nature portfolio

Peer Review File

Large-scale mega-analysis indicates that serial dependence deteriorates perceptual decision-making

Corresponding Author: Dr Ayberk Ozkirli

Version 0:

Decision Letter:

29th April 2024

Dear Mr Ozkirli,

Thank you once again for your manuscript, entitled "Rethinking serial dependence: A meta-analysis of its effects on the variability of perceptual estimates," and for your patience during the peer review process.

Your manuscript has now been evaluated by 3 reviewers, whose comments are included at the end of this letter. In the light of their advice, I regret that we cannot offer to publish your manuscript in Nature Human Behaviour.

While the reviewers find your work of some interest, they raise serious concerns about the strength of the novel conclusions that can be drawn at this stage as well as the appropriateness of the chose meta-analytic approach. We feel that these reservations are sufficiently important as to preclude publication of this work in Nature Human Behaviour.

I am sorry that we cannot be more positive on this occasion but hope that you will find our reviewers' comments helpful when preparing your paper for submission elsewhere.

Sincerely,

████████
██

Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Meta-analysis, evidence synthesis, perceptual decision-making

Reviewer #2: Serial dependence, perceptual decision-making

Reviewer #3: Serial dependence, perceptual decision-making

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors should explain in more detail what serial dependence is, having a general audience in mind. For instance, the opening sentence of the abstract takes for granted that the reader is already familiar with these effects. This is unlikely the case for the average reader of NHB. In general, the abstract is almost impossible to understand for a researcher that has never come across this literature before. The first paragraph of the introduction doesn't offer much more information. The authors should make an effort to adjust the text to the general audience of the journal.

I couldn't find anywhere information about how the authors identified the studies that they entered into the meta-analysis. The should explain the literature search strategy and the inclusion/exclusion criteria, to show, at least, that there was no inherent bias in the selection of studies.

It is also unclear why they rejected studies where the serial dependence effect on bias was lower than $g = 0.20$. In principle, studies without any perceptible effects of serial dependence on bias would be the ideal situation to test for effects on scatter.

If I am not mistaken, the authors only selected studies for which trial-level data was available. But, if so, why conduct a meta-analysis at all? Would it not make more sense to analyze the data with a (far simpler) multilevel model with participant level data? It is unclear what is gained by using meta-analytic methods, in this case.

Among other advantages, a multilevel analysis would allow to enter stimuli distance as a numerical predictor instead of categorizing trials in three levels (iso, middle, ortho) and conducting pairwise comparisons, losing power and information.

Reviewer #2:
Remarks to the Author:
Review by Matthias Fritsche:

Ozkirli et al. report a meta-analysis of 20 orientation/motion serial dependence datasets, investigating how serial dependence affects response variability across several orientation differences between current and previous stimuli. While optimal integration models predict that response variability should be lowest for successively similar orientations, they find that the empirical estimate is similar to an orthogonal baseline, indicating no change in perceptual performance. Consequently, the authors argue that serial dependence only deteriorates perceptual performance at intermediate orientation differences, and does not improve perceptual decision-making, thus challenging optimal integration models of serial dependence.

This study investigates an important prediction of optimal integration models of serial dependence. I agree with the authors that past research has been very focused on the biasing nature of serial dependence, and has mostly neglected other influences such as effects on response variability. Given that optimal integration models make quantitively testable predictions about response variability, this is an interesting dimension to investigate. I also would like to commend the authors on their efforts to make use of the quite large collection of open-access datasets in this field, demonstrating the usefulness of open science. However, there are a few points, which make me doubt the strength of the presented conclusions, and the challenge to existing models, in particular to Bayesian models of serial dependence. I believe that the manuscript would require more detailed analyses to seriously address these points. Please find my detailed points below.

Major points:

1. The "simple optimal integration" model by Cicchini et al. is not an appropriate model for serial dependencies in circular feature spaces. Presenting the predictions of this model as the predictions of optimal integration (Fig. 1D) and strongly refuting them in the manuscript has a bit of a strawman character. In particular, the model by Cicchini and colleagues is not developed for a circular feature space. As a result, the model severely misestimates biases at near-orthogonal orientation differences, predicting large opposite biases for clockwise and counterclockwise orientations with a discontinuity at +-90 deg (see Supplementary Fig. 1A). This is because the model considers +89 and -89 deg orientations to be maximally distinct (in a linear space) rather than almost the same (in a circular space). The prediction of opposite biases for highly similar orientations is very implausible a priori, and indeed not observed in human data. Therefore, the model deviates from empirical bias patterns at exactly those orientations which are the main focus of the current study (performance baseline). In that sense, it is not surprising that the "simple optimal integration" model does not capture the empirical response scatter either. Since the model does not reflect optimal integration in a circular feature space, I do not think that it should be used as a reference model for predicting error scatter (Fig. 1D), and it should not be used to draw conclusions about optimal integration in general. The authors note themselves that Bayesian models (developed for circular feature spaces) predict a more complicated non-monotonic pattern of response variability. This prediction should be shown in the main manuscript and should be tested against the empirical data.

