Neurosci* 27, 758-771 (2024).
- 55.Trautmann, E.M., et al. Large-scale high-density brain-wide neural recording in nonhuman primates. *bioRxiv* (2023).
- 57.Jensen, K.T., Kadmon Harpaz, N., Dhawale, A.K., Wolff, S.B.E. & Olveczky, B.P. Long-term stability of single neuron activity in the motor system. *Nature Neuroscience* 25, 1664-1674 (2022).
- 61.Papaioannou, S. & Medini, P. Advantages, Pitfalls, and Developments of All Optical Interrogation Strategies of Microcircuits in vivo. *Frontiers in Neuroscience* 16, 859803 (2022).
- 73.Guidera, J.A., et al. Regional specialization manifests in the reliability of neural population codes. *bioRxiv* (2024).
- 83.Sotomayor-Gómez, B., Battaglia, F.P. & Vinck, M. Temporal spiking sequences in visual cortex carry unique information about natural movies. *bioRxiv* (2023).
- 86.Garrett, M., et al. Stimulus novelty uncovers coding diversity in survey of visual cortex. *bioRxiv* (2025).
- 88.Bauer, J., et al. Sensory experience steers representational drift in mouse visual cortex. *Nature Communications* 15, 9153 (2024).
- 89.Nguyen, N.D., et al. Cortical reactivations predict future sensory responses. *Nature* 625, 110-118 (2024).
- 92.Failor, S.W., Carandini, M. & Harris, K.D. Visual experience orthogonalizes visual cortical stimulus responses via population code transformation. *Cell Reports* 44, 115235 (2025).

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors record neuronal responses in areas V1 and LM of the mouse for several days (up to 15) and test the stability of their tuning for various stimulus categories at both the single unit and population level. They found that, in general, the tuning is quite stable across days, although to a different degree for different units. Interestingly, the extent to which a neuron is able to maintain its tuning for visual stimuli over time correlates with its trial-by-trial reproducibility. Additionally, they show that, taking into account the dynamics of the response during stimulus presentation, as opposed to just computing the overall firing rate, provides more discriminatory information about the visual stimuli and such information is more stable across days of recordings.

Overall, the focus of the study and the conclusions reached by the authors are interesting and relevant. Not many studies have been able to track neuronal tuning across days. As such the results about the day-by-day stability of the visual representation are novel and important for the community, especially the relationship with the trial-by-trial reproducibility of the response. The results concerning the additional information provided by the dynamics of the responses (what the authors refer to as temporal code) are less surprising (by definition, a higher dimensional representation contains richer information than a lower dimensional one). Still, it is interesting how the authors investigate the various temporal profiles of the responses and the way they allow a better discrimination of the visual stimuli.

In summary, in terms of content, I think that the study is valuable and does provides some new knowledge to the vision science community. The problem, however, is the way the manuscript is written, the analyses are described, and the results are presented. Sentences are often very unclear and convoluted or syntactically wrong. The details of the analyses are often neglected. In the Results section, the single unit and population metrics used to assess stimulus discriminability and neuronal tuning are almost never explained intuitively. In the Methods section, if such metrics are explained, the text is often so short and confusing to be of little help (equations are rarely used). Same goes for the legends. This is a problem, especially for those analyses that are not very conventional and because the figures are extremely rich and complex (with tens of subplots and panels) and the authors jump from one analysis to the next without providing much explanation. This is a serious problem because, in many instances, I could only marginally grasp the meaning and the solidity of the results presented by the authors.

To sum up, the manuscript, as it is currently written, it is not suitable for publication, which is a pity, given that, as I wrote above, several of the findings presented here seem interesting. So, I urge the authors to carry out a deep revision of the way the write, present, explain and discuss their analyses and results. Below, I provided some more detailed feedback.

MAJOR CONCERNS

1. Clarity of the writing. As explained above, the lack of clarity in the writing is so pervasive that it is hard to point a few specific instances. Still, I provide some comments below for what concerns the title, abstract and introduction.

Title.

The title, "Temporal coding stabilizes visual cortical representations", lacks clarity and impact for two main reasons. First, the expression "temporal coding" is quite vague: it does not refer to a specific attribute of neural responses, but rather encompasses a broad set of features related to the temporal dimension of neural coding. For instance, (Panzeri et al., 2010) distinguish between "latency code", "interspike intervals", and "phase of firing", each pinpointing to different aspects of the temporal dynamics of neural activity. The fact that the title does not specify which exact dimension of temporal coding is being discussed introduces ambiguity and potentially undermines the paper's accessibility and comprehensibility. The second issue concerns the use of the term "stabilizes", which is somewhat ambiguous without further specification. Does it refer to stabilization over time or in space? Since the paper focuses on temporal stabilization, adding "over time" to the end of the title would have helped clarify the intended meaning. Also, stabilization over time can simply mean over the course of a few hundreds of ms or seconds - see for instance (Piasini et al., 2021; Rudelt et al., 2024). What is peculiar of this work is that stability is assessed across days. So it would be worth stating this upfront in the title.

A: Thanks for your detailed review of our manuscript. Following your kind suggestion and reviewer 1's suggestion (comment 1), we have revised the title into "Temporal coding carries more stable cortical visual representations than firing rate over time". We acknowledge that there are multiple ways of extracting temporal codes such as through latency, phase and interspike interval coding. Indeed, we wrote in limitation Line 608-609 (original submission) and cited (Panzeri et al., 2010) to acknowledge that future work could delve into different specific aspects of the temporal codes. Yet, our current approach of temporal code extraction is more general (finding structures in the temporal spiking patterns) and do not seem to employ a single category of temporal code above. We thus find it appropriate to keep a general term "temporal coding" in the title. In fact, our approach is more similar to another temporal code paper from Dr. Panzeri, (Zuo et al., 2015) where PCA based temporal components were extracted. Our use of a general term "temporal coding" is aligned with the use of a general term "spike timing" in their title. Finally, the use of "temporal coding" in the title also stays consistent with past literature, such as Huxter et al., Smith et al., and Jadhav et al.,

Zuo, Y., Safaai, H., Notaro, G., Mazzoni, A., Panzeri, S., & Diamond, M. E. (2015). Complementary contributions of spike timing and spike rate to perceptual decisions in rat S1 and S2 cortex. *Current Biology*, 25(3), 357-363. [https://www.cell.com/current-biology/fulltext/S0960-9822\(14\)01560-7](https://www.cell.com/current-biology/fulltext/S0960-9822(14)01560-7)

Huxter, J., Burgess, N. & O'Keefe, J. Independent rate and temporal coding in hippocampal pyramidal cells. *Nature* 425, 828-832 (2003). <https://doi.org/10.1038/nature02058>

Smith, E.H., Horga, G., Yates, M.J. et al. Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nat Neurosci* 22, 1883-1891 (2019). <https://doi.org/10.1038/s41593-019-0494-0>

Jadhav, S., Wolfe, J. & Feldman, D. Sparse temporal coding of elementary tactile features during active whisker sensation. *Nat Neurosci* 12, 792-800 (2009). <https://doi.org/10.1038/nn.2328>

Abstract.

The abstract assumes some specific prior knowledge and is somewhat convoluted. Expressions such as "representational drift" are quite specialized and may not be universally recognized. Expressions such as "temporal coding" on the other hand are too vague and broad (see the criticism above). It would be better if the authors simply tried to explain what they did rather than use jargon of this kind, which may be either too specialized or too vague for many readers. Additionally, some sentences are cryptic, such as: "Across various types of stimuli, we found that neurons exhibited varying degrees of day-to-day stability in their slower, second time-scale rate code tuning, with this stability being explained by the reliability of their tuning". The meaning of such sentences can only be grasped after reading them several times.

A: Thanks for your feedback. We understand that representational drift may not be generally understood from your perspective. Yet, despite not being universally recognized, representational drift is widely recognized. For example, Reviewer 2 believes that it is a hot topic in the field "Representational drift is a topic that is currently intensely discussed in the field..." and Reviewer 1 believes that it is already a field by itself "In the field of representational drift it is largely up to the authors...". To put our work into the relevant context and appropriately deliver our point, we have thus decided to align our usage of the term "representational drift" with relevant literature, such as (Zhang et al., 2024) (Marks et al., 2021) (Nguyen et al., 2024). These papers used the term in their title/abstract.

i) We have now made slight changes to our abstract such that the level of explanation will be similar to (de Snoo et al., 2023) and (Bauer et al., 2024).

"Previous studies have used slow dynamic (1-5 seconds) rate code measurements to study visual tuning, revealing varying degrees of gradual activity changes over time or "representational drifts."

ii) we have revised the sentence introducing "temporal coding" into "In this study, we extracted structures in fast (tens of milliseconds) temporal responses and explored the role of such temporal codes in supporting the stability of visual representations. This level of explanation is consistent with those in (Zuo et al., 2015).

iii) we have improved the clarity of this sentence "Across various types of stimuli, we found that neurons exhibited varying degrees of day-to-day stability in their slower, second time-scale rate code tuning, with this stability being explained by the reliability of their tuning" by breaking it into two sentences: "Across various types of stimuli, we found that neurons exhibited varying degrees of day-to-day stability in their firing rate-based tuning. The across day stability correlated with within day reliability."

