nature portfolio

Peer Review File

Neural correlates of perisaccadic visual mislocalization in extrastriate cortex

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this paper titled “Neural correlates of perisaccadic visual mislocalization in extrastriate cortex”, Weng and coauthors, have combined single unit neural recording in macaque monkeys with computational modeling in order to provide a systematic understanding of the perisaccadic events in the visual areas, middle temporal and V4. The authors have utilized a basic visually guided saccade task and during the task, bombarded the spatiotemporal visual field with a sequence of small visual probes while recording the neural responses in the extrastriate cortex. They then, used this data to build 4 dimensional spatiotemporal units and spatiotemporal kernels for the recorded neurons. Using these blocks, the authors built a computational structure that is used to quantitatively predict perisaccadic spatial bias.

The study is meticulously designed and performed. The central question of the paper is of utmost importance in the field, and I very much enjoyed reading it. I think the paper is well suited for a high rank journal with general readership like Nature Communications as the contents will be of interest to a wide range of readers from theoretical neuroscientists to experimentalists beyond the field of nonhuman primate physiology. In my view the paper is very close to its presentable form, although, following, I have only a few comments for improving the manuscript.

Major comments:

- The fixation point disappears randomly in an interwall between 700 to 1100 ms from its onset. At the beginning of this 400 ms period, the monkey faces a large degree of uncertainty about when to saccade. On the other hand, towards the end of this period, the monkey is getting more and more certain about the saccade time by the passage of each millisecond. Does this certainty affect the spatial bias? Specifically, if we compare the first 150ms of this period with its last 150ms, is there any difference in the neural measures of bias? There are many ways to shed light on this, I leave it to the authors to pick their torch.

- The current study provides a wonderful and systematic insight into the neural events happening in the extrastriate cortex. In that sense the paper is a first of its kind and that’s why it deserves to be published in Nature Communications. However, in order to fully explain the perceptual spatial bias two more steps need to be taken and the Discussion section should elaborate more on them. Firstly, the current paper does not provide any direct psychophysical measure from the monkey and

it only qualitatively explains the previous psychophysical results (that are contradicting in some cases). In order to put the paper's quantitative hypothesis to test, this should ideally be tested against direct behavioral measures of bias from the monkeys whose neurons are being recorded. Only then, the quantitative model could be validated against trial based quantitative data. Secondly, even then, the link between the underlying neural function to perception would be correlational and causal experiments would need to follow in order to close the loop and establish a mechanistic link between neurons and perceptual bias. Adding a discussion about these points would put the current paper in a bigger context and provide a sense of future direction.

Minor comment:

- Figure 1.b attempts to communicate the 4 dimensional nature of a SMU by putting a time axis on the bottom of a cube and linking them with a curved arrow. I had a hard time noticing what the figure is graphically communicating, I think it can be improved. Perhaps by having multiple cubes (all faded, only one highlighted) fanning out of the time axis?

Reviewer #2 (Remarks to the Author):

Review of "Neural correlates of perisaccadic visual mislocalization in extrastriate cortex"

Through a sophisticated analysis of spiking data from monkey MT and V4, this paper suggests that receptive fields are biased toward the saccadic target at/near the time of the saccade, and that the greatest contribution to distorted localizations in behavior likely stems from neurons with receptive fields far from the saccadic target. Finally, these RF changes appear to sacrifice veridicality of localization for increased sensitivity near the saccadic target.

This paper is far from what I do, but I do have interest in spatial coding in and around saccades (and have published on integration across saccades) and have followed some of the literature on spatial remapping at the time of saccades. Perhaps as a result, I found thie a VERY difficult paper to parse and found myself rereading sections of the paper multiple times, bouncing between the main paper, the Methods supplement and the Supplement in order to figure out what was meant. Maybe other readers will have the same problem, so I'm going to try to explain why I had such a hard time, in the hopes that it will help you clarify the paper for a broader audience.

First, my problem started with my strong expectations. From the abstract, this is a paper about representational bias and "response remapping". That led me to expect to see something about mapped spatial receptive fields and how they change near a saccade. Once I understood that you were attempting to estimate a full 4-d kernel for each neuron (probability of a spike at time t+\tau in response to a stimulus at time t relative to the saccade in position x/y), I assumed I'd eventually be shown RF maps (or RF centroids) and how they change across time. That's pretty much never analyzed per se and never plotted. I also expected at least some verbiage about what we mean by remapping. Normally, one thinks about remapping as a change in RF plotted in retinotopic coordinates. Yet, this entire paper discusses stimuli and RF characteristics in screen coordinates both before and after a large saccade (which I find vexing to think about). In the older literature (on V1/V2) that I know about such things, the idea is that just before a saccade a neuron might respond to the stimulus that will (post-saccade) land in the RF, i.e., that its RF now includes (by extending the RF or adding a 2nd lobe) a region shifted in the direction of the saccade. Or the "compression" story would have an RF be biased in the direction of the saccade pre-saccadically, and in the opposite direction afterward (for an RF the is now located on the other side of the saccadic target). It's very weird, to me, to think of all of this in screen coordinates, and that definitely tripped me up.

What follows are a bunch of other things that confused me on first reading, but for many of which I eventually figured it out after much work. I guess my point is that you should make it possible for a casual reader to just read from the beginning forward and get the gist of what you did and the analysis, without having to jump to Methods and Supplement in order to get anything out of the main text.

Specifics by line number

* 85: "relative to THAT timepoint": At first I wasn't sure what this sentence meant, so it might be good to phrase it more transparently, especially since space is barely mentioned: we estimated a spatiotemporal kernel for each neuron, i.e., if a stimulus appeared at position (x,y) at time t, we estimated how much that would increase the probability of firing at time t+\tau (ok, ok, this is not precisely true, but it's close).

* 123: "multiple basis functions": You have to go deep into the supplement to find out what this phrase refers to. The number of basis functions, initially, is the same as the raw data (for each x/y it's equal to the number of 7 ms stimulus time steps times the number of 7 ms delays you analyze). That is, before you drop "unused" basis functions (the "sparse" SVGLM), the basis functions only serve to allow you to interpolate a smooth kernel between discrete samples/bins. Anyway, this paragraph confused me because I wasn't sure what was an STU. Is it the 4-d data structure (function of x, y, t and \tau)? I thought saw, but then you say Figure 1c shows the "STUs" (i.e., plural)

at an example x/y. To me, that's just a 1-d slice through the single 4-d STU. An STU is "represented" by the weights, but isn't the STU still a 4-d matrix (the sum of the sparse basis functions you leave in, scaled by the weights)? Then the STUs "comprise kernels ... at that location". So is an STU just the 2-d kernel for a specific x/y, or is it the whole thing?

* 131: You say "convolved with", but then the kernels should be, indeed, kernels, not impulse response functon(i.e., flipped). I'm sure you don't mean that. And convolve isn't quite right either, because it's not shift-invariant. You measure a separate impulse response for every x, y and, most importantly, t.

* 143-144: This is the first hint that you aren't going to show me shifting/distorting retinotopic-coordinates RF maps. And, you aren't going to assume or even discuss how the population code is read out to provide a stimulus location estimate. Rather, you determine that for a give x/y/t/\tau combination, if the response across the subset of neurons with RFs at a similar location is similar to the same set of neural responses to a stimulus at location x+1, then there will be a rightward bias. In other words, near as I can make out, you are trying to make a strong statement about the predicted bias direction without saying anything about the specifics of downstream readout mechanisms. That's a bit obscure, and rather indirect logic, but I get why you are doing it. But, you don't advertise what you are doing nor why beyond this one rather telegraphic sentence. In fact, you never clearly state that your goal is to demonstrate readout bias without having to specify how the readout works.

* 146: "the representation of a probe": you might say explicitly that you mean the population response (for a particular time of stimulus and delay to looking at the response). To take this a little further, you compute your correlations (unless I'm confused) for vectors of responses across neurons for a fixed x/y and t/\tau. But, wouldn't it make some sense to compute the correlations across \tau values as well (the full temporal response to a stimulus at time t across the set of neurons to a stimulus at x/y is similar to that at x+1/y)?

* 156: Now that I'm writing the review, I must still be confused, or maybe not. The correlation coefficient (same as cosine similarity) is for a fixed x/y/t/\tau but across neurons in an ensemble (the vector length is the number of neurons in the ensemble), yes? It's not across weights on your basis functions, is it? I thought the sparse representation was summed to give a continuous response, but that continuous response was used to get the vectors (across neurons) that were correlated. This is really impossible to understand from the main text.

* 164: Again, it's a logical leap to jump from a summed set of correlations, turned into vectors, to call that sum "bias" (by which we'd usually mean a shift of an RF, or a set of RF, or a read out

location estimate). As I said above, I think I get your logic for acting like this is okay, but you don't justify it explicitly, so a casual reader will likely trip over this sentence.

* 173: I now realize that your "bias" estimate is really a relative bias between positions x and x+1 (and x-1, etc.), not an absolute bias. The overall population response could be shifted a LOT, so that a readout would get a hugely shifted estimate, but all you are noticing is that however the full population is interpreted, locally the response is similar to x as to x+1, even though the form of that population response might look overall like the stimulus is at location x+20 (yeah, yeah, you don't have an x+20).