Related to this point, when showing the predictions of both models in Supplementary Figure 1A and B, I am wondering why they authors plot model predictions for response scatter up to 50 deg when the average empirical response scatter is around ~10 deg and does not exceed 23.34 deg? For the Bayesian model, the examples with unusually high response scatter overemphasize the monotonic increase in response scatter with increasing orientation differences, which would not be expected under more realistic conditions (at empirically observed levels of response scatter). This comes at the expense of clearly visualizing the more nuanced non-monotonic relationship visible at lower levels of response scatter.

2. I feel very uneasy about the analysis choice to select the middle bin based on the peak from the individual error scatter. This procedure systematically biases the middle bin's estimates towards high values, as one always selects the bin with the highest value. Using this procedure on noisy data, I would expect that even in the complete absence of any true variation in response scatter, one would find that the middle bin exhibits larger response scatter than iso and ortho bins, simply due to this biased selection. In the current context, I think this procedure will exaggerate the genuine iso-middle and ortho-middle differences, and generally undermines the authors' statistical conclusions. When choosing the middle bin for a model-free analysis, it should be fixed to a particular orientation difference, in the same way the iso and ortho bins are fixed.

3. I am somewhat confused whether the meta-analysis does or does not support Bayesian integration models. I have previously looked at response variability in one of my papers (Fritsche et al., 2020), of which three studies feature in the current meta-analysis. You can find the response variability patterns in Figure 7—figure supplement 4 in the eLife paper. Two out of the three experiments showed exactly the pattern predicted by the Bayesian model (Exp. 1 and 3). That is, response variability is lowest for iso orientations, intermediate for ortho orientations and highest for middle orientations. In fact, the variability at ortho orientations was equal to the average variability across all orientations. This is exactly what the Bayesian model predicts. I replotted the panels here, together with the Bayesian model which looks remarkably similar: https://drive.google.com/file/d/1DhORgH-Z6VKFEp44tCn16M5RrXTvAqCp/view?usp=share_link. Based on these analyses, I would have concluded that the evidence is in favor rather than against the Bayesian model.

When looking at Fig. 3A of the current study, these two experiments (#10 and #12) appear very similar to many of the other studies,

which are collectively interpreted as evidence against the Bayesian model. To me this is observation is puzzling and makes me very doubtful whether or not the data supports Bayesian integration. I think we would first need to identify how the discrepancy between my previous and the current results arises before drawing strong conclusions in either direction. For instance, in my analyses I did not remove cardinal biases and serial dependence biases prior to computing response variability. It could be easily tested whether these two analysis steps have a sizable impact on the results and could explain the divergent findings. It should also be straightforward to create continuous plots of response variability as a function orientation difference for each of the 20 studies, similar to the plots that I have attached. This would allow to better understand the shapes of the response variability patterns and may clarify the differences between analysis strategies.

In my analyses, one of my experiments does not conform to the Bayesian model's prediction (Exp. 2), and yields more similar patterns to the ones reported by the authors (i.e., similar iso and ortho response variability). Curiously this is not reflected in Fig. 3A, where this studies' iso-ortho difference (#11) is in between studies #10 and #12. Is this due to an inconsistency in labelling, or due to a difference between analyses?

Overall, based on my observations the response variability patterns provide somewhat mixed evidence for a Bayesian model, although I found the similarity between Exp. 1 and 3 and the Bayesian observer quite remarkable. I am puzzled why the current analyses provide seemingly opposite evidence, despite using the same data.

Minor points:

4. I was wondering about the procedure to remove cardinal biases. The fitting of the two polynomials introduces a discontinuity around orientations 0 and +- 90 (cardinals?). This discontinuity is not real, right? There should be a smooth (but steep) transition of the bias around the cardinals. Perhaps I am misunderstanding, but would this procedure not lead to an overcorrection for orientations that are very close to the cardinals?

5. I am slightly confused by the "broader discussion on the concept of optimality and ideal models" (line 445 - 458) in relation to the current findings. Even if the authors would have found support for Bayes-optimal integration, this integration would still be suboptimal in the context of randomized stimulus sequences. It is only optimal when visual input is temporally correlated (and if the observer has accurate knowledge of these correlations). The suboptimality in laboratory settings does not specifically follow from the current findings, as the last sentence of the paragraph seems to suggest.