Zhang, Y., Wang, M., Zhu, Q., Guo, Y., Liu, B., Li, J., Yao, X., Kong, C., Zhang, Y., Huang, Y., Qi, H., Wu, J., Guo, Z. V., & Dai, Q. (2024). Long-term mesoscale imaging of 3D intercellular dynamics across a mammalian organ. *Cell*, 187(21), 6104–6122.e25. <https://doi.org/10.1016/j.cell.2024.08.026>

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature communications*, 12(1), 5169. <https://doi.org/10.1038/s41467-021-25436-3>

Nguyen, N. D., Lutas, A., Amsalem, O., Fernando, J., Ahn, A. Y., Hakim, R., Vergara, J., McMahon, J., Dimidschstein, J., Sabatini, B. L., & Andermann, M. L. (2024). Cortical reactivations predict future sensory responses. *Nature*, 625(7993), 110–118. <https://doi.org/10.1038/s41586-023-06810-1>

de Snoo, M. L., Miller, A. M. P., Ramsaran, A. I., Josselyn, S. A., & Frankland, P. W. (2023). Exercise accelerates place cell representational drift. *Current biology : CB*, 33(3), R96–R97. <https://doi.org/10.1016/j.cub.2022.12.033>

Bauer, J., Lewin, U., Herbert, E., Gjorgjieva, J., Schoonover, C. E., Fink, A. J. P., Rose, T., Bonhoeffer, T., & Hübener, M. (2024). Sensory experience steers representational drift in mouse visual cortex. *Nature communications*, 15(1), 9153. <https://doi.org/10.1038/s41467-024-53326-x>

Zuo, Y., Safaai, H., Notaro, G., Mazzone, A., Panzeri, S., & Diamond, M. E. (2015). Complementary contributions of spike timing and spike rate to perceptual decisions in rat S1 and S2 cortex. *Current Biology*, 25(3), 357–363. [https://www.cell.com/current-biology/fulltext/S0960-9822\(14\)01560-7](https://www.cell.com/current-biology/fulltext/S0960-9822(14)01560-7)

Introduction.

The Introduction fails to provide the necessary background for the reader to clearly understand the research question posed by the authors, its relevance within the field, and the logical progression that led to the experimental design. The concepts underpinning the entire work, such as rate coding and temporal coding, are never precisely defined, nor are they clearly contrasted as differential aspects of neural coding.

Furthermore, this section does not outline the predictions regarding potential results and their interpretations, making it challenging for the reader to follow the underlying rationale of the study.

Moreover, the summary of the results is way too dense, squeezed in poorly readable sentences such as: "Neurons showed different day-to-day rate code tuning stability, with the degree of stability best explained by tuning reliability" (lines 61-62), and "Accordingly, adding temporal dynamic information stably improved the longitudinal causal decoding accuracy. Lastly, we found that the reliability of temporal code was associated with static and dynamic properties of network functional connectivity." (lines 67-69).

A: We now explain core concepts underpinning the entire work and outline our prediction with the following modifications:

i) "...Consequently, these studies mostly use seconds-long rate codes (firing rates, spike counts normalized by elapsed time) to compute visual representations, largely eliminating structures and information in the timings of the spikes.

~~Indeed,~~ Information and structures in the timing of stimulus-evoked spikes, referred to as the temporal code¹⁷, may describe unique aspects of external stimuli. We thus hypothesize that temporal code and rate code have different rates of representational drift over extended periods.

Temporal code is supported by both anatomical and functional evidence...."

ii) We now break Lines 61-62 into two sentence to improve readability. "Neurons displayed varying levels of day-to-day stability in their rate code tuning. This stability correlated with their tuning reliability."

iii) We have revised Lines 67-69 into the following "Accordingly, adding temporal dynamic information consistently improved the ~~longitudinal-causal~~ stimulus decoding accuracy in future days. Lastly, we found that the ~~reliability~~ stability of temporal code was associated with ~~static and dynamic properties of network functional connectivity and its changes.~~" (lines 67-69).

Several syntactical imprecisions in sentence construction undermine readability. For example, the sentence "These snapshot studies, however, have not elucidated whether the temporal dynamics are consistent chronically or their role in supporting the stability of visual representations." (lines 49-51) is particularly unclear.

A: We have changed this to "These snapshot studies, however, have not determined whether temporal dynamics remain consistent over days or how they contribute to the stability of visual representations."

Discussion.

This section of the manuscript also contains sentences whose meaning is difficult to grasp. Examples include:

Lines 568-570: "A breakdown characterization for each individual component (Supplementary Fig. 8a) showed that the stabilization effect was not due to individual components themselves drifted less than firing rates.". This sentence is grammatically awkward and unclear, making it challenging to understand the intended point about the stabilization effect and the individual components.

A: we followed reviewer 2 (comment 9)'s suggestion and changed the verb "drifted" into "drifting".

"A breakdown characterization for each individual component (Supplementary Fig8a) showed that the effect was not due to individual components themselves drifting less than firing rates."

Lines 571-574: "We showed that (Supplementary Fig. 8c) the reduction in representational drift (Figure 4f) offered by temporal code was not trivially due to temporal codes increasing reliability (Supplementary Fig. 4) and hence increasing the denominator (CCws) of the representational drift index.". The explanation here is highly technical and convoluted, with key elements left unexplained. The connection between temporal code reliability and the denominator of the representational drift index requires clearer elaboration.

A: We now have second thoughts in this claim and have remove it. We explain below the motivation of this claim.

It can be seen that CCws and reliability are closely related concepts: (Marks et al., 2021) note that "Where CCws is the Pearson correlation of the trial-averaged activity of two random halves of trials within a session". We defined reliability in our Methods section "Similar to past literature[36], the reliability of a neuron's rate code or temporal code to a particular stimulus type on any given day was defined as the correlation coefficient between its tuning function derived from one trial versus the mean tuning function computed from the rest of the trials, averaged over all such single trials". Thus, CCws and reliability conveys the same concept of "within-day stability" with the difference being how trials within that one session is partitioned. Since the denominator of representational drift index is (CCws + CCbs), we find this claim less solid as of now and have removed it.

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. Nature communications, 12(1), 5169. <https://doi.org/10.1038/s41467-021-25436-3>

Lines 588-589: "Therefore, network functional connectivity metrics better explained temporal components based single neuron tuning stability than that of firing rate based.". This sentence is syntactically flawed and lacks clarity, making it hard to discern how functional connectivity metrics relate to the stability of single-neuron tuning.

A: We have changed this to "Therefore, network functional connectivity metrics better explained the stability of temporal code than rate code at the single unit level."

These examples reflect the broader need for improved clarity and precision throughout the Discussion section.

2. Lines 119-121. "We used a published method, an actively-learned ensemble of hierarchical clustering trees to

match neurons ...". Too little info is provided about how neurons were matched across sessions. Fine to refer to a published method but some qualitative explanation of how the method works should be provided. Especially if a figure is shown with the outcome of the matching procedure. The figure by itself is very poorly explained. The legend is basically useless. The authors should explain what they are plotting and why in a way that is detailed enough to be useful and understandable.

A: We understand it might be a complicated method. Previously we explained the tracking procedure in the supplemental information – Supplemental Methods (directly after all supplemental figures) section. Here we further strengthen the explanation by adding another supplementary figure and another paragraph in the supplemental methods to illustrate the core principles behind the tracking.

Later part of the first paragraph in the supplemental methods:

"...To reduce the amount of manual effort, three strategies were applied: 1. Distance thresholding (masking) of invalid pixels in the adjacency matrix ([Supplementary Figure 9](#)). 2. Unsupervised Mutual nearest neighbor[1] match (MNN) between units from temporally adjacent sessions for track initialization ([Supplementary Figure 9](#)). 3. Active learning (user feedback) with an ensemble of hierarchical clustering trees whose leaf nodes were initialized by MNN ([Supplementary Figure 9](#)). 4. Finally, we checked manual labeling consistency with cross-validated deep learning and linear discriminant analysis methods. ([Supplementary Figure 10](#))"