* 176: I'm not sure how this normalization works. Presumably it's a single scalar per ensemble? If you divide by the max positive correlation, so the normalized number range up to +1, that doesn't mean they'll range all the way to -1. I assume you don't do an affine stretch because it would be silly to shift a zero correlation to be nonzero.

* 207: phenomenon -> phenomena

* 210: 60:90 is now post-saccade, right? This is the first time you really need to unpack how to think about this stuff, since you plot in x/y in screen coordinates, but the eyes have now moved, so one expects the RF in screen coordinates to shift with the length of the saccade. Clarification please.

* 227: Remind the reader here (I think it's buried in Methods somewhere) that you flip things for leftward vs. rightward saccades so that it makes sense to compare these in the same plot, i.e., it's independent of which side of the saccadic target you started on

* 235: I didn't dig into what "p < .001" meant here, but this was all about model predictions. Did the p value come from a collection of stochastic simulations of the model? That is, are you making the mistake of computing a p value for a simulation when you could have easily, and pointlessly, inflated that p value by running more simulations?

* 261: Again you are looking at postsaccadic values of t, when the eyes have moved and one expects the RF, in screen coordinates, to have moved a lot. So please clarify what we are looking at and what to expect.

* 293: FF = feed-forward? Note: I'm not a fan of having to learn a bunch of acronyms when I read a paper (STU, ST, etc.).

* 311: You have neurons from MT and V4 but never discuss any are differences! Also, you never discuss whether the biases you find might be inherited wholly from modulations of responses in upstream areas.

* Another point: your predictions of "bias" from correlations assumes that these similar population vectors (from x and x+1) drive a distorted/biased readout. That is, it assumes that the readout is blind to the effects of time relative to a saccade. It would be good to talk about that explicitly.

* This stimulus ensemble is sparse, but not "spherically symmetric" and thus if you were just doing a spike-triggered average to estimate a receptive field, that estimate would be biased (as you know from teaching MathTools!)

* 533: Another place that you might clarify that x/y is in screen coordinates, but flipped for leftward saccades and that "0:200" is looking backward in time to the stimulus from the response bin. "9 probe locations around the neuron's RF" also bumps into the issue of screen vs. retinotopic coordinates and whether flipped for leftward vs. rightward saccades.

* 615: A trial consists of one of each of the 81 probes in a random permutation?

* 621: Was the grid set up so that both leftward and rightward saccades would leave the RFs inside the grid?

* 632: You mean similar RF position but identical ST and grid position? So a session only had leftward or only had rightward saccades???

* 643: "sparse collection of STUs": This sounds like a non-deleted basis function is what you are calling a STU, which I don't think is what you mean.

* 652: "k_{x,y}(t,\tau) represent the stimulus kernels". Huh? isn't k a stimulus kernel (i.e., the 4-d thing)? You say "kernels" (plural) and treat that inline equation like its plural. I can't parse that. Do you ever say how you estimate h(\tau) and using what parameterization?

* You don't ever show temporal evolution of a spatial RF and don't show anything like a spatial RF until Figure 4

* 665: smoothed over t, over \tau or both?

* 672-673: This is a range of time t, but what range of delays \tau figure into this? Separate correlations computed for each value of \tau???

* 678: Again, this bias vector from correlations is pretty indirect, and it's not clear that it uses or ignores the overall strength of a response or the spatial RF itself except that if you are outside of the RF, responses become noise and correlations should drop, leading to low, noisy bias.

* 682: How does one see the shift of the RF in Fig. 2a-b?

* 684: This is probably the only place you allude to a postsaccadic RF, i.e., to the fact that you plot in display coordinates.

* 691-696: Give that I'm still unclear what an STU is, this section lost me entirely. I thought an STU was the full 4-d linear kernel. What's the group of items that your are using to determine the subset that are "modulating"? What does the word "prevalence" mean here (it's never defined)? Why does Eq. 2 only mention time, not space? I'm completely lost here. The same confusion applies to the units of "bias-relevant" STUs and the AUC computed from them.

* Fig. 1c: the y-axis label should be a clearly Greek \tau. This same weird "T" occurs in the running text (Supplement, 25). This figure's legend should emphasize that these are screen coordinates.

* Suppl. Eq. 1: Is B_{i,j} what you call an STU?

* Suppl. 30: The full set of B_{i,j} is as big as the data, so I don't know what you mean by binning and two orders of magnitude reduction

* Does the subset of the B_{i,j} retained depend on probe location?

Reviewer #3 (Remarks to the Author):

The authors record responses in MT and V4 to brief visual probes around the time of a saccade. They find a spatial bias in the direction opposite the saccade and enhanced sensitivity around the area of the saccade target. This is an interesting and broad investigation of perisaccadic neural response and I have only three comments.

First, the central measure is the correlation between a RFs response and its neighbors. When the correlation is higher on one side, that is labeled a "spatial bias". Here is the logic they used:

"We made the assumption that similar neural sensitivity to probes appearing at neighboring locations could create uncertainty in a readout of the stimulus location by a downstream area, which would lead to a bias in spatial perception. In other words, if during a saccade, the representation of a probe becomes similar to that of a neighboring probe, we can assume a representational bias toward that neighboring location." Lines 143-147.

This sounds plausible but it is not clear what is meant by "representational bias". How does it translate to perceived location? The bias they find is away from the saccade target, specifically RFs on the side of a cell opposite the saccade direction are responding like the cell. This could be taken as a shift of receptive fields toward the saccade target – the RF further away from the saccade target is responding as if it were closer. This shift has been reported by others. But if receptive fields shift toward the saccade target, perceived location must shift away according to classic labeled line reasoning. This is the wrong direction. The authors nevertheless claim that their results "account for changes in spatial perception during eye movements." They have not shown how this happens. Yes, the receptive fields may be shifting (showing a representational bias) toward the saccade target but that does not account for a shift in location toward the saccade target. It predicts the opposite. Perhaps the explanation is there somewhere in the text but I couldn't find it. The authors need to make a clear link between their spatial bias and perceive location if they wish to claim that they account for the psychophysical results of perisaccadic location perception.

Second, the authors present highly processed results in terms of normalized arbitrary units. These show a spatial bias in the expected direction at a reasonable time widow after saccade onset. However, we have no idea whether the scale of the effect is in the order expected for perisaccadic mislocalization. Perhaps 0.2 arbitrary units is a very small shift and if so that would suggest that these anatomical areas are showing only initial hints of the large displacements that may come later in the visual system. Please do some reverse modeling and convert the spatial bias to actual degrees of visual angle so we can decide if this is the site where the shifts arise or not. Also, and again arguing against a relation to perceived shifts, the results show no shifts before the time of the saccade even though behavioral results do.

Finally, the authors record from MT and V4 but do not give any breakdown or summary of the responses for the two areas.

Reviewer #4 (Remarks to the Author):

Weng et al. use a previously established rigorous encoding modeling framework to characterize the spatiotemporal kernel of neural responses to visual probes flashed at various times relative to horizontal saccades. The encoding model developed by the same authors previously is robust and statistically sound. However, there is a major concern in the interpretation of what the authors label as spatial bias. I outline them point-wise below. They must be addressed before the paper is ready for publication.

Major (conceptual):

The spatial bias is measured in terms of correlations between responses to probes at a central location and probes at areas surrounding the central location. Throughout the paper, the authors choose probes near the ST as the central location. Furthermore, the correlations are projected onto the horizontal axis – the axis of the saccade direction. Based on these metrics, the authors find a spatial bias in the neural representation around the time of a saccade and that the bias is in the direction opposite to the saccade. The authors are attempting to link the neural representation of visual space to spatial perception around the time of saccades, but their specific goal of linking ST remapping to the peri-saccadic compression effect is not supported by the results. I find the results compatible with FF remapping because the direction of spatial bias is unidirectional as opposed to omnidirectional, as you'd expect with ST target remapping. Because the authors only look at the horizontal projection of spatial bias, it is difficult to dissociate whether the bias they report is due to FF remapping or ST remapping. At face value, it looks like FF remapping, but this could be due to the choice of horizontal projection of the data. Any statement on ST remapping must be assessed with a parallel analysis of the vertical projection of spatial bias of neuron ensemble. Since the probes were not limited to horizontal direction, this should be possible with the existing dataset.

For near probe vs. far probe analysis – again, the results are compatible with FF remapping. The authors should compare responses to the three probes along the vertical axis to assess whether the responses are compressed towards the ST.

Line 287-289: I don't understand the description of the sensitivity index. As a result, I don't understand Fig 6 and don't understand the motivation clearly. Spatial bias and sensitivity are quantified from the same dataset, and it has a feel of double dipping. But this could be clarified in the main text and in methods.