Furthermore, I don't understand the difference in logic to multisensory integration. Multisensory integration is only optimal when the signals from different modalities in fact originate from the same source, which is often the case in natural environments but not necessarily the case in laboratory settings. Thus, the degree to which the logic of optimal integration holds for temporal and multisensory signals depends on the temporal and spatial correlations of signals in the natural environment of the observer. Overall, it seems that the paragraph is mixing a few ideas from optimal and suboptimal integration to randomized (laboratory) and natural sensory input. I think the discussion would benefit from separating these ideas more clearly.

Reviewer #3:
Remarks to the Author:
Using meta-analysis on serial dependence studies, the authors addressed an important question: How do past-present similarities affect the variability of present perceptual performance? With large data samples from 20 datasets, they showed that the variability of estimates in the current trials is modulated by the stimulus difference between previous and present trials. This modulation occurs not monotonically, but in an inverted u-shape. Although I think the topic is interesting and the results are sound, there is a lack of theoretical explanation for the present findings. Without digging deep into the mechanism underlying these effects, I don't find the present study suitable for publication in Nature Human Behavior. My comments in details are attached below.

When people say "serial dependence," they don't always agree on its definition. Some researchers think it specifically refers to the influence of past stimuli on the present perception when the stimulus similarity between past and present is high. In this condition, it usually leads to an attractive bias. In orientation tasks, the past-present stimulus similarity for the peak bias is less than 30° (e.g., studies by Whitney, Burr, and de Lange groups). In such cases, serial dependence can be well explained by Bayesian integration model. However, "serial dependence" can also refer to the influence of past trial on present perception more generally. Indeed, there is repulsive serial dependence when past-present stimulus similarity is low (e.g., Fritsche, Mostert, & de Lange, 2017, peaked at ~75° difference). Compared to the extensive research on attractive serial dependence induced by high past-present stimulus similarity, the study on serial dependence with low past-present stimulus similarity is much less. Consequently, our understanding of serial dependence under such condition remains limited. Although the low past-present stimulus similarity induces no serial bias (or minor repulsive bias), it does not imply there is no influence from the past on present. The authors' choice to use the "ortho" condition (low stimulus similarity) as the baseline for perceptual performance and compare it with the "iso" condition seems somewhat arbitrary. They find comparable errors between two conditions and argue against the functional advantage of serial dependence under high stimulus similarity condition as predicted by Bayesian integration model. I think the authors need to rule out other possibilities before reaching this conclusion.

1. Serial dependence can occur at multiple cognitive stages. In fact, recent findings suggest that there is an adaptation-like repulsive effect during sensory processing (Sheehan & Serences, 2022; Hajonides et al., 2023; Luo, Zhang, & Luo, 2024). And it has been shown that brief adaptation to orientation can enhance the identification of orthogonal orientations by sharpening neuronal selectivity (Dragoi et al., 2002, nature neuroscience). It is possible that in the "ortho" condition of serial dependence, the present orientation identification is also enhanced by previous orthogonal orientation, resulting in a small error scatter. That it, although both "iso" and "ortho" conditions exhibit lower variability in perceptual estimates, they might be accounted by difference mechanisms, through Bayesian integration and adaptation, respectively.

2. The Bayesian integration model for serial dependence works only if the past and present stimuli are of high similarity. How does the brain deal with the situation that past and present are dissimilar? It's possible that like Bayesian casual inference which is typically demonstrated in multisensory cue combination tasks (Körding et al., 2007), the brain also infers whether past and present

trials belong to a same event first. If they are similar (same event), the brain would integrate past with present, inducing attractive serial dependence. If not, the brain would separate them. Moreover, it has been shown that visual perception is retrospective Bayesian inference from high level to low level (Ding et al., 2017). If the sequentially presented stimuli are different, to preserve this ordinal information, perceptual difference would be exaggerated. Could this operation result in the small error scatter in the "ortho" condition?

3.In either case, there would be potential decrease in the variability of perceptual estimates for "ortho" condition. Indeed, as shown by the authors as well as Cicchini, Mikellidou, and Burr (2018), the peaked variability/"root-mean square error" is around 45° past-present difference, larger than that for the peaked bias (within 30° past-present difference). Is the peak error due to the fact it is far away from both Bayesian integration improvement for "iso" condition and another improvement mechanism for "ortho" condition?