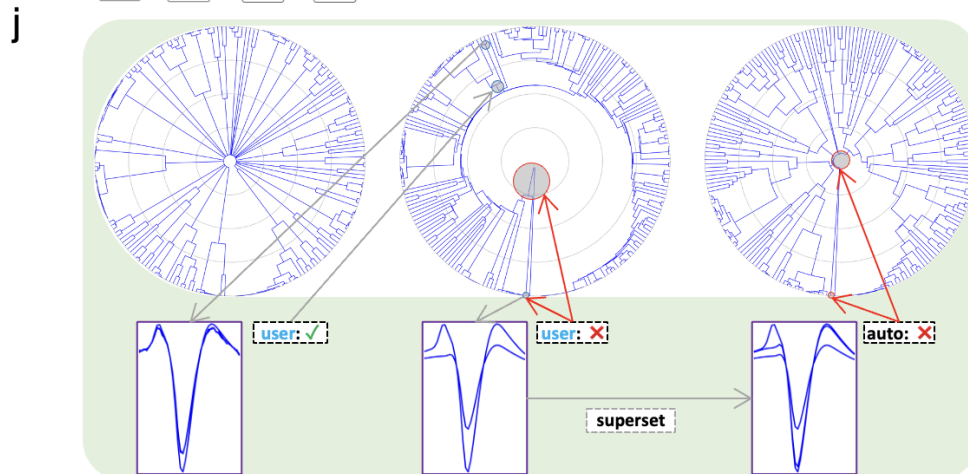
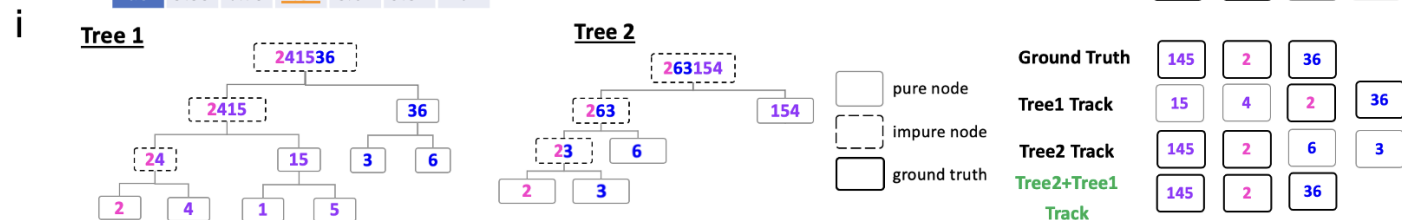
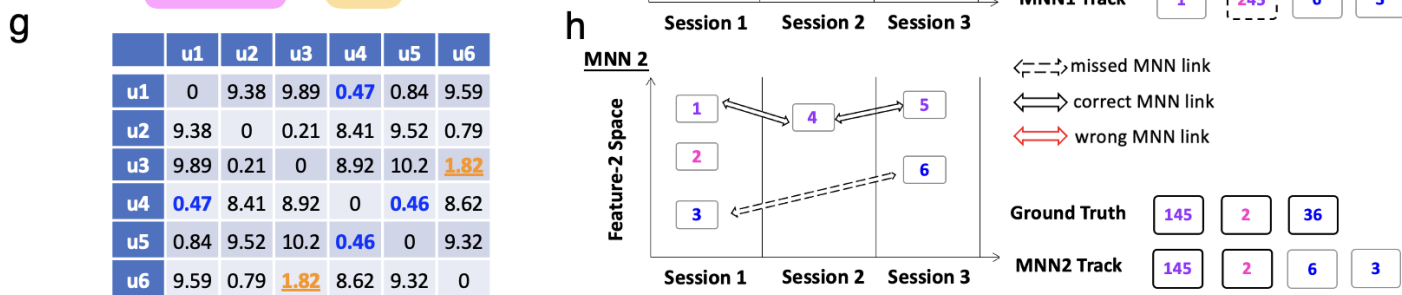
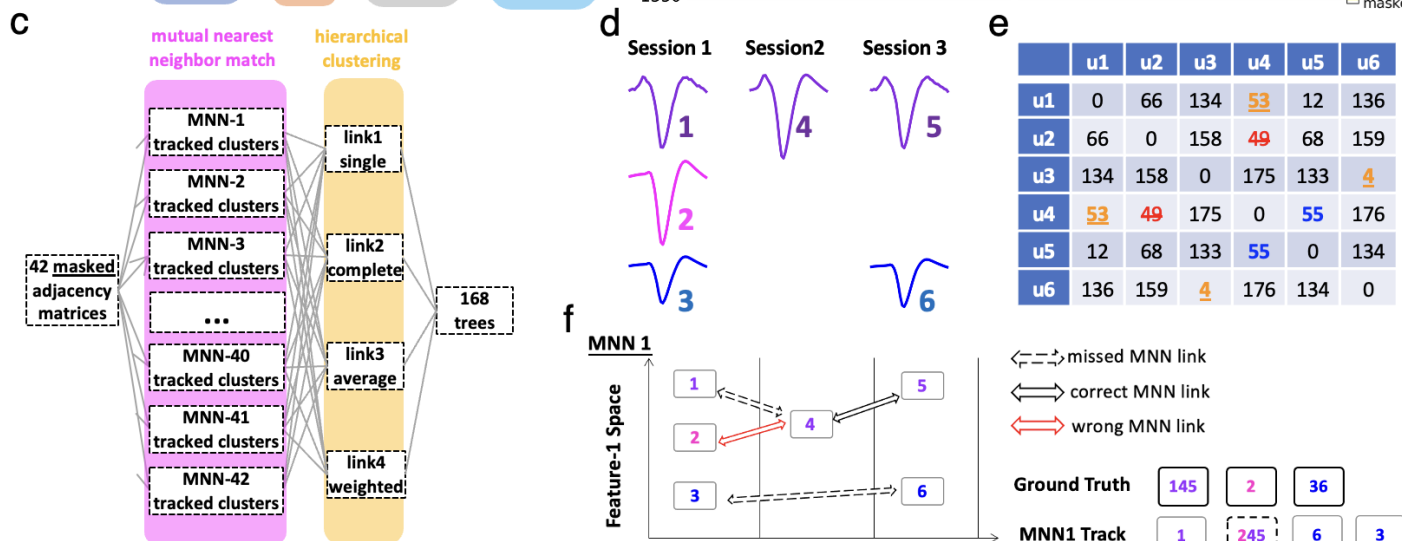
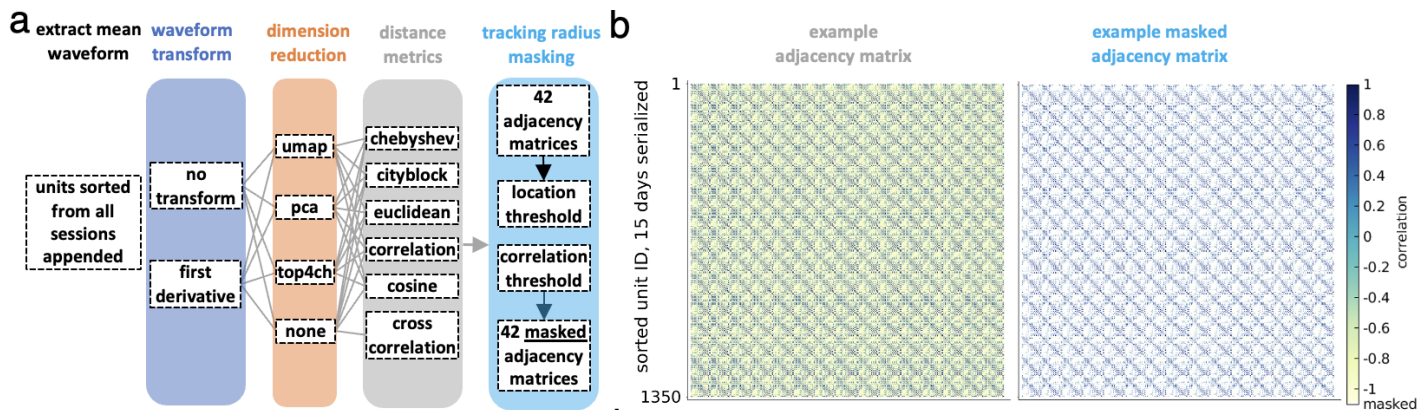
Added paragraph immediate below the first paragraph:

"Here we further explain the 3 core principles behind the tracking.

1. Distance thresholding: a large number of the $\binom{n}{2}$ pairwise decisions in the adjacency matrix did not merit further visual inspection. They were automatically masked as 0 following distance thresholds, such as having correlation coefficient between unit waveforms smaller than 0.7. [Supplementary Figure 9b](#)

2. Mutual nearest neighbor match: Unit X in session N is matched/linked to another unit Y in the next immediate session N+1 if a) Y is the closest unit in session N+1 relative to unit X and b) unit X is the closest unit in session N relative to unit Y. The process was repeated until all possible links/tracks of units were formed. An example matching process for 6 units across 3 sessions using 2 different distance metrics were shown in [Supplementary Figure 9.d-h](#). Despite the algorithmic simplicity, MNN could not link units that show up on non-consecutive days ([Supplementary Figure 9f,h](#)). In addition, false positive matches could still be generated ([Supplementary Figure 9f](#)).

3. For each MNN matched unit tracks, a hierarchical clustering tree initialized by those tracks ignored session labels and attempted to form increasingly larger clusters/tracks that eventually merged all units across all sessions. ([Supplementary Figure 9i](#)). This solved the issue of MNN which fails to link units showing up on non-consecutive sessions. Users were prompted to stop the merging at appropriate heights of each tree branches. Finally, combining tree cutting results from multiple trees improved the final tracking performance compared to only using individual trees ([Supplementary Figure 9i](#)). To improve reviewing efficiency, the next node to be prompted to the user was intelligently selected (actively learned) given the user's feedback on the current node. ([Supplementary Figure 9j](#))."