To truly assess the perceptual bias, monkeys could be trained to report the probe locations. Such an experiment would make the paper exponentially better and behaviorally relevant. However, I understand that the behavioral experiment is out of the scope of this study. Nonetheless, I encourage the authors to pursue this path given their excellent modeling tools at hand. More broadly, the question of readout is crucial to correctly interpret the implication of RF remapping, namely whether the downstream decoder is aware or unaware of RF shifts. Mickey Goldberg and colleagues have recently written about this. The authors' argument would be more understandable if they reframed their paper in terms of the language in Goldberg's paper:

Tuning curves vs. population responses, and perceptual consequences of receptive-field remapping. Ning Qian, Michael E. Goldberg, and Mingsha Zhang. Frontiers in Computational Neuroscience (2022). DOI: 10.3389/fncom.2022.1060757

Major (technical):

Can the authors compare the V4 and MT populations and compare the % of neurons with spatial bias?

Keeping the probe locations constant, can the authors compare trials in opposite saccade directions? The bias should also have opposite directions. In other words, can the authors disaggregate the saccade directions instead of combining all saccade directions? Even though this will sacrifice statistical power, it will make the case more convincing.

Minor:

Fig 1d vs 1e: it is confusing to look at these two figures side-by-side because it gives the impression that they represent the same probe cluster, but 1e is around ST, and 1d is around RF. Please label them clearly on the respective figures.

Fig 2c is missing the color bar.

March 13, 2024

Dear reviewers,

We have revised our manuscript NCOMMS-23-31269, entitled “Neural correlates of perisaccadic visual mislocalization in extrastriate cortex”, based on the reviewers’ comments. We thank the reviewers for their feedback, as these revisions have made the manuscript clearer and more thorough. We have added new analysis and figures, expanded the scope of the study to include behavioral results, and clarified numerous methodological details. Point by point responses to each of the reviewer’s concerns are included below.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this paper titled “Neural correlates of perisaccadic visual mislocalization in extrastriate cortex”, Weng and coauthors, have combined single unit neural recording in macaque monkeys with computational modeling in order to provide a systematic understanding of the perisaccadic events in the visual areas, middle temporal and V4. The authors have utilized a basic visually guided saccade task and during the task, bombarded the spatiotemporal visual field with a sequence of small visual probes while recording the neural responses in the extrastriate cortex. They then, used this data to build 4 dimensional spatiotemporal units and spatiotemporal kernels for the recorded neurons. Using these blocks, the authors built a computational structure that is used to quantitatively predict perisaccadic spatial bias.

The study is meticulously designed and performed. The central question of the paper is of utmost importance in the field, and I very much enjoyed reading it. I think the paper is well suited for a high rank journal with general readership like Nature Communications as the contents will be of interest to a wide range of readers from theoretical neuroscientists to experimentalists beyond the field of nonhuman primate physiology. In my view the paper is very close to its presentable form, although, following, I have only a few comments for improving the manuscript.

Major comments:

- The fixation point disappears randomly in an interwall between 700 to 1100 ms from its onset. At the beginning of this 400 ms period, the monkey faces a large degree of uncertainty about when to saccade. On the other hand, towards the end of this period, the monkey is getting more and more certain about the saccade time by the passage of each millisecond. Does this certainty affect the spatial bias? Specifically, if we compare the first 150ms of this period with its last 150ms, is there any difference in the neural measures of bias? There are many ways to shed light on this, I leave it to the authors to pick their torch.

Our kernels were organized according to time of response relative to saccade onset and stimulus presentation. We did not take into account the effect of uncertainty and have

sorted the trials without marking the different lengths of fixation, at the earliest stage in the data processing. We apologize that we cannot separate the trials based on the fixation point disappearance time unless we go back to our raw data and fit the model separately for these sets of trials. For our newly added behavioral results, the duration of fixation was fixed, but we compared the localization errors for slow and fast reaction times (Supplementary Fig. 1a). We divided the trials in each session by the median reaction time of the monkey, and found no difference between the two. This at least suggests that mislocalization is tied to saccade execution rather than to factors which modulate the reaction time (arousal, strength of motor planning, etc).

- The current study provides a wonderful and systematic insight into the neural events happening in the extrastriate cortex. In that sense the paper is a first of its kind and that's why it deserves to be published in Nature Communications. However, in order to fully explain the perceptual spatial bias two more steps need to be taken and the Discussion section should elaborate more on them. Firstly, the current paper does not provide any direct psychophysical measure from the monkey and it only qualitatively explains the previous psychophysical results (that are contradicting in some cases). In order to put the paper's quantitative hypothesis to test, this should ideally be tested against direct behavioral measures of bias from the monkeys whose neurons are being recorded. Only then, the quantitative model could be validated against trial based quantitative data. Secondly, even then, the link between the underlying neural function to perception would be correlational and causal experiments would need to follow in order to close the loop and establish a mechanistic link between neurons and perceptual bias. Adding a discussion about these points would put the current paper in a bigger context and provide a sense of future direction.

We have added direct psychophysical measurements of mislocalization from two monkeys as shown in the new Fig. 1. We observed mislocalization in the direction opposite to saccade direction for stimuli presented around -100:0 ms prior to saccade onset. The spatial bias computed from kernels is most negative at around 50 ms from saccade onset and 50:100 ms delay relative to that time, which translates to around -50:0 ms for time of stimulus presentation, similar to what we observed behaviorally.

Simultaneous behavioral data and recordings are certainly part of our future plans. The reviewer is correct that manipulations of the computational model cannot substitute for causal experiments. We have now mentioned these points at the end of discussion section in line 434-436.

"Our results provide a potential explanation of the neural basis of mislocalization, which could be tested most definitively through experiments combining psychophysical measurements in macaques with causal manipulations of neural activity."

Minor comment:

- Figure 1.b attempts to communicate the 4 dimensional nature of a SMU by putting a time axis on the bottom of a cube and linking them with a curved arrow. I had a hard time noticing what the figure is graphically communicating, I think it can be improved. Perhaps by having multiple cubes (all faded, only one highlighted) fanning out of the time axis?

We have updated the figure (now Fig. 2b) to clarify that there is a cube at each time from saccade onset.

Reviewer #2 (Remarks to the Author):

Review of "Neural correlates of perisaccadic visual mislocalization in extrastriate cortex"

Through a sophisticated analysis of spiking data from monkey MT and V4, this paper suggests that receptive fields are biased toward the saccadic target at/near the time of the saccade, and that the greatest contribution to distorted localizations in behavior likely stems from neurons with receptive fields far from the saccadic target. Finally, these RF changes appear to sacrifice veridicality of localization for increased sensitivity near the saccadic target.

This paper is far from what I do, but I do have interest in spatial coding in and around saccades (and have published on integration across saccades) and have followed some of the literature on spatial remapping at the time of saccades. Perhaps as a result, I found thie a VERY difficult paper to parse and found myself rereading sections of the paper multiple times, bouncing between the main paper, the Methods supplement and the Supplement in order to figure out what was meant. Maybe other readers will have the same problem, so I'm going to try to explain why I had such a hard time, in the hopes that it will help you clarify the paper for a broader audience.

First, my problem started with my strong expectations. From the abstract, this is a paper about representational bias and "response remapping". That led me to expect to see something about mapped spatial receptive fields and how they change near a saccade. Once I understood that you were attempting to estimate a full 4-d kernel for each neuron (probability of a spike at time t+\tau in response to a stimulus at time t relative to the saccade in position x/y), I assumed I'd eventually be shown RF maps (or RF centroids) and how they change across time. That's pretty much never analyzed per se and never plotted. I also expected at least some verbiage about what we mean by remapping. Normally, one thinks about remapping as a change in RF plotted in retinotopic coordinates. Yet, this entire paper discusses stimuli and RF characteristics in screen coordinates both before and after a large saccade (which I find vexing to think about). In the older literature (on V1/V2) that I know about such things, the idea is that just before a saccade a neuron might respond to the stimulus that will (post-saccade) land in the RF, i.e., that its RF now includes (by extending the RF or adding a 2nd lobe) a region shifted in the direction of the saccade. Or the "compression" story would have an RF be biased in the direction of the saccade pre-saccadically, and in the opposite direction afterward (for an RF the is now located on the other side of the saccadic target). It's very weird, to me, to think of all of this in screen coordinates, and that definitely tripped me up.

What follows are a bunch of other things that confused me on first reading, but for many of which I eventually figured it out after much work. I guess my point is that you should make it possible for a casual reader to just read from the beginning forward and get the gist of what you did and the analysis, without having to jump to Methods and Supplement in order to get anything out of the main text.

Specific edits are addressed below. We hope that the collective revisions allow for a clearer, more straightforward readthrough of the paper.

Specifics by line number

* 85: "relative to THAT timepoint": At first I wasn't sure what this sentence meant, so it might be good to phrase it more transparently, especially since space is barely mentioned: we estimated a spatiotemporal kernel for each neuron, i.e., if a stimulus appeared at position (x,y) at time t, we estimated how much that would increase the probability of firing at time t+\tau (ok, ok, this is not precisely true, but it's close).

We have simplified this sentence and left the details in the later sections. Kernels are defined in line 84, and are explained more in details in the context of STUs later in the results in line 148-155.