In short, "serial dependence" is a complex phenomenon, which involve interactions between different systems (memory, perception, and decision-making) and multiple computational mechanisms (efficient coding, bayesian integration, or even casual inference). I appreciated the authors effort on studying serial dependence across the whole range of past-present similarity. But I am not convinced by the present finding of comparable perceptual variability between "iso" and "ortho" conditions as evidence against Bayesian integration. A more unified framework is needed.

Version 1:

Decision Letter:

Dear Mr Ozkirli,

Thank you for your correspondence asking us to reconsider our decision on your Article, "Rethinking serial dependence: A meta-analysis of its effects on the variability of perceptual estimates". After careful consideration we have decided that we would be willing to consider a revised version of your manuscript.

Along with your revised manuscript, you should also submit a separate point-by-point response to all of the concerns raised by the referees, in each case describing what changes have been made to the manuscript or, alternatively, if no action has been taken, providing a compelling argument for why that is the case. If we feel that a substantial attempt has been made to address the referees' comments, this response will be sent back to the referees - along with the revised manuscript - so that they can judge whether their concerns have been addressed satisfactorily or otherwise.

When revising your paper:

- ensure it complies with our format requirements as set out in our Guide to Authors.

- state in a cover note the length of the text, methods and figure legends; the number of references and the number of display items.

Please ensure that all correspondence is marked with your Nature Human Behaviour reference number in the subject line.

Please use the following link to submit your revised manuscript:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to receive your revised paper within four weeks. If you cannot send it within this time, please let us know so that we can close your file. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Human Behaviour or published elsewhere in the meantime. Should you miss the four-week deadline and your paper is eventually published, the received date will be that of the revised, not the original, version.

I look forward to hearing from you soon.

Best regards,

██████████
████████

Nature Human Behaviour

Version 2:

Decision Letter:

23rd June 2025

Dear Dr Ozkirli,

Thank you once again for your revised manuscript, entitled "Rethinking serial dependence: A large-scale analysis," and for your patience during the re-review process.

Your manuscript has now been evaluated by the same reviewers who evaluated your original manuscript. All reviewer feedback is included at the end of this letter. Although the reviewers found your manuscript to have improved during revision, they also maintain some important outstanding concerns. We remain interested in the possibility of publishing your study in Nature Human Behaviour, but would like to consider your response to these outstanding concerns in the form of a revised manuscript before we make a decision on publication.

In particular, Reviewer #3 questions whether sufficient evidence has been presented to support the claim that serial dependence deteriorates, rather than improves, perceptual decision-making. We ask you to address this concern in full.

We hope you will find the referees' comments useful as you decide how to proceed. If you wish to submit a substantially revised manuscript, please bear in mind that we will be reluctant to approach the referees again in the absence of major revisions. We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

Finally, your revised manuscript must comply fully with our editorial policies and formatting requirements. Failure to do so will result in your manuscript being returned to you, which will delay its consideration. To assist you in this process, I have attached a checklist that lists all of our requirements. If you have any questions about any of our policies or formatting, please don't hesitate to contact me.

If you wish to submit a suitably revised manuscript, we would hope to receive it within 4 months. I would be grateful if you could contact us as soon as possible if you foresee difficulties with meeting this target resubmission date.

With your revision, please:

• Include a "Response to the editors and reviewers" document detailing, point-by-point, how you addressed each editor and referee comment. If no action was taken to address a point, you must provide a compelling argument. When formatting this document, please respond to each reviewer comment individually, including the full text of the reviewer comment verbatim followed by your response to the individual point. This response will be used by the editors to evaluate your revision and sent back to the reviewers along with the revised manuscript.

• Highlight all changes made to your manuscript or provide us with a version that tracks changes.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

Thank you for the opportunity to review your work. Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further.

Sincerely,

██████████████

Nature Human Behaviour

Reviewer expertise:

Reviewer #1: Meta-analysis, evidence synthesis, perceptual decision-making

Reviewer #2: Serial dependence, perceptual decision-making

Reviewer #3: Serial dependence, perceptual decision-making

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have done a good job at addressing my comments. The present version is far more accessible for non-expert readers

(like me). I also think that the decision to replace the meta-analysis with a trial-level analysis simplifies the text and analyses substantially. Just a few minor comments:

- Title: Consider presenting your work as a "mega-analysis" instead of a "large-scale analysis"

- Figure 1: This is the first time your reader is presented with an experimental paradigm for the study of sequential effects. Please, mention explicitly that the current and previous orientation refer to consecutive trials (not to consecutive stimuli within a trial). This is explained in the main text only afterwards.