Supplementary Fig9: Principles of the semi-automatic method for tracking the same neurons across sessions

- a. A flow chart for the initial tracking process. An ensemble of diverse adjacency matrices was created.
- b. Example raw adjacency matrix formed by units from one probe pooled from 15 days using correlation as the distance metric. Units were sorted by session. Within each session, units were sorted by depth locations along the shank. $(14 \times 2) + 1 = 29$ stripes can be seen, representing potentially repeatedly identifiable units that were similar to themselves between sessions. A large number of elements in each adjacency matrix were masked after applying simple location and correlation thresholds.
- c. A flow chart for later tracking procedures. The pairwise adjacency matrices were initially flattened (forming clusters) using the mutual nearest neighbor algorithm between all adjacent sessions. The initial clusters were further linked recursively to form 168 hierarchical clustering trees. See (d) to (j) for expanded explanation of these processes.
- d. A simulated scenario demonstrating the tracking procedure: six hypothetical units sharing the same top channel were detected/sorted in 3 individual sessions. Unit group 1-4-5 represented the same unit across time, while unit group 3-6 were another neuron which did not appear in the second session. Unit 2 was a different unit from groups 1-4-5 and 3-6, appearing only in session 1. For simplicity, only the top channel waveform of each unit is shown; in actual tracking, all available channels from a 32-channel shank were used. Unit 2 was obtained from a real recording, while unit group 1-4-5 (characterized by a positive peak before depolarization with slight amplitude fluctuation across sessions) and unit group 3-6 (major amplitude change from unit 2 with slightly different depolarization and repolarization slopes) were derived from Unit 2 to simulate a difficult tracking scenario.
- e. Distance matrix comparing the mean squared error (distance feature 1) between all pairs of units. This serves as the starting point of the past unsupervised method and the initialization portion of the current method. Mutual nearest neighbor (MNN) pairing across adjacent sessions was performed on this matrix. Red strikethrough: False positive pairs detected by the MNN. Orange: Missed pairs by the MNN. Blue: Correctly detected MNN pairs.
- f. The distance matrix in (e) was transformed into a one-dimensional feature space (e.g., using a multidimensional scaling algorithm) and plotted across time to illustrate the MNN matching result. Unit 2 incorrectly prevented Unit 1 from linking with Unit 4, as Unit 2 was closer to Unit 4 than Unit 1. Similarly, Unit 3 failed to link with Unit 6 due to the absence of a mutual nearest neighbor in session 2, preventing a bridge between Unit 3 (session 1) and Unit 6 (session 3). Consequently, MNN tracking based on this feature resulted in three fragmented units and one incorrectly tracked unit group.
- g. Another distance matrix comparing correlation distance (distance feature 2) between all pairs of units. For ease of presentation, the data was transformed into $(1 - \text{correlation coefficient}) \times 1000$. Mutual nearest neighbor (MNN) pairing across adjacent sessions was performed on this matrix. Orange: Missed pairs by the MNN. Blue: Correctly detected MNN pairs.
- h. Same as (f) but for summarizing the MNN result in (g). Using this distance feature resulted in different tracking outcomes (two correctly tracked units and one fragmented unit) but still failed to reach the ground truth.
- i. Hierarchical clustering trees ignore session labels and attempt to form multiple possible links, ultimately joining all units. This approach solves the issue of missing session. (e.g., Tree 1, where unit3 is successfully linked with unit6). Leaf nodes in Trees 1 and 2 were initialized based on all valid MNN pairs found in (f) and (h) respectively. Invalid MNN pairs (identified after visual inspection) in (f) and (h) were broken into individual elements before tree formation. As trees progress from leaf nodes (bottom) to root nodes (top), they become increasingly impure (containing incorrect merges). However, since users manually verify tree nodes, tracking terminates before reaching impure root nodes, ensuring that no false linkages are automatically generated. Nevertheless, hierarchical clustering trees, like MNN, face limitations as no single distance metric fits all cases. Tree 1 failed to correctly track Unit 1-4-5, while Tree 2 failed for Unit 3-6.
Green: Combining results from both Tree 1 and Tree 2 produced the ground truth track. This justifies the approach outlined in (a) and (c), where multiple trees were formed using different combinations of distance metrics and hierarchical clustering methods.
- j. The procedure for cutting multiple trees at appropriate heights. Units in (b) were used to create Hierarchical clustering trees (only 3 shown for simplicity; 168 trees were used in total). The trees were visualized in circular layouts with root nodes at the center. Waveforms of the tracked units were overlaid and presented to users for confirmation. (Insets: top channel waveforms of units contained in tree nodes; actual tracking used all electrode channels.) User confirmation of unit merging results in selecting the parent node as the next query target. Conversely, rejecting a merge leads to automatically cutting all parent nodes in that tree. Nodes from other trees containing a superset of the rejected node's units are also automatically cut. This process continues until all nodes across all trees are verified, ultimately yielding a final set of correctly tracked units.

Finally, we have slightly revised the legend of Supplementary Fig1, we also place it after the newly added Supplementary Fig9 above, since Supplementary Fig9 describes the tracking itself where Supplementary Fig1 describes the validation of the tracking method by/against other methods.

"Supplementary Fig10: Validation of tracking same neurons, related to Fig 1 and Extended Fig1 and Supplementary Fig9

"a. Validation of using unit autocorrelogram (ACG) as an additional feature for tracking the same neurons. Top: Example ACGs from 6 units (rows) across 15 days (columns). ACGs maintained nearly identical shapes over time. Scale bar: 50 ms. Bottom: Single-unit ACGs were more like themselves (red) across sessions than to other units from the same probe within the same session (black, all valid single units across all mice, $p < 0.001$ see Supplementary Table 1). **b.** Flowchart of the tracking validation process. Diverse pair-wise distance metrics (see Supplemental Methods and Supplementary Figure 9) were computed for all sorted units pooled from individual days in a probe. 780,000 user-labeled unit pairs (Supplemental Methods and Supplementary Figure 9) were collected across 25 probes as ground truth. Each pair contained 43 distance features and a binary user label of "same" or "different" units. A dense neural network was trained to reproduce user decisions outputting the probability of merging any two units. **c-d.** Comparison of tracking performance between Mutual nearest neighbors-based tracking[1] (MNN, Supplemental Methods and Supplementary Figure 9, Figure 1d) and the current method. MNN method (c) was applied to units in an example probe. The horizontal lines (tracks) corresponded to the same units. Units could be tracked for limited duration. The resulting unit longevity tally plot fragmented into multiple short pieces. **d.** Unit longevity tally for the method: The number of fragmented tracks was drastically reduced, and average longevity increased compared to (c). **e.** Percentage of manual decisions of a given probe that were reproduced by either the deep learning method or a simple linear discriminant analysis method trained with labels from the other 24 probes. High reproducibility indicates strong internal labeling consistency. Paired t-test $p < 0.01$."

3. Lines 170-171. "The tuning similarity across all four stimulus types ...". What similarity metric was used here? A Pearson corr. coeff? If so, it should be stated somewhere more clearly. Also, in the histograms, a given cell how many data points contributed? ?? It should be explained. Also, in panels of Fig. 2 >=b the tuning similarity metric seems different from a, it is not bounded anymore between 0 and 1, so it cannot be a simple Pearson corr coeff. What is it then? This is a typical example of how poorly the metrics and the analyses are explained.

A: We have now stated it more clearly "The tuning similarity across all four stimulus types, defined as the Pearson's correlation coefficient between the tuning curves, displayed a wide range of variation" We have now added in the legend of this figure that "each neuron contributed 1 point, averaging all possible instances tuning curve similarity after 7 days interval". We further added in legend after Fig.2 >=b that "correlation coefficient was Fisher z-transformed". We performed the transform https://en.wikipedia.org/wiki/Fisher_transformation to better satisfy assumptions in the linear regression model (dependent variables should not be capped at a constant level i.e. 1)

4. Lines 224-226. "We fitted a decoder with a 90 ms window length at 160 ms after stimulus onset (from 70 ms to 160 ms) to ensure that different stimulus categories were readily decodable." The writing of the manuscript is in general quite poor, sentences are too long and convoluted. This makes it hard to understand some of the analyses. This is an example. I guess what the authors mean is that they built a linear decoder based on a rate codes, i.e., firing rates computed between 70 and 160 ms. If so the sentence "We fitted a decoder with a 90 ms window length at 160 ms after stimulus onset" is very misleading, because it seems that the integration window started at 160 ms and not at 70.

A: Thanks for pointing out the potentially misleading sentence in Line 224-226. We have revised it to "We built the linear decoder based on rate code at high temporal resolution, i.e., firing rates computed between 70 and 160 ms. This temporal window accounted for the time delay of visual signal propagation and ensured different stimulus categories were readily decodable."