* 123: "multiple basis functions": You have to go deep into the supplement to find out what this phrase refers to. The number of basis functions, initially, is the same as the raw data (for each x/y it's equal to the number of 7 ms stimulus time steps times the number of 7 ms delays you analyze). That is, before you drop "unused" basis functions (the "sparse" SVGLM), the basis functions only serve to allow you to interpolate a smooth kernel between discrete samples/bins. Anyway, this paragraph confused me because I wasn't sure what was an STU. Is it the 4-d data structure (function of x, y, t and \tau)? I thought saw, but then you say Figure 1c shows the "STUs" (i.e., plural) at an example x/y. To me, that's just a 1-d slice through the single 4-d STU. An STU is "represented" by the weights, but isn't the STU still a 4-d matrix (the sum of the sparse basis functions you leave in, scaled by the weights)? Then the STUs "comprise kernels ... at that location". So is an STU just the 2-d kernel for a specific x/y, or is it the whole thing?

We removed the mention of “basis functions” from the main text, and only left them in the supplementary information. We see that our language describing STUs was inconsistent in the previous draft. We now clarify that STUs are discrete units in a 4-dimensional space (horizontal probe location x, vertical probe location y, time bin t, delay bin tau). The weights of STUs are scalar and have single numerical values. We have revised the original lines (now line 145-154) to read:

“First, we discretized the space into 81 locations in a two-dimensional plane (representing the 81 probe locations) and discretized all times of the neural response relative to saccade onset and delays (times of the stimulus relative to each response time) into 7-ms bins, to form spatiotemporal units (STUs) in a 4-dimensional space (Fig. 2b). These STUs were used to construct the spatiotemporal kernels of each neuron, which represent the neuron’s sensitivity across locations, times and delays. The neuron’s spatiotemporal kernels were constituted by a weighted linear combination of STUs, where the weight by which each STU contributed to the kernel was estimated using a model that could capture the neuron’s spiking response. Figure 2c shows the STU weights at an example probe location around the RF of an example neuron across time and delay.”

* 131: You say "convolved with", but then the kernels should be, indeed, kernels, not impulse response functon(i.e., flipped). I'm sure you don't mean that. And convolve isn't

quite right either, because it's not shift-invariant. You measure a separate impulse response for every x, y and, most importantly, t.

We define kernels in a more general sense as 4-dimensional objects that can vary over space, time, and delay dimensions. The 'Encoding model framework' section and Eq. (1) in Methods clarify the operations applied to the stimulus signal using these kernels.

We removed the convolution term to keep the text more descriptive and accurate, and referred to the Methods for the technical details. We edited the following sentences in line 158-162 to be clearer:

"A signal representing the stimulus time and location passes through spatiotemporal kernels (representative of the neuron's time-varying spatiotemporal sensitivity), and is added to the time-varying baseline neural activity relative to saccade onset (captured by an offset kernel) and the feedback signal generated using a post-spike filter (representing the effects of spiking history)."

* 143-144: This is the first hint that you aren't going to show me shifting/distorting retinotopic-coordinates RF maps. And, you aren't going to assume or even discuss how the population code is read out to provide a stimulus location estimate. Rather, you determine that for a give x/y/t/\tau combination, if the response across the subset of neurons with RFs at a similar location is similar to the same set of neural responses to a stimulus at location x+1, then there will be a rightward bias. In other words, near as I can make out, you are trying to make a strong statement about the predicted bias direction without saying anything about the specifics of downstream readout mechanisms. That's a bit obscure, and rather indirect logic, but I get why you are doing it. But, you don't advertise what you are doing nor why beyond this one rather telegraphic sentence. In fact, you never clearly state that your goal is to demonstrate readout bias without having to specify how the readout works.

We now introduced the approach sooner in the introduction as in following lines 86-88:

"We quantified the representational spatial bias using the spatiotemporal kernels of populations of neurons, based on the similarity in neural sensitivity to neighboring probe locations, without assumptions about the downstream readout mechanisms."

and clarified again in the below lines 172-175:

"In other words, if during a saccade the population response to one probe becomes similar to that of a neighboring probe, we can assume a representational bias toward that neighboring location, without specifically modeling downstream readout mechanisms."

* 146: "the representation of a probe": you might say explicitly that you mean the population response (for a particular time of stimulus and delay to looking at the response). To take this a little further, you compute your correlations (unless I'm confused) for vectors of responses across neurons for a fixed x/y and t/\tau. But, wouldn't it make some sense to compute the correlations across \tau values as well (the full temporal response to a stimulus at time t across the set of neurons to a stimulus at x/y is similar to that at x+1/y)?

We have updated the wording as suggested:

"In other words, if during a saccade the population response to one probe becomes similar to that of a neighboring probe, we can assume a representational bias toward that neighboring location, without specifically modeling downstream readout mechanisms."

The correlations were computed across t and tau, as shown in Figure 3c. We then averaged across delays of interest (50-100 ms) to get the spatial bias (Figure 3d).

* 156: Now that I'm writing the review, I must still be confused, or maybe not. The correlation coefficient (same as cosine similarity) is for a fixed x/y/t/\tau but across neurons in an ensemble (the vector length is the number of neurons in the ensemble), yes? It's not across weights on your basis functions, is it? I thought the sparse representation was summed to give a continuous response, but that continuous response was used to get the vectors (across neurons) that were correlated. This is really impossible to understand from the main text.

Correct, the correlations are across neurons in an ensemble (at a single time and delay value), now clarified in line 180-186:

"For each ensemble, we measured the similarity between the neural sensitivity at neighboring locations by taking cosine correlations between the kernel weights at the center probe and the kernel weights at each of its eight neighboring probes; these correlations were calculated at each time and delay value. For the rest of the paper, correlation always refers to cosine similarity between the kernel weights for neighboring probes across neurons in an ensemble."

* 164: Again, it's a logical leap to jump from a summed set of correlations, turned into vectors, to call that sum "bias" (by which we'd usually mean a shift of an RF, or a set of RF, or a read out location estimate). As I said above, I think I get your logic for acting like this is okay, but you don't justify it explicitly, so a casual reader will likely trip over this sentence.

We removed the reference to "bias" from the sentence, and explicitly define "spatial bias" in the following sentence in line 193-196:

"We then averaged over the eight vectors at each probe location to get one vector (Fig. 2e purple arrow). Since the saccade direction was either to the left or the right horizontally, we focused on the horizontal projections of the average vectors, which we defined as the spatial bias."

* 173: I now realize that your "bias" estimate is really a relative bias between positions x and x+1 (and x-1, etc.), not an absolute bias. The overall population response could be shifted a LOT, so that a readout would get a hugely shifted estimate, but all you are noticing is that however the full population is interpreted, locally the response is similar to x as to x+1, even though the form of that population response might look overall like the stimulus is at location x+20 (yeah, yeah, you don't have an x+20).

That's correct. There are many different ways one might define a bias measure based on neural responses, and this is the one we used. Sampling RFs across the entire visual hemifield requires substantially larger data collection. We clarified this definition by adding the following sentence in line 188-189:

“We defined a spatial bias measure based on relative ensemble sensitivity to probes near the ST.”

* 176: I'm not sure how this normalization works. Presumably it's a single scalar per ensemble? If you divide by the max positive correlation, so the normalized number range up to +1, that doesn't mean they'll range all the way to -1. I assume you don't do an affine stretch because it would be silly to shift a zero correlation to be nonzero.

The maximum and minimum bias for each ensemble vary which could make averaging across ensembles not helpful for visualization, so we chose to normalize the spatial bias map of each ensemble using the following formula:

normalized_bias = 2*(bias-min(bias))/(max(bias)-min(bias))-1

The range of original values of each map was different but they were quite symmetric around zero; none or negligible shifting of a zero correlation due to normalization. We clarified the procedure in line 208-210:

“We then averaged the bias maps for 15 ensembles (447 neurons) (Fig. 3c), where the individual maps were first normalized so that each ensemble had bias values ranging from -1 to 1 (see Methods).”

* 207: phenomenon -> phenomena

Updated.

* 210: 60:90 is now post-saccade, right? This is the first time you really need to unpack how to think about this stuff, since you plot in x/y in screen coordinates, but the eyes have now moved, so one expects the RF in screen coordinates to shift with the length of the saccade. Clarification please.

The **response** is 60:90 ms post-saccade, and the delay between stimulus and response is 80:110 ms, so the stimulus onset is between 50 ms before to 10 ms after saccade onset (-50:10 ms). Fig. 6 breaks down responses to the same screen locations during fixation 1, around saccade onset (perisaccadic), and during fixation 2 (postsaccadic). The analysis focuses on the sensitivity and bias around the ST location.

Within 10 ms, for saccades of amplitude 13 dva, which take on average ~30 ms to complete, the eyes will have moved ~4 dva. With RF eccentricities of ~10 dva, this would not be enough to actually bring the probe into the post-saccadic RF.

* 227: Remind the reader here (I think it's buried in Methods somewhere) that you flip things for leftward vs. rightward saccades so that it makes sense to compare these in the same plot, i.e., it's independent of which side of the saccadic target you started on

Added the following sentence in line 263-264 to clarify:

“Data from ensembles recorded with leftward saccades have been flipped to be combined with those recorded with rightward saccades.”