- Line 164: why use a standardized effect size (Hedges' g) instead of the raw number of degrees? Given that the experimental paradigms and dependent variables are quite homogeneous, I think it makes sense to report exclusively degrees.

- any evidence of heterogeneity across studies or participants? Although the analysis is no longer presented as a meta-analysis, heterogeneity is still relevant. It can be easily assessed with the random slopes of the LMMs.

Reviewer #2 (Remarks to the Author):

I would like to thank the authors for their thorough revisions and replies to my concerns. I particularly appreciate the detailed control analyses to understand the interactions between stimulus-specific and serial dependence biases. The authors demonstrate quite convincingly that removing versus not removing stimulus-specific biases during preprocessing explains the differences between current and previous results. I have a few follow-up questions related to this explanation and some minor suggestions.

Major points

I can see how response variance can be reduced by accounting for stimulus-specific biases. This explains the orientation-dependent up-down shift in response variance in Figure S3D. However, it does not by itself explain the iso-ortho difference in response scatter. To explain this, the authors assume a history dependency in the stimulus-specific bias, which becomes central to the argument: "This [iso-ortho difference] likely reflects differences in the magnitude of stimulus-specific bias after repeating the same stimulus ($\Delta = 0°$) or changing by 90° ($\Delta = \pm90°$)." Do the authors think that this would occur due to "classical" serial dependence or would this be an additional history-dependency? Wouldn't this reduction of oblique bias for $\Delta = 0°$ constitute a "superiority effect", as responses would be more accurate (less biased) than for $\Delta = \pm90°$? It would be helpful to explain how such a superiority effect would be similar or different to that posited by Bayesian integration. Also, to my knowledge such a history-dependency in stimulus-specific biases has not been previously shown. Therefore, it would be important to assess whether such a history dependence is present in the empirical data.

Furthermore, the authors' procedure for removing stimulus-specific biases is history-free. It removes the average bias, regardless of stimulus history. Therefore, it would not appear to be able to account for the history-dependent variation in stimulus-specific bias magnitude. I suppose the procedure would overestimate stimulus-specific biases for $\Delta = 0°$ and underestimate those for $\Delta = 90°$. Is the idea that an overestimation would inflate variance at $\Delta = 0°$ to account for the failed removal of variance at $\Delta = 90°$? It is not immediately obvious to me that this would work. It would be important to clarify the logic behind why the removal of history-*independent* stimulus-specific biases manages to provide an unbiased estimate of error scatter.

Minor points

It would be interesting to mention separate results for orientation and motion estimation tasks, to show that the results hold for different stimulus features.

There are some errors in labeling the supplemental figures both in captions and in text.

Reviewer #3 (Remarks to the Author):

I appreciate that the authors have expanded their datasets and refined their analysis, which increases my confidence in their findings. As I stated in the first round of review, the authors raise an important question that has been largely overlooked in the serial dependence literature. However, my primary concern remains unaddressed. Only reporting the phenomenon that variance follows an inverted U-shape as past-present similarity decreases provides limited insight without a mechanistic explanation. I believe the conclusion remains overstated. The "inverted U-shape" of variance is insufficient to support the statement that 'serial dependence deteriorates rather than improves perceptual decision-making'.

Given that the authors have clarified they are examining only the attractive form of serial dependence—which manifests when past and present stimuli are similar—the 'ortho' condition is not an appropriate baseline for this analysis. To properly investigate the functional role of serial dependence, the study should compare conditions where past and present stimuli are similar against conditions where no prior influence exists, rather than using orthogonal conditions that are inherently irrelevant to the attractive serial dependence phenomenon. For example, since serial dependence is context-dependent (Fischer et al., 2020), the influence of previous orientation under different contextual conditions might provide a more appropriate baseline for comparison.

The interpretation of serial dependence as 'interference' is speculative and potentially contradictory. The authors claim that 'when consecutive stimuli share identical features, interference is minimal because their representations fully overlap.' However, overlapping representations could plausibly enhance perception by strengthening the signal rather than minimizing interference. This raises a fundamental question: if overlapping representations improve perceptual performance, how can this effect be

characterized as 'minimized interference'? Supporting this alternative interpretation, Cicchini et al. (2018) demonstrated that reaction times decrease when past and present stimuli share identical features, suggesting facilitation rather than interference.