Leaving this aside, it is not so clear what happened next. The authors somehow project the pop response vectors (measured over time in small 10 ms time bins) over a space obtained by the LDA. But how is this space defined? LDA should return a single weight vector that defines a direction in the representational space. If data points are projected along this vector/direction, they are maximally discriminable. But this is a single direction (a single dimension), while the authors seem to extract multiple directions/axes. From the main text, it is impossible to understand how these further directions were extracted. The legend, the methods and the figure itself provide some hints. There appear to be 3 dimensions here. But while I can understand how the authors ended up having two separate axes for orientation and direction or for orientation and SF (they could have built separate LDAs to discriminate orientation and direction

or orientation and SF), I really don't understand how they can get two different axes for the same property (i.e., two axes for orientation). They need to explain this. It looks like they did something equivalent to a PCA but using and LDA, but how this was achieved is really unclear. I would suggest using some equations in the Methods section to explain it in details and then provide an intuitive explanation in the Results, perhaps by adding a figure with a cartoon to illustrate the procedure.

A: We first would like to clarify that LDA can return multiple directions and axes. This has been exemplified in the literature. E.g. Tsao et al., 2018 used 2D embeddings (their Figure 2) and Magloire et al., 2019 used 3D embeddings (their Figure 1). The literature also implies that LDA is well-established and are often used without comprehensive explanation. E.g. Tsao et al., directly stated the number of embedding dimensions used without elaboration in their Methods: "followed by applying linear discriminant analysis to the dimension-reduced data and taking the top two linear discriminants". In the same vein, Magloire et al., stated "For 2D visualisation of features and their corresponding states, linear discriminant analysis was used to project the 20 features to 3 dimensions." without additional explanation.

We thank the reviewer for bringing it to our attention that multi-dimensional LDA may not be a method that everyone is familiar with. To improve the readership for the board audience, we have now cited a tutorial paper from Tharwat et al., 2017 about LDA. Their equation 8 and figure 3 (cartoon) are particularly useful in elaborating how multiple dimensions are extracted. Meanwhile, we have now included a brief overview of how LDA works in our Methods.

"LDA is a supervised dimensionality reduction technique that finds axes maximizing the linear separability of data based on class labels[ref1]. LDA computes eigenvectors (axes) corresponding to the eigenvalues of the matrix $(S_W)^{-1} S_B$ where S_W is the within-class scatter matrix, representing the variance within each class. S_B is the between-class scatter matrix, capturing the variance of class means relative to the global mean. However, when working with high-dimensional data, LDA can suffer from overfitting[ref2]. To address this, a regularization term is added to stabilize the computation. Instead of directly computing eigenvectors from $(S_W)^{-1} S_B$, they are now computed from $((1 - \gamma)S_W + \gamma * \text{diag}(S_W))^{-1} S_B$ where $\text{diag}()$ is the diagonal matrix operation, γ is the regularization factor between 0 and 1. (determined through cross-validation)."

Tsao, A., Sugar, J., Lu, L. et al. Integrating time from experience in the lateral entorhinal cortex. *Nature* 561, 57–62 (2018). <https://doi.org/10.1038/s41586-018-0459-6>

Magloire, V., Cornford, J., Lieb, A. et al. KCC2 overexpression prevents the paradoxical seizure-promoting action of somatic inhibition. *Nat Commun* 10, 1225 (2019). <https://doi.org/10.1038/s41467-019-08933-4>

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190.

[ref1] Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190.

[ref2] Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8(1), 86-100.

5. Lines 265-266. "To effectively extract temporal codes, we utilized cross-validated, regularized linear discriminant analysis (LDA) to determine the temporal components that best separated the different stimulus conditions." Again, this section is not understandable because it is not explained how this LDA based decomposition works.

A: We have now provided some details on how LDA works in our Methods:

"LDA is a supervised dimensionality reduction technique that finds axes maximizing the linear separability of data based on class labels[ref1]. LDA computes eigenvectors (axes) corresponding to the eigenvalues of the matrix $(S_W)^{-1} S_B$ where S_W is the within-class scatter matrix, representing the variance within each class. S_B is the between-class scatter matrix, capturing the variance of class means relative to the global mean. However, when working with high-dimensional data, LDA can suffer from overfitting[ref2]. To address this, a regularization term is added to stabilize the computation. Instead of directly computing eigenvectors from $(S_W)^{-1} S_B$, they are now computed from $((1 - \gamma)S_W + \gamma * \text{diag}(S_W))^{-1} S_B$ where $\text{diag}()$ is the diagonal matrix operation, γ is the regularization factor between 0 and 1. (determined through cross-validation)."

Previously, we described in Methods that we used a Matlab function for performing the regularized LDA: "... fitting a cross-validated, regularized LDA model (searching for best gamma with Bayesian optimization using MATLAB's fitcdiscr function) to classify stimuli identity". This level of details is largely consistent with literature that use the same method E.g. Perl et al., "We decoded narrative conditions ('PTSD' or 'Sad' script) from multivoxel spatial patterns data using an rLDA classifier, 'fitdiscr' function in MATLAB, which shows superior performance over LDA in high-dimensional imaging data that may present issues of multicollinearity and overfitting."

[ref1] Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, 30(2), 169-190.

[ref2] Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8(1), 86-100.

Perl, O., Duek, O., Kulkarni, K.R. et al. Neural patterns differentiate traumatic from sad autobiographical memories in PTSD. *Nat Neurosci* 26, 2226–2236 (2023). <https://doi.org/10.1038/s41593-023-01483-5>

6. Lines 288-290. "Across all stimuli, unit tuning curves were more similar to themselves after 7 days (Fig. 4e) when derived jointly from temporal components 1-3 than those derived from firing rates". Again, it is not explained what the similarity metrics are on the axes here, whether they are bounded, etc.

A: We have revised the sentence to "Across all stimuli, unit tuning curves were more similar to themselves after 7 days (Fig. 4e, similarity quantified by correlation coefficient after Fisher z-transformation, unbounded) when derived jointly from temporal components 1-3 than those derived from firing rates".

7. Lines 392-394. "adding temporal components in addition to firing rate in the fixed decoder consistently reduced decoding error for all four stimuli across 8 future testing days when trained on the first 7 days ($P < 0.0089$, Extended Fig6e ...". It is very unclear what this figure shows. The y axis labels do not help. They also seem inconsistent across classes of stimuli. For some classification, apparently a decoding performance is reported, while for others, an error rate is reported. In the case of the accuracy, it is very odd that accuracy is higher for NI classification than for SG. So, I guess that, whatever this accuracy is, it is not the percentage of correct choices of the classifier. So, I wonder why the authors did make it so hard to understand this decoding analysis and so difficult to compare it across stimulus sets. Cannot they simply report (as customary in this type of analyses) the classification accuracy of the classifier (with both firing rates and temporal components) as the fraction (or percentage) of correct choices? Putting a dashed line indicating the chance level? And also reporting the performance of the classifier with shuffled labels as a reference? I urge them to do so, otherwise the reader cannot really understand and judge the solidity of their claims.

A: i) We believe that the decoding accuracy is higher for NI classification than SG may be attributed to more neurons being tuned to NI than SG. This was shown in both our result (Extended Fig.2a) and in other literature, de Vries et al., 2020 <https://www.nature.com/articles/s41593-019-0550-9#Fig2> (Panel h). NI decoders thus have more information available than that of SG. Yet our goal is not to compare decoding result between different stimuli but to show that temporal code consistently improves stimulus decoding across future days.