* 235: I didn't dig into what "p < .001" meant here, but this was all about model

predictions. Did the p value come from a collection of stochastic simulations of the model? That is, are you making the mistake of computing a p value for a simulation when you could have easily, and pointlessly, inflated that p value by running more simulations?

No, the *p*-values were computed across the number of neurons (paired comparison of near vs. far predicted responses for each neuron). We added “n = 94 neurons” after the *p*-value to clarify.

* 261: Again you are looking at postsaccadic values of t, when the eyes have moved and one expects the RF, in screen coordinates, to have moved a lot. So please clarify what we are looking at and what to expect.

The analysis focuses on the probes around the ST location. The perisaccadic window includes the stimulus appearing -20:10ms relative to the saccade onset, when the location of the perisaccadic RF is expected to be similar to that of the presaccadic RF (as explained in our response to the earlier comment). For many neurons, the near-ST probe locations fall in the postsaccadic RF, and not the presaccadic RF, as seen in Fig. 6b. Perisaccadic responses resemble the postsaccadic response more than presaccadic response.

* 293: FF = feed-forward? Note: I'm not a fan of having to learn a bunch of acronyms when I read a paper (STU, ST, etc.).

We have updated all “FF” to “future field”, but would like to keep the abbreviations for STU and ST to be consistent with our previous papers.

* 311: You have neurons from MT and V4 but never discuss any are differences! Also, you never discuss whether the biases you find might be inherited wholly from modulations of responses in upstream areas.

The results in the two areas look similar (Supplementary Fig. 6), and so our focus is on their shared properties. We added a sentence in the main text in line 294 noting this result:

“Results look similar for MT and V4 neurons (Supplementary Fig. 6)”

We also added a sentence to discuss the possibility of inheriting the modulation from upstream areas in the discussion section, line 421-423:

“We also cannot definitively state whether these spatial biases in responses arise first in MT and V4 or are inherited from upstream areas.”

* Another point: your predictions of "bias" from correlations assumes that these similar population vectors (from x and x+1) drive a distorted/biased readout. That is, it assumes that the readout is blind to the effects of time relative to a saccade. It would be good to talk about that explicitly.

We added a sentence in line 397-400 to clarify:

“Like most previous studies attempting to understand this connection, our approach for measuring spatial bias assumes the same decoding algorithm is used during fixation and

around the time of saccades, with altered visual responses driving the perisaccadic perceptual changes."

* This stimulus ensemble is sparse, but not "spherically symmetric" and thus if you were just doing a spike-triggered average to estimate a receptive field, that estimate would be biased (as you know from teaching MathTools!)

With regards to the model fitting: to address possible challenges regarding an unbiased estimation of the model components including the stimulus kernels, we took advantage of several procedures and properties in both the experimental design and in the model framework and fitting, to minimize and trace possible biases in the estimation. The high resolution and complexity of the stimulation paradigm, the high dimensionality of the data, the scarcity of perisaccadic spiking activity, the fast dynamics of spatial and temporal receptive fields, required us to follow a systematic procedure. Those procedures and the validation results have been presented in the two previous publications from the lab [ref 42 and 49], which in part included pseudo spatial and temporal white noise stimulation, maximum likelihood optimization framework, controlling for overfitting or prediction variability, data resampling and permutation, robustness measures for non-convexity, statistical sparsity, data-informed discretization, data-driven dimensionality reduction, etc. To independently verify that our procedure was not subject to significant bias due to the binary white noise stimulation, we calculated the empirical RFs directly from the neural responses in our earlier studies of this dataset. We ensured that the model-estimated and data-estimated RFs were consistent. At any rate, the trial-averaged and stimulus-aligned nature of the empirically estimated RFs made them of much lower resolution than the model-predicted RFs on a millisecond timescale.

We also added a paragraph for "Empirical RF estimation" in Supplementary Information:

"The centers of RFs were assigned based on responses to the probes that generated the maximum firing rate during the fixation period before the saccade. For each probe location, the probe-aligned responses were calculated by averaging the spike trains over repetitions of the probe before or after the saccade (greater than 100 ms before or after the saccade onset), from 0:200 ms following probe presentation, across all trials. The response was then smoothed using a Gaussian window of 5-ms full width at half maximum."

* 533: Another place that you might clarify that x/y is in screen coordinates, but flipped for leftward saccades and that "0:200" is looking backward in time to the stimulus from the response bin. "9 probe locations around the neuron's RF" also bumps into the issue of screen vs. retinotopic coordinates and whether flipped for leftward vs. rightward saccades.

We updated the following sentences in line 627-635 to clarify:

"Time refers to the time of response relative to saccade onset (-540:540 ms), and delay refers to the time of stimulus relative to particular response time (0:200 ms before), discretized in bins of 7 ms. Bottom cubes at specified time points denote the spatiotemporal kernels obtained by the weighted combination of STUs. Each kernel has spatial (x, y), time (t), and delay (τ) dimensions, and the spatial coordinates are based on locations on the screen. The line plot within each cube shows a kernel across its delay dimension only. d. Each layer represents the kernel map of one neuron along the

time and delay dimensions for 9 probe locations around the neuron's RF during the initial fixation. These kernels are calculated for each neuron in each ensemble (denoted by several layers in the plot)."

Within each ensemble, there is only one saccade direction. We have added clarification on flipping direction when combining ensembles in line 263-264.

Figure 2 utilizes the RF kernels to illustrate the methods used for calculating the spatial bias, however, the rest of the paper and analysis focus on the ST region independent of the RF locations or saccade directions.

* 615: A trial consists of one of each of the 81 probes in a random permutation?

Yes, but in pseudorandom order. A trial consists of a sequence of probe stimuli appearing randomly at any of the 81 possible probe locations, each turned on for 7 ms. We updated the following sentence in line 734-736 to clarify:

"Throughout the length of each trial, a complete sequence of 81 probe stimuli flashed on the screen in pseudorandom order, one at a time for 7 ms each."

* 621: Was the grid set up so that both leftward and rightward saccades would leave the RFs inside the grid?

Each session only has either leftward or rightward saccades, and both the presaccadic and postsaccadic RF centers are inside the grid. We have clarified in the Methods in line 740-742:

"During each neurophysiological recording session, the grid of possible locations of the probes was placed and scaled to cover the estimated presaccadic and postsaccadic RF centers of the neurons recorded, the FP, and the ST."

* 632: You mean similar RF position but identical ST and grid position? So a session only had leftward or only had rightward saccades???

Yes, each session only had either leftward or rightward saccades. We have clarified in the Methods in line 730-732 and line 786-787:

"While the monkey was holding fixation, a ST appeared 13 dva away from the FP horizontally. In each recording session, there was only one saccade direction (leftward or rightward)."

"Neurons recorded with the same ST position and probe arrangements (grid positioning and spacing), and with similar RF locations were grouped as an ensemble."

* 643: "sparse collection of STUs": This sounds like a non-deleted basis function is what you are calling a STU, which I don't think is what you mean.

Each STU is assigned a weight parameter and the STUs whose corresponding weight parameter is found to be statistically significant are included in the model fitting and other analyses. We have edited the phrasing in this sentence in line 762-767:

"The SVGLM can capture the neurons' sensitivity varying over space and time with high temporal resolution by providing a parameterized representation of the neuron's

spatiotemporal kernels using discrete STU components. Through a dimensionality reduction process, the model selects statistically significant STUs to achieve sparsity for fitting the parameters to the data (see Supplementary Information)."

* 652: "k_{x,y}(t,\tau) represent the stimulus kernels". Huh? isn't k a stimulus kernel (i.e., the 4-d thing)? You say "kernels" (plural) and treat that inline equation like its plural. I can't parse that. Do you ever say how you estimate h(\tau) and using what parameterization?

Yes, 'k' represent stimulus kernels, defined here as 4-dimensional functions/objects, which take values over {t,tau,x,y}. In the paper projections/slices of kernels across a subset of these dimensions may have been referred to as kernels too. For example, the term kernels as used here emphasizes that there are distinct kernels for each (x,y) location.

We omitted the details of the model components not related to this study (including the post-spike filter), and instead described them briefly and referred to our previous publication on the model development for full details [ref 49]. Below are some details about the post-spike filter adapted from that paper:

$$h(\tau) = -\sum_i \eta_i^2 . H_i[\tau],$$

where the basis functions H_i (τ) representing the post-spike kernel were chosen to be B-spline functions of order two with non-uniformly distributed 23 knots over the delay variable τ; the spacing of the knots around zero, which indicates the spike time, was smaller and increased further away from the spike time and its associated refractory period (the knots were spaced at {1, 2, 3, 4, 6, 8, 15, 22, 29, 36, 43, 50, 57, 64, 71, 78, 92, 106, 120, 134, 148, 162, 176} ms, in total 20 basis functions).

We also included that reference to the end of this paragraph in line 783-784:

"More details about the model structure, estimation, and evaluation can be found in [49]."

* You don't ever show temporal evolution of a spatial RF and don't show anything like a spatial RF until Figure 4

This is true. Rather than looking at how the spatial RF of a particular neuron changes over time, we look at how the response of a neuron to visual stimuli at one fixed location (in the world/screen coordinates) changes perisaccadically.