Examining serial dependence across the full range of past-present differences would be more comprehensive, though this would shift the focus away from investigating the functional role of serial dependence specifically. The 'ortho' condition likely involves distinct types of past-trial influences that differ qualitatively from attractive serial dependence. As I reasoned above, the 'iso' condition likely benefits from past information. Since performance (bias and variance) is comparable between 'iso' and 'ortho' conditions, the 'ortho' condition may also benefit from past information, albeit through a different mechanism. Without empirical verification, we cannot assume this condition represents an absence of past-trial influence, and using it directly as a 'no serial dependence' baseline may confound the findings. Thus, there appears to be a gap between their investigation of variance across the full range of past-present differences and their conclusion about whether serial dependence deteriorates or improves perception.

Both behavioral evidence, including work by one of the authors (Pascucci, PLoS Biol, 2019), and recent neural findings (Luo et al., PLoS Biol, 2025; Shan, Hajonides, & Myers, bioRxiv, 2024) indicate that attractive serial dependence is linked to decision-making processes. The decision-making system likely employs flexible strategies to guide current decisions under varying conditions. As noted in my initial review, Bayesian causal inference models provide a standard framework for determining whether two cues originate from the same or different sources (as in the 'ortho' condition here). Furthermore, Ding et al. (PNAS, 2017) demonstrated a repulsive bias in sequential orientation reproduction similar to the repulsive effect observed near the 'ortho' condition (Fig. 3B, upper), which could be successfully explained through retrospective Bayesian inference from higher to lower processing levels.

The Bayesian model is a very general and powerful framework. I agree with the authors that the current Bayesian model of serial dependence is not good enough to explain the patterns of variance under conditions when the past and present are dissimilar. The authors can extend current simple model to provide more mechanistic explanations for various conditions.

Minor:
The model simulation (Fig. 3) is misleading due to scale discrepancies. The simulated bias ranges from 0-10, which is substantially larger than the observed bias range of 0-2. The scales should be made comparable to enable meaningful comparison between model predictions and empirical data.

Version 3:

Decision Letter:

Our ref: NATHUMBEHAV-24020543C

17th September 2025

Dear Dr Ozkirli,

Thank you for submitting your revised manuscript "Rethinking serial dependence: A mega-analysis" (NATHUMBEHAV-24020543C). It has now been seen by two of the original referees and their comments are below. As you can see, the reviewers find that the paper has improved in revision. We will therefore be happy in principle to publish it in Nature Human Behaviour, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements within two weeks. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Please do not hesitate to contact me if you have any questions.

Sincerely,

████████████
████████

Nature Human Behaviour

Reviewer #2 (Remarks to the Author):

Major point 1: The authors have convinced me that the history dependency in stimulus-specific bias and serial dependence are likely two distinct phenomena. This is mainly because of their point I (no correlation between serial dependence and stimulus-specific bias reduction).

Major point 2: I am not convinced by the authors' reply to my comment, but I am inclined to think that this issue does not invalidate their conclusions. I will try to explain what I found unclear, and perhaps it may help to provide a more specific explanation in the

paper.

I am fully on board with the authors' reasoning that their debiasing method removes most of the stimulus-specific bias, and therefore reduces artifactual error scatter in the serial dependence analysis. This is what the authors reiterated in their reply. However, my focus is on the specific question about the iso-ortho difference in error scatter in the serial dependence analysis. The authors claim that this difference occurs because the stimulus-specific bias is slightly reduced in the iso *relative* to the ortho condition, leading to a slightly smaller artifactual increase in error scatter in the iso condition. In their reply, the authors show that this *relative difference* in stimulus-specific bias between iso and ortho remains intact. When the relative difference in bias remains intact, this raises the question how the confound in iso-ortho error scatter in the SD analysis, which came about due to the relative difference in stimulus-specific bias, has been removed. I do not think that the authors have answered this question. Here is my interpretation: After debiasing, there remains a small attraction to the oblique in the ortho condition. This will lead to a small increase in error scatter in the SD analysis. In the iso condition, there now is a small *repulsion* from the obliques, which will also lead to a small increase in error scatter in the SD analysis. Ortho repulsion and iso attraction biases, although of opposite sign, have a similar magnitude and therefore will lead to similar (small) artifactual increases in error scatter. Therefore, by matching the amplitude of iso and ortho biases (but not their relative difference), the small remaining artifactual increase in error scatter cancels out.

Reviewer #3 (Remarks to the Author):

I appreciate the authors' great revisions and clarifications. I have no further questions. Congratulations on the excellent work!