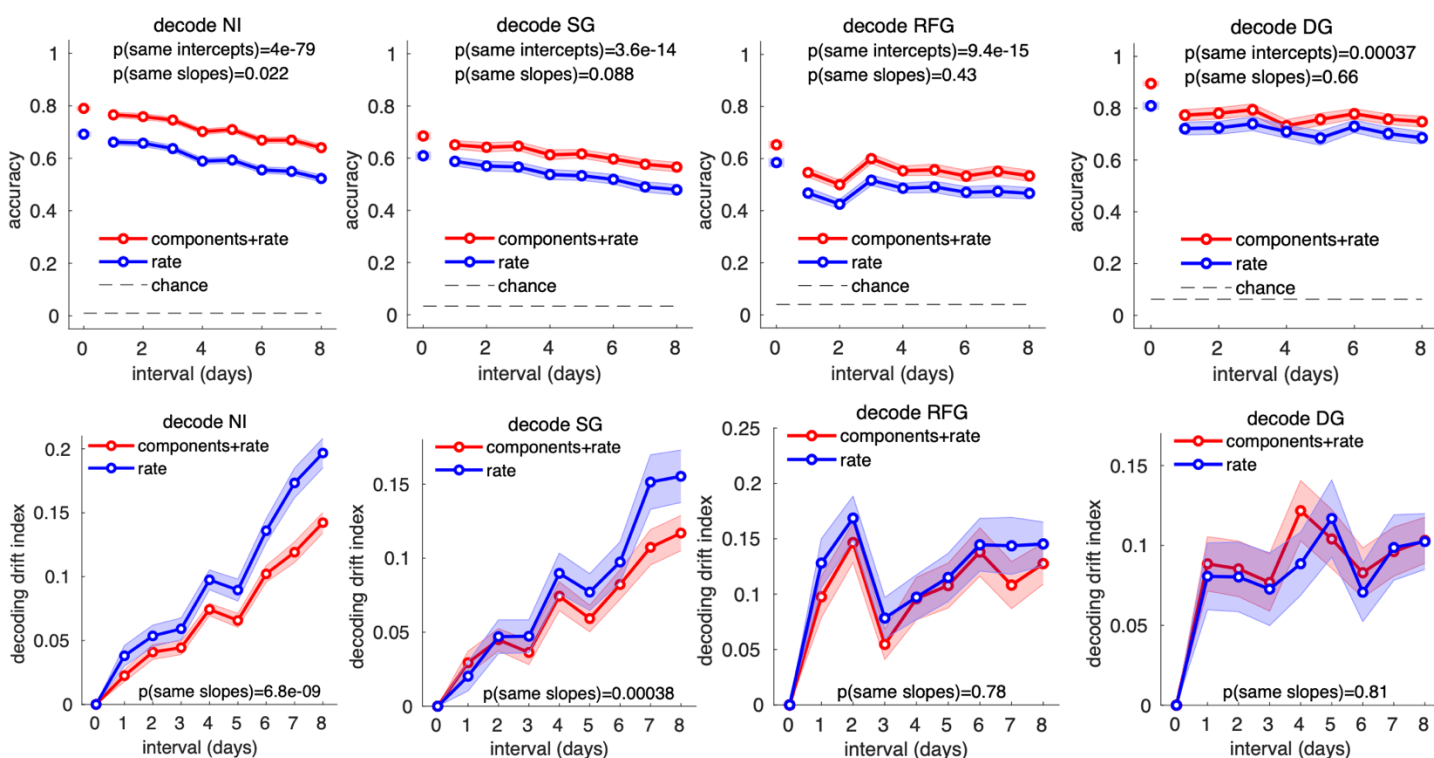
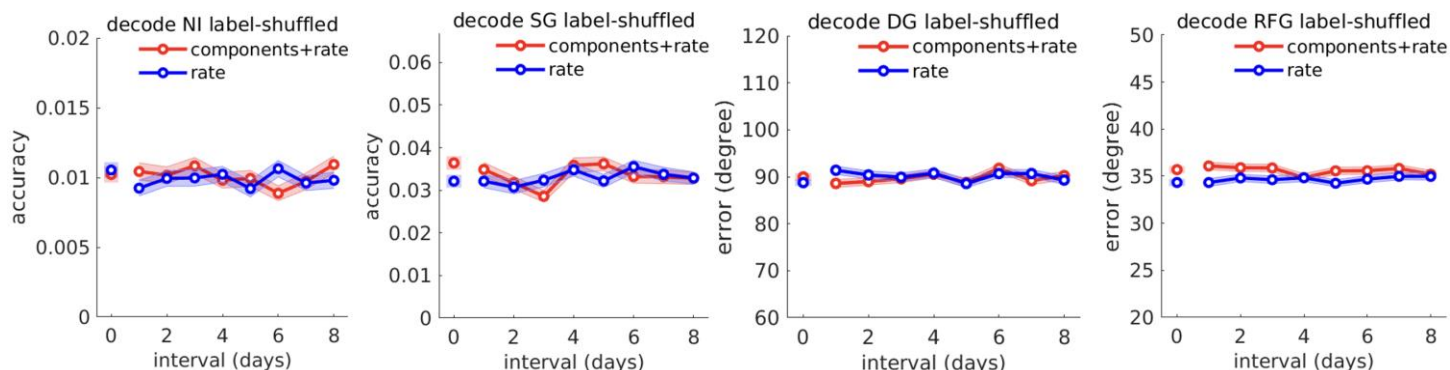
ii) Regarding the inconsistency in decoding performance metrics, since DG and RFG are parametric stimuli where decoding errors carry clear meaning, we used error as the performance metric. Another benefit of using error is that it better captures subtle differences in decoding performance compared to accuracy and hence could better capture drift in stimulus decoding. (E.g. when the true angle is 0 degree for DG, a predicted angle of 25 vs 180 degrees means very differently in terms of drift in stimulus decoding. However, since both predictions are wrong, they decrease the accuracy calculation by the same amount, obscuring the possible differences in drift across days). For SG and NI however, it's hard to define error: SGs are controlled by spatial frequencies and orientation which carry different units, NIs were not generated from parameters. We have thus selected an accuracy measure (after normalizing by chance level) to report, which is similar to Jeon et al., 2022 <https://elifesciences.org/articles/80361> (their Fig3).

iii) We reported these theoretically calculated chance level in the legend of Extended Fig.5. We assumed all types of error (pairs of true vs predicted labels) happen completely at random with equal chance. In response to your comment, here we also present the label-shuffled decoding result as a reviewer figure. The plot below shows that the empirically derived chance level is similar to the theoretical calculation.

iv) We understand your point is also very important. We have now included a separate plot according to your guidance in our Supplementary Fig 5c (below).

de Vries, S.E.J., Lecoq, J.A., Buice, M.A. et al. A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nat Neurosci* 23, 138–151 (2020). <https://doi.org/10.1038/s41593-019-0550-9>

Jeon, B. B., Fuchs, T., Chase, S. M., & Kuhlman, S. J. (2022). Visual experience has opposing influences on the quality of stimulus representation in adult primary visual cortex. *Elife*, 11, e80361.



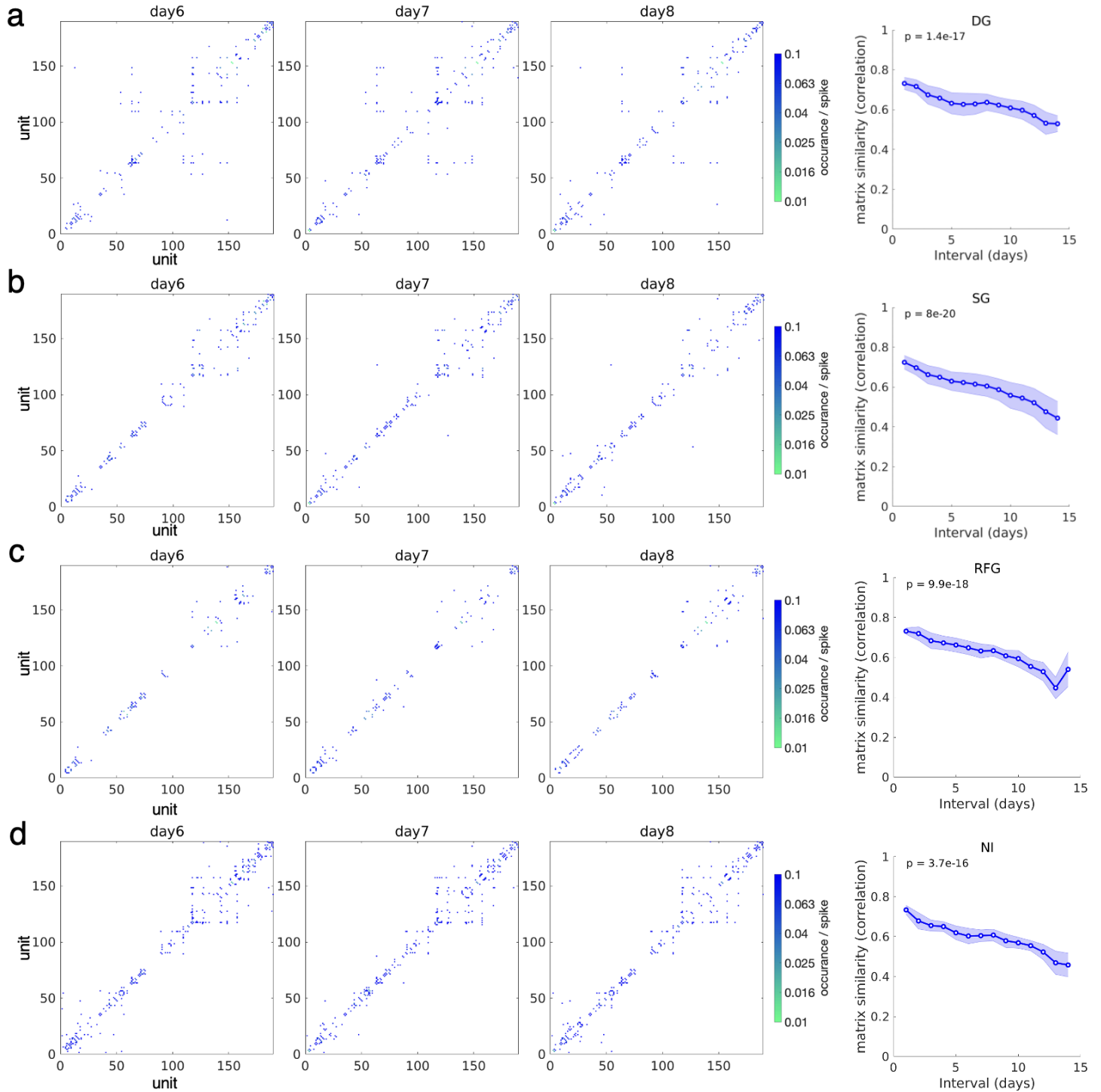
Legend: Same decoding analysis as in Extended Fig. 6e, but with all decoding quantification reported in terms of accuracy, decoding drift index calculated from accuracy.