This analysis of responses to one probe location (defined in screen coordinates) matches the behavioral analysis: the behavioral mislocalization measure is a difference in reported location for a probe in one screen location during fixation vs. perisaccadically; we look at the corresponding neural responses to the same probe location during fixation vs. perisaccadically.

* 665: smoothed over t, over \tau or both?

Both, we smoothed over t with windows of 50 ms, and over tau with windows of 20 ms. We have clarified the sentence in line 788-790:

“Before any analysis, kernels of all neurons were smoothed by moving average using boxcar windows of length 50 ms across time t and 20 ms across delay τ to reduce noise.”

* 672-673: This is a range of time t, but what range of delays \tau figure into this? Separate correlations computed for each value of \tau???

We subtracted the correlation at each (t, τ) by its corresponding baseline correlation measured by averaging the correlations at the same delay, over the fixation window (-441:-141 ms).

* 678: Again, this bias vector from correlations is pretty indirect, and it's not clear that it uses or ignores the overall strength of a response or the spatial RF itself except that if you are outside of the RF, responses become noise and correlations should drop, leading to low, noisy bias.

As a second method of predicting changes in location decoding, we have established a maximum a posteriori decoder based on response of single neurons to directly measure the confusion in decoding. As shown by the newly added Supplementary Fig. 4, we observed that the decoder is confused in the direction opposite to the saccade perisaccadically, which validates the observed negative spatial bias predicted by the correlation measure.

* 682: How does one see the shift of the RF in Fig. 2a-b?

There is not a simple intuitive relationship between our bias measure and the shift of RFs. For example, either saccade target or future field remapping of RFs could potentially underly the spatial bias.

* 684: This is probably the only place you allude to a postsaccadic RF, i.e., to the fact that you plot in display coordinates.

We hope our additions to the text in response to the previous comments have helped clarify this point. We note again that mislocalization, at a behavioral level, is measured in display coordinates, and we are trying to measure neural responses in the most comparable format.

* 691-696: Give that I'm still unclear what an STU is, this section lost me entirely. I thought an STU was the full 4-d linear kernel. What's the group of items that your are using to determine the subset that are "modulating"? What does the word "prevalence" mean here (it's never defined)? Why does Eq. 2 only mention time, not space? I'm completely lost here. The same confusion applies to the units of "bias-relevant" STUs and the AUC computed from them.

We hope our previous clarification of our STU definition has been helpful.

Modulated STUs are determined based on the difference of distribution of STU weights obtained in the fixation vs. perisaccadic period, for all probe locations, delays and times. We have replaced the term “prevalence” with “the fraction of contributing STUs”, and added the following sentence in line 816-820 to clarify the “contributing STUs” (also described in the ‘Dimensionality reduction’ section in Supplementary Information).

"These contributing STUs were identified using the dimensionality reduction process during the model fitting, based on a statistical significance criterion, which measures the contribution of individual STUs to capturing the stimulus-response relationship (see Supplementary Information, and Niknam et al.[49] for details). "

Eq. (2) was applied to each individual STU, which corresponds to a single probe location.

The bias-relevant STUs are the subset of modulated STUs whose inclusion generates a bias index greater than a threshold value. The bias index is defined based on the contribution of a STU to the difference between the kernel over times and delays associated with its location and that of the neighboring locations, according to Eq. (2).

* Fig. 1c: the y-axis label should be a clearly Greek \tau. This same weird "T" occurs in the running text (Supplement, 25). This figure's legend should emphasize that these are screen coordinates.

We have updated the figure and the caption in line 630-631:

"Each kernel has spatial (x, y), time (t), and delay (τ) dimensions, and the spatial coordinates are based on locations on the screen."

* Suppl. Eq. 1: Is B_{i,j} what you call an STU?

We hope our previous clarification of STU, varying over four dimensions, was helpful. B_{i,j} are two-dimensional basis functions varying across time and delay at 1-ms resolution. The linear combination of these basis functions scaled by the STU weights construct the stimulus kernel. We have updated the 'Dimensionality reduction' section in Supplementary Information to further clarify the equations.

There was also a typo in Eq. (3) for indexing S and we have corrected it.

* Suppl. 30: The full set of B_{i,j} is as big as the data, so I don't know what you mean by binning and two orders of magnitude reduction

This line refers to the discretization process by down-sampling the time and delay dimensions using 7-ms bins to define STUs.

The indexing of B_{i,j} has been described in line 34-43 in Supplementary Information.

* Does the subset of the B_{i,j} retained depend on probe location?

The same basis functions are used for constructing the kernels corresponding to different probe locations, but a different subset of weights for the basis functions B_{i,j} are estimated for each probe location.

Reviewer #3 (Remarks to the Author):

The authors record responses in MT and V4 to brief visual probes around the time of a saccade. They find a spatial bias in the direction opposite the saccade and enhanced

sensitivity around the area of the saccade target. This is an interesting and broad investigation of perisaccadic neural response and I have only three comments.

First, the central measure is the correlation between a RFs response and its neighbors. When the correlation is higher on one side, that is labeled a “spatial bias”. Here is the logic they used:

“We made the assumption that similar neural sensitivity to probes appearing at neighboring locations could create uncertainty in a readout of the stimulus location by a downstream area, which would lead to a bias in spatial perception. In other words, if during a saccade, the representation of a probe becomes similar to that of a neighboring probe, we can assume a representational bias toward that neighboring location.” Lines 143-147.

This sounds plausible but it is not clear what is meant by “representational bias”. How does it translate to perceived location? The bias they find is away from the saccade target, specifically RFs on the side of a cell opposite the saccade direction are responding like the cell.

This could be taken as a shift of receptive fields toward the saccade target – the RF further away from the saccade target is responding as if it were closer. This shift has been reported by others. But if receptive fields shift toward the saccade target, perceived location must shift away according to classic labeled line reasoning. This is the wrong direction. The authors nevertheless claim that their results “account for changes in spatial perception during eye movements.” They have not shown how this happens. Yes, the receptive fields may be shifting (showing a representational bias) toward the saccade target but that does not account for a shift in location toward the saccade target. It predicts the opposite. Perhaps the explanation is there somewhere in the text but I couldn’t find it. The authors need to make a clear link between their spatial bias and perceive location if they wish to claim that they account for the psychophysical results of perisaccadic location perception.

We do not directly model or measure shifts in RFs in this paper. Instead, we assumed that when neural sensitivity to two probe locations becomes similar, those two locations are more likely to be confused. When this similarity in sensitivity to perisaccadic probes was systematically biased in a particular direction relative to the saccade, we assumed that probes would be mislocalized in that direction.

First of all, thanks for the comments. In response to this comment and the first reviewer’s comment now we have added psychophysical evidence for spatial mislocalization around the time of saccade (See Fig. 1). We observed mislocalization in the direction opposite to saccade direction for stimuli presented around -100:0 ms prior to saccade onset. The spatial bias computed from kernels is most negative at around 50 ms from saccade onset and 50:100 ms delay relative to that time, which translates to around -50:0 ms for time of stimulus presentation, similar to what we observed behaviorally.

Moreover, we added a second method of relating response changes to predicted changes in the readout of location: we established a maximum a posteriori decoder based on response of single neurons to directly measure the confusion in decoding. Perisaccadically, neural response at neighboring probes become similar, and thus become more difficult for the decoder to differentiate, leading to a bias in perception. As

shown by the newly added Supplementary Fig. 4, we observed that this decoder is confused in the direction opposite the saccade perisaccadically, which validates the observed negative spatial bias from correlations.

We also note that the relationship between RF shifts and decoding is not always intuitive and will depend on assumptions about the decoder's 'awareness' of RF shifts—see additions to the discussion section line 390-393.

Second, the authors present highly processed results in terms of normalized arbitrary units. These show a spatial bias in the expected direction at a reasonable time widow after saccade onset. However, we have no idea whether the scale of the effect is in the order expected for perisaccadic mislocalization. Perhaps 0.2 arbitrary units is a very small shift and if so that would suggest that these anatomical areas are showing only initial hints of the large displacements that may come later in the visual system. Please do some reverse modeling and convert the spatial bias to actual degrees of visual angle so we can decide if this is the site where the shifts arise or not. Also, and again arguing against a relation to perceived shifts, the results show no shifts before the time of the saccade even though behavioral results do.

We have updated the figures to write "(unit)" instead of "(a.u.)" to make it clear that we have normalized the spatial bias to range from -1 to 1. Whereas for the previous measure, it would be inaccurate to convert the values into actual degrees of visual angle; the newly presented behavioral data (Fig. 1) and the new decoding analysis can give a more tangible description of the magnitude of the mislocalization in terms of dva (behavioral data) and representational confusion (decoder).

Note that we are reporting spatial bias in terms of **response times** relative to the saccade, which corresponds to probe onsets prior to the saccade. The spatial bias computed from kernels is most negative around 50 ms from saccade onset with 50:100 ms delay, which translates to around -50:0 ms for time of stimulus presentation, similar to what we observed behaviorally (new Figure 1).

Finally, the authors record from MT and V4 but do not give any breakdown or summary of the responses for the two areas.