Version 4:

Decision Letter:

Dear Dr Ozkirli,

We are pleased to inform you that your Article "Large-scale mega-analysis indicates that serial dependence deteriorates perceptual decision-making", has now been accepted for publication in Nature Human Behaviour.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles or the NIH public access policy) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement – including the applicable embargo period – publication under the subscription model isn't suitable for authors whose funders require no embargo.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see http://www.nature.com/nathumbehav/info/gta). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

We welcome the submission of potential cover material (including a short caption of around 40 words) related to your manuscript; suggestions should be sent to Nature Human Behaviour as electronic files (the image should be 300 dpi at 210 x 297 mm in either TIFF or JPEG format). Please note that such pictures should be selected more for their aesthetic appeal than for their scientific content, and that colour images work better than black and white or grayscale images. Please do not try to design a cover with the Nature Human Behaviour logo etc., and please do not submit composites of images related to your work. I am sure you will understand that we cannot make any promise as to whether any of your suggestions might be selected for the cover of the journal.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a

subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

In approximately 10 business days you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

We look forward to publishing your paper.

With best regards,

███████████
██████████
Nature Human Behaviour

P.S. Click on the following link if you would like to recommend Nature Human Behaviour to your librarian
http://www.nature.com/subscriptions/recommend.html#forms

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**

Response to Reviewers

Reviewer #1

Remarks to the Author:
The authors should explain in more detail what serial dependence is, having a general audience in mind. For instance, the opening sentence of the abstract takes for granted that the reader is already familiar with these effects. This is unlikely the case for the average reader of NHB. In general, the abstract is almost impossible to understand for a researcher that has never come across this literature before. The first paragraph of the introduction doesn't offer much more information. The authors should make an effort to adjust the text to the general audience of the journal.

We would like to thank the Reviewer for this suggestion, which has considerably helped to improve our manuscript. We have now largely revised our manuscript, starting from the abstract and introduction, with the main aim of making it easily accessible to a broader range of readers, such as those of *Nature Human Behavior*.

I couldn't find anywhere information about how the authors identified the studies that they entered into the meta-analysis. The should explain the literature search strategy and the inclusion/exclusion criteria, to show, at least, that there was no inherent bias in the selection of studies.

We thank the Reviewer for this very important point. In the revised version, we added a dedicated section explaining in detail the criteria and steps in the identification of relevant studies (*Dataset selection and exclusion criteria*). As now clarified, the datasets were selected and included based on specific criteria devised to avoid any inherent bias:
"*We conducted a search on PubMed using the keyword "serial dependence" for studies published between 2014 (from the initial study of Fischer and Whitney) and 2024. Studies were selected based on the following criteria:*
1. *Single-trial data publicly available online,*
2. *Adjustment tasks involving orientation or motion stimuli with circular responses (e.g., participants reproducing orientation of a Gabor patch or motion direction of dot clouds),*
3. *Trial-by-trial orientation or motion direction changes (Δ) covering 0 to ±90°, using uniform/randomly determined trial-by-trial changes in Δ.*"

It is also unclear why they rejected studies where the serial dependence effect on bias was lower than g = 0.20. In principle, studies without any perceptible effects of serial dependence on bias would be the ideal situation to test for effects on scatter.

We thank the Reviewer for the opportunity to clarify this point. The main aim of our analysis was to assess the effects of **positive** serial dependence on scatter. Initially, we applied an additional exclusion criterion to include only datasets that exhibited positive serial dependence effects in the bias measure. This decision was based on previous studies suggesting that the effects on bias and scatter are linked to the same underlying mechanism supporting a beneficial role of serial dependence.

However, in accordance with the Reviewer's request, we have now relaxed this assumption and included all available datasets, regardless of the strength of the serial dependence bias.

If I am not mistaken, the authors only selected studies for which trial-level data was available. But, if so, why conduct a meta-analysis at all? Would it not make more sense to analyze the data with a (far simpler) multilevel model with participant level data? It is unclear what is gained by using meta-analytic methods, in this case.
Among other advantages, a multilevel analysis would allow to enter stimuli distance as a numerical predictor instead of categorizing trials in three levels (iso, middle, ortho) and conducting pairwise comparisons, losing power and information.

Following the Reviewer's request, we have substantially revised our analysis approach and now employ a multilevel model with a maximal random effects structure (see Methods and Results). We appreciate the Reviewer's suggestion, as this approach is indeed better suited to our analysis and has improved the clarity of the results.