8. Fig. 6b. These plots are esthetically nice but hard to parse, if one has to judge the stability of the strength of the functional correlations across days. Also, they do not lean themselves to an easy way to quantify how stable these patterns of functional correlations are. The quantification done in Fig. 6c and d is very rough. Simply checking that the number of significantly correlated pairs is stable or that the magnitude of the correlation is stable does not quantify if it is the same pairs that remain connected over time. So here is my suggestion.

Instead of the plots shown in Fig. 6a,b, the authors should simply report a squared correlation matrix, where the intensity of the color codes the intensity of the correlation between each pair of units. This matrix can be made across days, making it easier to look for similarities of the connection patterns over time. More importantly, the authors can then compute the correlation between these correlation matrixes as a measure of the stability of the connection patterns across days. E.g., they can compute the correlation between day 6 and 7, day 6 and 8, day 6 and 9, etc, and produce a line plot like the one in c, d.

A: Thanks for the great suggestion. These plots have been added as a supplementary figure accompanying figure6.

"Within an animal, the connections showed stimulus-specific patterns that were visually similar across adjacent days (Fig. 6b, Extended Fig7a, [Supplementary Figure 6](#)). Quantitatively, we observed indications of overall network stability: the number of connections ($P > 0.11$) and average peak amplitude per animal ($P > 0.39$) remained stable for each stimulus, except that the peak amplitude during drifting gratings had a marginal rate of change at $-8.5\text{e-}5/\text{day}$ ($P = 0.005$) (Fig. 6c-d, Extended Fig7b-c). Despite the overall network stability, we also observed gradual and subtle drift in the connectivity matrix similarity across days for all 4 stimuli. $P < 3.7\text{e-}16$ ([Supplementary Figure 6](#))"



Legend: Supplementary Figure 6: Stability of functional neuronal network **a-d**. Left: All significant CCGs between neurons in one mouse during the 4 stimuli (DG(a) , SG(b), RFG(c), NI(d)) showed similar connection patterns across days. Neurons were sorted by depth locations along each electrode shank and appended across all shanks in one animal. The neuron order was fixed across days for each stimulus type. Only neurons that had at least 1 significant connection during either DG or NI stimuli across 15 days were included ($n=189$). Right: similarity (correlation coefficient) of connectivity matrices dropped gradually across days. Linear mixed effect, duration as fixed effect, individual animals as random effect. Lines are in population mean $\pm$ s.e.m.

9. Fig. 6f. It is hard to understand what is being plotted here. What does “attendance” mean exactly in the figure? How can the range between the first and last seen session also be expressed as a percentage/fraction? More importantly, it is not clear if all these analyses refer to pairs that were first detected on day 1, as it should be done, or if neurons first detected in later days were also included. Fig. 6g is equally hard to understand. What is showing exactly. The fraction of pairs lost in a given day, compared to the previous day? Or compared to the first day? This should be stated clearly.

A: We noted in Line 425 (original submission) that Supplementary Fig6 was also used to further explain the calculation of Fig6f. Supplementary Fig6 contained unit counts before the conversion into tracked portion. We have now changed the word “attendance” to “appear” such Fig6f is consistent with Supplementary Fig6 and Fig1e. Fig6f tracked individual pairs of connections rather than individual units, but the calculation and presentation remain similar. In the legend of Fig1e, we explained that “Number of tracked single units for certain intervals based on 3 criteria: total number of sessions appeared (appear, blue), consecutive sessions appeared (consecutive, black), and the day range between the first and last seen session (range, red). The purple dashed line indicates the maximal possible range for a given unit (e.g., a unit first appearing on day 14 can only be tracked for a maximum day range of two days).”

To provide further clarity, we have added a more concrete example in the Supplemental Methods section:

For example, if a neuron missed day1, day6 and day15 but appeared in all other days, then:

a) The range of this neuron was 13 days since the neuron’s first and last appearance was on day2 and day14 respectively: $14-2+1=13$.

b) The range bound of this neuron was 14 days, since the neuron first appeared on day2, the maximal possible tracking range this neuron was $15-2+1=14$ days.

c) The appearance/attendance of this neuron was 12 days, since the neuron missed a total of 3 days: $15-3=12$.

d) The (maximal) consecutive (appeared) days of this neuron was 8 days (from day7 to day14), while shorter segment of consecutive appearance (e.g., 4 days, from day2 to day5) were disregarded.

The four colored lines in Fig1e and Supplementary Fig6 represented a cumulative histogram of the number of the neurons that at least satisfy any given longevity criterion. Finally, Fig6f presented the probability of tracking a neuron for given amount of duration. Conceptually, it’s the day-wise division of the other 3 curves against the range bound curve. Formally, it is calculated by dividing the number of neurons that satisfy a given longevity criterion by the total number of neurons that could have been tracked for that long (range bound). For example, a neuron first appearing on day 14 (range bound = 2) is only included in the calculation for P(tracked for 1 day) and P(tracked for 2 days) but is not counted in the numerator for P(tracked for 10 days). Fig6f tracked individual pairs of connections rather than individual units, but the calculation and presentation remain similar, with the exception that the range bound of a appear were decided based on the range between the first and last co-appearing day of the two units forming the pair. Fig6g represents the expected fraction of non-trackable (lost) pairs after N future days, normalized by the elapsed days N. Fig6g (portion lost) is inherently related to Fig6f (portion tracked), it’s defined as $(1-\text{portion tracked})/\text{elapsed days}$. For the best comparison with past literature, our tracked portion analysis is similar to those in Chung et al., 2019, (their Fig4f)

Chung, J. E., Joo, H. R., Fan, J. L., Liu, D. F., Barnett, A. H., Chen, S., ... & Frank, L. M. (2019). High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays. *Neuron*, 101(1), 21-31.
[https://www.cell.com/neuron/fulltext/S0896-6273\(18\)30993-0](https://www.cell.com/neuron/fulltext/S0896-6273(18)30993-0)

10. The authors wrote that the electrodes were implanted in two different visual areas, V1 and LM, which are known to be in a hierarchical relationship – e.g. see (Wang and Burkhalter, 2007; Wang et al., 2012). However, these areas are never directly compared in any of the reported results. This raises the question of why both areas were considered in the first place, as why their respective roles and differences are not addressed or discussed.

A: We previously explained (Line 859, original submission) our motivation to implant in V1 and LM. This is due to the report (Deitch et al., 2021) (their Fig2) that LM and V1 were among the most stable regions in terms of within-day tuning stability. However, since we only targeted 2 out of the 9 visual sub-regions with a disproportionate 18 probes targeting V1 and 7 probes targeting LM, we decided not to focus our discussion in systematically comparing the stability between different visual areas, we have stated in limitation (previously Line 613) that the stability of other visual areas are to be studied in the future. However, we understand this an important point, in response to your comment, here we add another reviewer figure (this journal publishes all reviewer report after 2022) below showing the representational drift in both LM and V1. The plot shows that temporal code was more stable than rate code in both LM and V1. These effects held true for all 4 stimuli: DG, SG, RFG and NI. Finally, we have now stated in the

limitation section that “We observed that temporal code was more stable than rate code in both LM and V1 (data not shown).”

FIGURE REDACTED

MINOR CONCERNS

1. legend of Fig. 1j. “Neuron identity shuffled decoder approached chance ...”. This sentence is misleading because it seems that all neurons have been shuffled, while, based on the main text, only 10% per session were shuffled. Please specify here that only 10% of cells were shuffled.

A: We have now specified in the legend that “10% of cells were shuffled per session, resulting in an expected $(0.9)^{\text{day}}$ fraction of unshuffled neurons at any given day.”

2. Lines 143-144. “... we observed visually consistent day-to-day most preferred. Conversely, we also detected longitudinal drifts in units’ tuning to less preferred images (Fig. 1i, zoomed-in view)”. This does not seem a fair description of what shown in the zoomed-in views. It seems to me that what is changing over time in several neurons is the response to the originally preferred stimulus, which ceases to be such after 7 days. This does not happen for all the units, but some of them seem to actually loose their tuning.

A: Agreed. We have now changed the text to “Conversely, we also detected longitudinal drifts in the tuning, including some units losing their tuning to the most preferred images (Fig. 1i, zoomed-in view).”

3. Lines 151-152. "We observed similar decoding error indexed stability and drift for all other stimulus types (Supplementary Fig2)." Not clear which stimuli the authors are referring to here.