The results in the two areas look similar (Supplementary Fig. 6), and therefore we focus on their shared properties. We added a sentence in the main text noting this result:

"Results look similar for MT and V4 neurons (Supplementary Fig. 6)."

Reviewer #4 (Remarks to the Author):

Weng et al. use a previously established rigorous encoding modeling framework to characterize the spatiotemporal kernel of neural responses to visual probes flashed at various times relative to horizontal saccades. The encoding model developed by the same authors previously is robust and statistically sound. However, there is a major concern in the interpretation of what the authors label as spatial bias. I outline them point-wise below. They must be addressed before the paper is ready for publication.

Major (conceptual):

The spatial bias is measured in terms of correlations between responses to probes at a central location and probes at areas surrounding the central location. Throughout the paper, the authors choose probes near the ST as the central location. Furthermore, the correlations are projected onto the horizontal axis – the axis of the saccade direction. Based on these metrics, the authors find a spatial bias in the neural representation around the time of a saccade and that the bias is in the direction opposite to the saccade. The authors are attempting to link the neural representation of visual space to spatial perception around the time of saccades, but their specific goal of linking ST remapping to the peri-saccadic compression effect is not supported by the results. I find the results compatible with FF remapping because the direction of spatial bias is unidirectional as opposed to omnidirectional, as you'd expect with ST target remapping. Because the authors only look at the horizontal projection of spatial bias, it is difficult to dissociate whether the bias they report is due to FF remapping or ST remapping. At face value, it looks like FF remapping, but this could be due to the choice of horizontal projection of the data. Any statement on ST remapping must be assessed with a parallel analysis of the vertical projection of spatial bias of neuron ensemble. Since the probes were not limited to horizontal direction, this should be possible with the existing dataset.

For near probe vs. far probe analysis – again, the results are compatible with FF remapping. The authors should compare responses to the three probes along the vertical axis to assess whether the responses are compressed towards the ST.

It is important to note that this dataset is not optimal for differentiating between ST and FF remapping, because the relative arrangement of RFs and STs puts the future field close to the ST for many neurons, and lack of sampling of the full range of probe and RF positions. For this reason, we do not attempt to make a claim about the nature of remapping. We do see an enhanced perisaccadic sensitivity to probes appearing near the ST, even when attempting to exclude neurons whose future field covers these locations (Fig. 7), but we now avoid using the term 'ST remapping' to describe this effect to avoid confusion. The issue of the exact nature of remapping is an important question which we hope to address in the future with a more complete dataset. We now acknowledge this limitation more explicitly in the discussion in line 416-418:

"It should be noted that future field remapping could also contribute to perisaccadic spatial bias, and our dataset was not optimized to definitively differentiate between these two forms of remapping."

Per the reviewer's request, we have now included the vertical components of the spatial bias in Supplementary Fig. 3a, which shows no average bias for probes above or below the ST. Supplementary Fig. 3b compares the neural responses along the vertical axis by averaging the response of the top, middle, and bottom two probes for the 6 probes around ST, which also demonstrate no compression towards the ST. Note that our new behavioral data also shows no net vertical mislocalization (Supplementary Fig. 1c). However, this analysis is still limited to probes near the ST and cannot rule out compression towards the ST at greater eccentricities.

Line 287-289: I don't understand the description of the sensitivity index. As a result, I don't understand Fig 6 and don't understand the motivation clearly. Spatial bias and

sensitivity are quantified from the same dataset, and it has a feel of double dipping. But this could be clarified in the main text and in methods.

Certainly both measures are from the same dataset and fitted models. The sensitivity index looks at the average change in sensitivity to all near-ST probe locations over time, whereas the spatial bias measure looks for patterns in the similarity of responses between various near-ST probe locations. Sensitivity could theoretically change with no change in spatial bias, and spatial bias could theoretically change with no change in overall sensitivity. In our data, both change, and we wanted to see if those changes had a common cause.

We used the kernels to quantify the sensitivity index so we could compare the indices computed from kernels before and after removing the bias-relevant STUs, and thus demonstrated that the bias-relevant STUs are also responsible for the enhanced ST representation. To make a smoother transition, we have added a few panels showing the perisaccadic increase of neural firing rate around ST (Fig. 7a-b). We also added the following sentence to the text (line 331-333): “This sensitivity index quantifies the average changes in sensitivity for near-ST probes and can theoretically vary independently of the spatial bias measure.”

To truly assess the perceptual bias, monkeys could be trained to report the probe locations. Such an experiment would make the paper exponentially better and behaviorally relevant. However, I understand that the behavioral experiment is out of the scope of this study. Nonetheless, I encourage the authors to pursue this path given their excellent modeling tools at hand. More broadly, the question of readout is crucial to correctly interpret the implication of RF remapping, namely whether the downstream decoder is aware or unaware of RF shifts. Mickey Goldberg and colleagues have recently written about this. The authors’ argument would be more understandable if they reframed their paper in terms of the language in Goldberg’s paper:

Tuning curves vs. population responses, and perceptual consequences of receptive-field remapping. Ning Qian, Michael E. Goldberg, and Mingsha Zhang. Frontiers in Computational Neuroscience (2022). DOI: 10.3389/fncom.2022.1060757

We have included behavioral results measured from monkeys (in separate experiments) in the new Fig. 1. We observed mislocalization in the direction opposite to saccade direction for stimuli presented around -100:0 ms prior to saccade onset. The spatial bias computed from kernels is mostly negative around 50 ms from saccade onset with 50:100 ms delay, which translates to around -50:0 ms for time of stimulus presentation, similar to what we observed behaviorally. We did not observe vertical localization error perisaccadically (Supplementary Fig. 1c).

Thanks for suggesting this very relevant paper by Qian, Goldberg, and Zhang. To be consistent with our previous work and clear about the distinction between behavior and neural response, we have decided to keep the terminology as “future field remapping” and “saccade target (ST) remapping”. We updated the sentence in line 60-64 to include the other terminology:

“Some neurons in the extrastriate visual areas and prefrontal cortex show a sensitivity shift to the postsaccadic receptive field (RF) even before the saccade, a phenomenon often referred as future field remapping or forward remapping[27,28]. There is also another

phenomenon, called ST remapping or convergent remapping, in which neural RFs shift towards the ST around a saccade[29–38].”

We also added a few sentences in the discussion in line 393-397 referring to the results of Qian, Goldberg, and Zhang:

“In a theoretical study done by Qian and colleagues, a decoder predicted that future field remapping would result in mislocalization opposite to the saccade direction when the decoder is unaware of the RF shift[59]. When there is ST remapping, the unaware decoder predicted divergent mislocalization, and the aware decoder predicted convergent mislocalization.”

Major (technical):
Can the authors compare the V4 and MT populations and compare the % of neurons with spatial bias?

The results in the two areas look similar (Supplementary Fig. 6). We added a sentence in line 294 noting this result:

“Results look similar for MT and V4 neurons (Supplementary Fig. 6).”

The spatial bias was measured from each ensemble of neurons, and thus we cannot measure the percentage of individual neurons with spatial bias. However, we included a new decoding analysis which was based on responses of single neurons (Supplementary Fig. 4).

Keeping the probe locations constant, can the authors compare trials in opposite saccade directions? The bias should also have opposite directions. In other words, can the authors disaggregate the saccade directions instead of combining all saccade directions? Even though this will sacrifice statistical power, it will make the case more convincing.

In each session, saccades were only in one direction, so we cannot compare saccades in opposite directions for the same neurons.

We have added Supplementary Table 1 which compares the spatial bias for ensembles with RF center close to vs. far from ST, and leftward vs. rightward saccades. The comparisons show that magnitude of spatial bias is affected by the distance between RF and ST, but is consistent for leftward and rightward saccades.

Minor:
Fig 1d vs 1e: it is confusing to look at these two figures side-by-side because it gives the impression that they represent the same probe cluster, but 1e is around ST, and 1d is around RF. Please label them clearly on the respective figures.

We have updated the figure (now Fig. 2d-e) to clarify the probe locations.

Fig 2c is missing the color bar.

The color bar is on top of the spatial bias map (now Fig. 3c).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have sufficiently addressed my concerns. I find the paper in good shape for publication.

Reviewer #2 (Remarks to the Author):

Re-review of "Neural correlates of perisaccadic visual mislocalization in extrastriate cortex"

This version is FAR more intelligible and clearly motivated to me than the first one. There are still a few analyses that I haven't fully understood, but this is much easier to wade through. I'm still confused by the AUC here (I'm not sure which curve based on which threshold that leads to an AUC.

Other things:

Fig. 7 legend: RF1 -> RF

Supplement Eq. 3: I'm not sure what "conditional intensity function" means (conditional on what, what sense of "intensity"). The equation seems weir. Why is there no sum over i,j? What is k_{x,y}, i.e., why does it no longer have subscripts i,j?