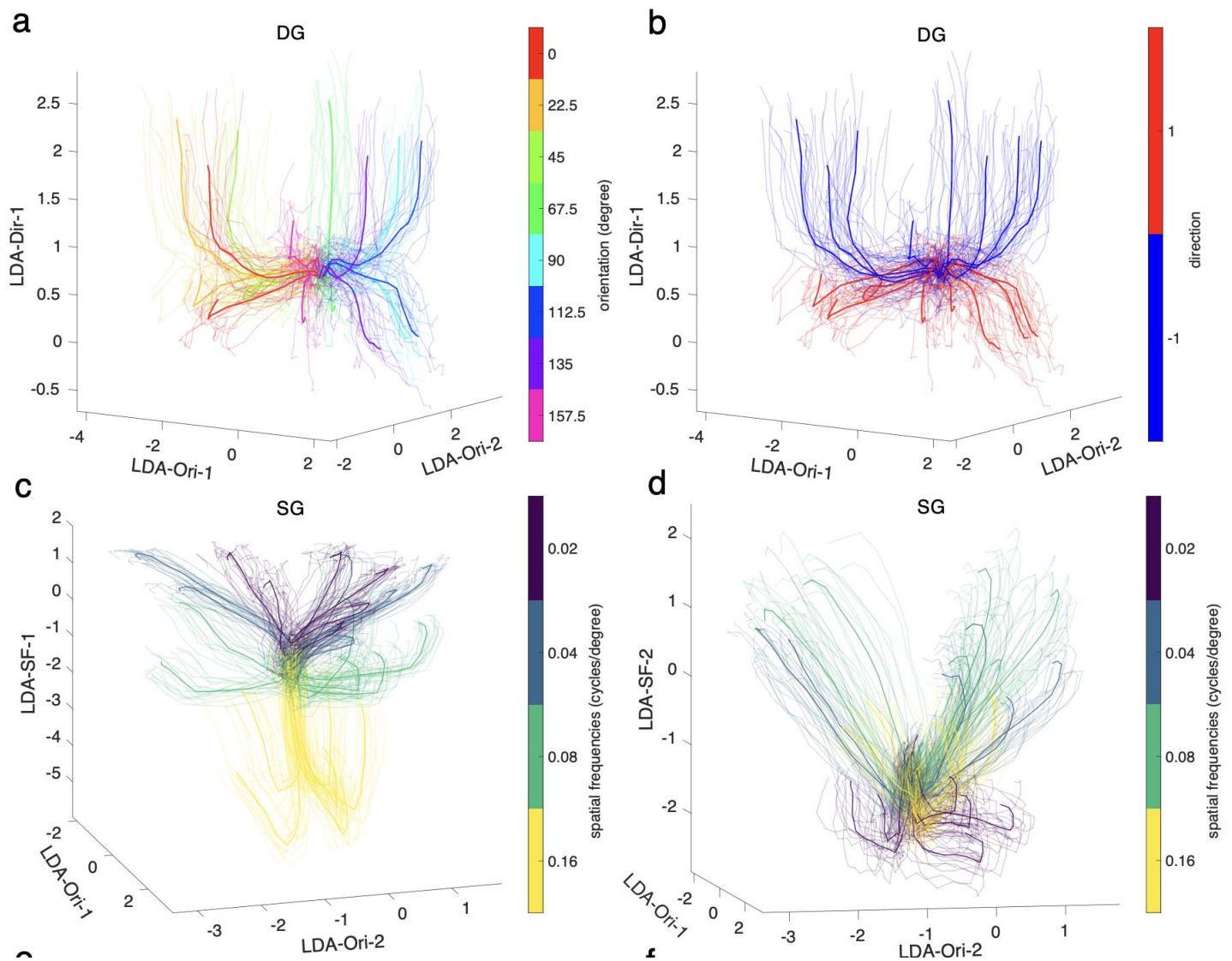
A: We have now improved this by stating "We observed similar decoding error indexed stability and drift for all other stimulus types (DG,SG,RFG stimuli, Supplementary Fig2).

4. Lines 199-201. "Neurons exhibited diverse temporal response profiles to an example stimulus, showing 0° drifting grating evoked daily averaged, smoothed firing time courses from example neurons pooled from 5 mice over 7 different experiment days." The writing of this sentence is truly terrible! The result is that it is hard to understand what is it that the figure actually shows. Is it the response to a single, specific stimulus (the 0° DG?) of a population of neurons across consecutive days?

A: Yes. We have now changed the sentence to "The stimulus-evoked spiking time course to a single, specific stimulus (i.e. 0° drifting grating) of a population of neurons across multiple days. Neurons were pooled from 5 mice over 7 different experiment days"

5. Lines 234-236. "Similarly, the four spatial frequencies of static gratings stimuli were established in four parallel planes, where the six orientations evolved radially outward forming a hexagonal formation (Fig. 3c, Supplementary Fig3b, Supplementary Movie 2)." The figure does not really allow to distinguish easily the traces referring to different SFs. The dashed lines connecting the traces to the images of the gratings do not help much. The authors could perhaps try to encircle those traces referring to the same SF and connecting the circle to the image of the grating with the corresponding SF.

A: Thanks for the suggestion. We have now included additional supplementary figures and supplementary movies to further illustrate our point. They have the same traces as in Fig3b-c but color-coded differently. The panel c and d below were color-coded using spatial frequencies. Collectively they show that stimulus of different spatial frequencies could be clearly separated with the first two spatial frequency LDA encoding axes.



Supplementary Fig3: Additional visualization on temporally resolved populational representation. **a-b.** Disentangled population representation dynamics during drifting gratings from an example animal. The plot shows daily trial-averaged (lighter lines) and 15-day-averaged (darker lines) traces, projected onto the first two linear discriminant orientation coding dimensions and the first direction coding axis. Orientations (a) or directions (b) were used to color-code the traces. **c-d.** Same as in (a) but for static grating stimuli, projected onto the first two orientation coding dimensions and the first (c) or second (d) spatial frequency coding axis. Stimulus spatial frequencies were coded in color.

6. Lines 296-297. Another very convoluted and odd sentence "Consistent with the finding in firing rate-based tuning (Fig. 2b), stability in temporal components defined tuning was also associated with component tuning reliability, strength, and selectivity." This is very hard to interpret. Please revise it.

A: We have revised this into "Figure 2b showed that the stability of single-neuron firing rate-based tuning curve was associated with tuning reliability, strength, and selectivity. Similarly, these associations held true when stability, reliability, strength, and selectivity were derived from tuning curves of temporal components."

7. Lines 310-312. "Components increased stimuli separation than firing rates, showing an example neuron's daily trial-averaged firing rates (Left) and first 3 components (Right) distribution (n=15 days) under 16 different stimuli." The writing here is terrible ... please revise it making sure a subject precedes a verb and then a complement or whatever follows. But please make things clearer.

A: We have revised this into "Components provided greater separation between stimuli compared to firing rates. This is illustrated by an example neuron's daily trial-averaged firing rates (left) and the distribution of the first three components (right) across 15 days under 16 different stimuli."

8. Lines 316-318. "Across all stimuli, temporal components-based unit tuning curves were more similar to themselves after 7 days than that of firing rate-based. Paired t- test." Be more specific about what is plotted. Do the dots refer to the neurons in the population? I guess so, but it should be explicitly reported along with the number of units being shown.

A: We have improved this sentence by adding "dots represent neurons in the population, n = 481, 533, 455, 684 stimulus-tuned single units that reappeared after 7 days for the 4 stimuli DG,SG,RFG,NI respectively."

REFERENCES CITED IN THE REVIEW

Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33:111–120.

Piasini E, Soltuzu L, Muratore P, Caramellino R, Vinken K, Op de Beeck H, Balasubramanian V, Zoccolan D (2021) Temporal stability of stimulus representation increases along rodent visual cortical hierarchies. Nat Commun 12:4448.

Rudelt L, Marx DG, Spitzner FP, Cramer B, Zierenberg J, Priesemann V (2024) Signatures of hierarchical temporal processing in the mouse visual system. PLOS Comput Biol 20:e1012355.

Wang Q, Burkhalter A (2007) Area map of mouse visual cortex. J Comp Neurol 502:339–357.

Wang Q, Sporns O, Burkhalter A (2012) Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. J Neurosci 32:4386–4399.

Reviewer #4 (Remarks to the Author):

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

A: Thank you for your effort reviewing our manuscript.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have done a great job addressing all the points raised.

[Reply: Thank you for you great effort reviewing and improving our manuscript.](#)

Reviewer #2 (Remarks to the Author):

The authors have addressed all my comments.

[Reply: Thank you for you great effort reviewing and improving our manuscript.](#)

Reviewer #3 (Remarks to the Author):

I am satisfied with the revision of the manuscript and I thank the authors for having added many of the descriptions and clarifications I felt were missing in the previous version of their manuscript.

[Reply: Thank you for you great effort reviewing and improving our manuscript.](#)

Reviewer #4 (Remarks to the Author):

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

[Reply: Thank you for you great effort reviewing and improving our manuscript.](#)

Reviewer #5 (Remarks to the Author):

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

[Reply: Thank you for you great effort reviewing and improving our manuscript.](#)