Supplement line 136: earlier [around] time

Reviewer #2 (Remarks on code availability):

N/A

Reviewer #3 (Remarks to the Author):

The authors have made extensive changes to the text and added a behavioral experiment with the monkeys. As before, they report a spatial bias in the direction opposite the saccade and enhanced sensitivity around the area of the saccade target. Importantly, the spatial bias is opposite the direction of the saccade in the period following the saccade onset (Figure 3d, e, f). The conversion of the units of this shift to actual dva is said to be complicated and is not attempted. The behavioral experiment showed mislocalization of probes in the 100 ms before a 10 dva saccade also in the direction opposite to the saccade, by a maximum amount of 0.5 dva.

There is a mismatch between the authors' behavioral results and the neural results. The neural bias is seen after the saccade whereas the behavior bias is seen before the saccade. The results are also at odds with publish behavioral data for non-human primates (Jefferies et al., 2007; Klingenhoefer & Krekelberg, 2017) where the mislocalization is much larger, about 5 dva rather than the 0.5 dva reported by the authors here. Finally, the reported increase of neural sensitivity near the saccade target around the time of the saccade is in conflict with the saccadic suppression typical found for humans and non-human primates (Klingenhoefer & Krekelberg, 2017).

My overall evaluation is quite mixed. The procedures and analyses are very sophisticated and extensive but the results do not hold together and are in conflict with other published results.

June 5, 2024

Dear editor,

We have revised our manuscript NCOMMS-23-31269, entitled “Neural correlates of perisaccadic visual mislocalization in extrastriate cortex”, based on the reviewers’ comments. We thank the reviewers for their feedback, as these revisions have made the manuscript clearer and more thorough. We have corrected the typos, clarified methodological details, and added discussion of how our findings compare to other studies. Point by point responses to each of the reviewer’s concerns are included below for your reference.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have sufficiently addressed my concerns. I find the paper in good shape for publication.

We thank the reviewer again for their comments, which helped us make the manuscript clearer and more thorough.

Reviewer #2 (Remarks to the Author):

Re-review of "Neural correlates of perisaccadic visual mislocalization in extrastriate cortex"

This version is FAR more intelligible and clearly motivated to me than the first one. There are still a few analyses that I haven't fully understood, but this is much easier to wade through. I'm still confused by the AUC here (I'm not sure which curve based on which threshold that leads to an AUC.

Each ST probe was treated as a center probe, and each center probe has 8 neighbor probes. Supplementary Fig. 5a shows the kernel of an example center probe (green line) and the kernel of one of its neighbor probes (blue line). The AUC was defined as the area (yellow shaded area) under curve (yellow line)—the difference of these two kernels, which represents the dissimilarity between kernels at these two neighboring probes.

We have updated line 217-219 to clarify:

“The AUC (yellow shaded area) represents the dissimilarity between kernels at two neighboring probes over time and delay.”

Other things:

Fig. 7 legend: RF1 -> RF

We have updated this typo. Good catch!

Supplement Eq. 3: I'm not sure what "conditional intensity function" means (conditional on what, what sense of "intensity"). The equation seems weir. Why is there no sum over i,j? What is k_{x,y}, i.e., why does it no longer have subscripts i,j?’

A sequence of neural spiking events occurring over time can be described as a binary point process time series. A point process, defined by a sequence of discrete events at random points in time, is fully characterized by its the conditional intensity function (CIF) (Brillinger, 1988; Daley and Vere-Jones, 2008). The CIF is defined using the probability of the counting process

associated with a point process, conditional on the history of the process and the covariates, and can be considered as the instantaneous rate given that history.

K_(x,y,i,j) should have the subscripts i,j and we apologize for the typo. The subscripts {i,j} represent the time and delay bin indices, and Eq. (3) calculates the instantaneous firing rate at each time t, summing over all delays τ. {κ_(x,y,i,j)} is the weight of the STU represented by the basis function B_(i,j) (t,τ) associated with location (x,y), and time and delay bin indices {i,j}.

We have updated the following sentences in line 53-61 of the supplementary information to be clearer:

"We obtained the weight distribution of each STU, by fitting a simpler generalized linear model (GLM) composed of individual STUs on 100 subsets of randomly selected spike trains (35% of all trials), and compared it to a control distribution obtained from 100 subsets of shuffled trials where the relationship between stimulus and response was altered. The conditional intensity function (CIF)[2,3] of this GLM was defined as:

$$\lambda_t = f\left(\sum_{\tau=1}^{T} s_{x,y}(t-\tau) \cdot \kappa_{x,y,i,j} \cdot \mathcal{B}_{i,j}(t,\tau)\right) \quad (3)$$

where λ_t is the instantaneous firing rate of the neuron at time t, $s_{x,y}(\tau)$ is the stimulus history of length T at location (x, y), $\{\kappa_{x,y,i,j}\}$ is the weight of the STU represented by the basis function $\mathcal{B}_{i,j}(t,\tau)$ defined in Eq. (1), and f is a static nonlinearity."

Supplement line 136: earlier [around] time

We have deleted the extra word.

Reviewer #2 (Remarks on code availability):

N/A

Reviewer #3 (Remarks to the Author):

The authors have made extensive changes to the text and added a behavioral experiment with the monkeys. As before, they report a spatial bias in the direction opposite the saccade and enhanced sensitivity around the area of the saccade target. Importantly, the spatial bias is opposite the direction of the saccade in the period following the saccade onset (Figure 3d, e, f). The conversion of the units of this shift to actual dva is said to be complicated and is not attempted. The behavioral experiment showed mislocalization of probes in the 100 ms before a 10 dva saccade also in the direction opposite to the saccade, by a maximum amount of 0.5 dva.

There is a mismatch between the authors' behavioral results and the neural results. The neural bias is seen after the saccade whereas the behavior bias is seen before the saccade. The results are also at odds with publish behavioral data for non-human primates (Jefferies et al., 2007; Klingenhoefer & Krekelberg, 2017) where the mislocalization is much larger, about 5 dva rather than the 0.5 dva reported by the authors here. Finally, the reported increase of neural sensitivity near the saccade target around the time of the saccade is in conflict with the saccadic suppression typical found for humans and non-human primates (Klingenhoefer & Krekelberg, 2017).

My overall evaluation is quite mixed. The procedures and analyses are very sophisticated and extensive but the results do not hold together and are in conflict with other published results.

Our behavioral and neural results were measured from different paradigms which might contribute to the difference in timing; also note that behavioral results are plotted based on probe

time, and neural results based on the time of the neural response. It is also possible that the spatial biases in MT and V4 responses are inherited from upstream areas and the neural correlates of the behavioral mislocalization we observed will be further processed by downstream areas to generate the final behavior. Line 503-505 has discussed this possibility:

“We also cannot definitively state whether these spatial biases in responses arise first in MT and V4 or are inherited from upstream areas.”

The difference in paradigms might also contribute to the different magnitudes of mislocalization or its other characteristics between our own study and previous reports (Jefferies et al., 2007; Klingenhoefer & Krekelberg, 2017). Jefferies et al. had 12 dva saccade vectors, 12 dva offset between probes, and their probe locations range over 36 dva horizontally. Their reported mislocalization was normalized for each session and did not subtract the fixation baseline. Klingenhoefer and Krekelberg had 16 dva saccade vectors and 8 dva offset between probes that range over 30 dva. They subtracted fixation baseline, but did not aggregate over mislocalization for the three stimulus locations. Our paradigm had 10 dva saccade vectors and 4 dva offset between probes that range only 12 dva horizontally, which means the animal had a narrower range of possible locations for the probe and could reduce overall error. We subtracted the fixation baseline and averaged over all probe locations for the reported mislocalization over time. Additionally, we only reported mislocalization from probes that are witin 5 dva of ST while the other two studies used probes that are at least 8 dva from ST.

We have added a sentence in line 425-429 to address this difference:

“The magnitude of mislocalization we observed is smaller than that reported in the previous literature, which might be due to various differences in the paradigm, including a shorter saccade vector, smaller offset between probes, smaller range of possible probe locations, and limiting the probes to within 5 dva from the ST.”

Saccadic suppression has been reported as an overall perisaccadic phenomenon, especially for probes far from ST. We found an increase of neural sensitivity near ST, which is consistent with previous psychophysical and neurophysiological studies showing enhanced perception and neural responses at the ST itself. Since we didn’t measure neural sensitivity at other probe locations, our data does not conflict with reports of saccadic suppression at locations far from ST.

We have updated line 405-414 and added a few sentences in line 474-489 to clarify:

“In addition, we demonstrated that the spatial bias is accompanied by the perisaccadic enhancement of neural sensitivity around the ST, with a matching time course and underlying neural response components. This enhancement of ST representation is consistent with previous behavioral and neural findings[24,26,35,37,50–52], suggesting that the brain prioritizes ST representation at the expense of biases in location perception.”

“Previous psychophysical studies have reported enhanced discrimination performance at the ST[35,37] and neurophysiological studies have shown a perisaccadic increase in neural sensitivity at the ST[24,26,38,39,51,52]. The exact spatial extent of this improved perception near the saccade target is unknown, and locations farther from the saccade endpoint instead display an overall decrease in visual sensitivity around the time of saccades, known as saccadic suppression[4–7,35]. Our model-based measure shows increased sensitivity to perisaccadic stimuli within 5 dva of the ST.”