Reviewer #2

Remarks to the Author:
Review by Matthias Fritsche:

Ozkirli et al. report a meta-analysis of 20 orientation/motion serial dependence datasets, investigating how serial dependence affects response variability across several orientation differences between current and previous stimuli. While optimal integration models predict that response variability should be lowest for successively similar orientations, they find that the empirical estimate is similar to an orthogonal baseline, indicating no change in perceptual performance. Consequently, the authors argue that serial dependence only deteriorates perceptual performance at intermediate orientation differences, and does not improve perceptual decision-making, thus challenging optimal integration models of serial dependence.
This study investigates an important prediction of optimal integration models of serial dependence. I agree with the authors that past research has been very focused on the biasing nature of serial dependence, and has mostly neglected other influences such as effects on response variability. Given that optimal integration models make quantitively testable predictions about response variability, this is an interesting dimension to investigate. I also would like to commend the authors on their efforts to make use of the quite large collection of open-access datasets in this field, demonstrating the usefulness of open science. However, there are a few points, which make me doubt the strength of the presented conclusions, and the challenge to existing models, in particular to Bayesian models of serial dependence. I believe that the manuscript would require more detailed analyses to seriously address these points. Please find my detailed points below.

We sincerely appreciate the Reviewer's thorough feedback and the quality of his comments, which prompted us to conduct additional detailed analyses. As outlined below, this led us to identify the

precise nature of the confounding effect of stimulus-specific biases in estimating error scatter—a critical aspect of this type of analysis that was previously unknown. This has significantly strengthened our confidence in the results, and we hope the Reviewer will find the analyses reported in the Supplementary Materials to be a decisive addition, not only for interpreting the current results but also for guiding future research on serial dependence in addressing similar problems.

Major points:

1. The "simple optimal integration" model by Cicchini et al. is not an appropriate model for serial dependencies in circular feature spaces. Presenting the predictions of this model as the predictions of optimal integration (Fig. 1D) and strongly refuting them in the manuscript has a bit of a strawman character. In particular, the model by Cicchini and colleagues is not developed for a circular feature space. As a result, the model severely misestimates biases at near-orthogonal orientation differences, predicting large opposite biases for clockwise and counterclockwise orientations with a discontinuity at +-90 deg (see Supplementary Fig. 1A). This is because the model considers +89 and -89 deg orientations to be maximally distinct (in a linear space) rather than almost the same (in a circular space). The prediction of opposite biases for highly similar orientations is very implausible a priori, and indeed not observed in human data. Therefore, the model deviates from empirical bias patterns at exactly those orientations which are the main focus of the current study (performance baseline). In that sense, it is not surprising that the "simple optimal integration" model does not capture the empirical response scatter either. Since the model does not reflect optimal integration in a circular feature space, I do not think that it should be used as a reference model for predicting error scatter (Fig. 1D), and it should not be used to draw conclusions about optimal integration in general. The authors note themselves that Bayesian models (developed for circular feature spaces) predict a more complicated non-monotonic pattearn of response variability. This prediction should be shown in the main manuscript and should be tested against the empirical data.

   We fully agree that the "simple optimal integration" model is limited in capturing the complex patterns in serial dependence. Therefore, we have now included the Bayesian model as a reference model upfront in the main text. Since the cue integration model remains a well-recognized reference in the field, we have decided to retain it alongside the Bayesian model as the two primary reference models. As the Reviewer will see, in the revised version of the manuscript, the main conclusions remain unchanged when comparing the observed pattern in empirical data against either model.

   Related to this point, when showing the predictions of both models in Supplementary Figure 1A and B, I am wondering why they authors plot model predictions for response scatter up to 50 deg when the average empirical response scatter is around ~10 deg and does not exceed 23.34 deg? For the Bayesian model, the examples with unusually high response scatter overemphasize the monotonic increase in response scatter with increasing orientation differences, which would not be expected under more realistic conditions (at empirically observed levels of response scatter). This comes at the expense of clearly visualizing the more nuanced non-monotonic relationship visible at lower levels of response scatter.

   Following the Reviewer's comments, this aspect of the study has been substantially revised, both in the main text and Supplementary Material. About this specific point, we now present simulations in the main text, obtained using σ values ranging from 5° to 20°, in 1° increments. To ensure a clear visualization of superiority effects across different σ values, we normalize the

4

error scatter by dividing all values across Δ by the corresponding simulated σ value (see Figure 3A).

2. I feel very uneasy about the analysis choice to select the middle bin based on the peak from the individual error scatter. This procedure systematically biases the middle bin's estimates towards high values, as one always selects the bin with the highest value. Using this procedure on noisy data, I would expect that even in the complete absence of any true variation in response scatter, one would find that the middle bin exhibits larger response scatter than iso and ortho bins, simply due to this biased selection. In the current context, I think this procedure will exaggerate the genuine iso-middle and ortho-middle differences, and generally undermines the authors' statistical conclusions. When choosing the middle bin for a model-free analysis, it should be fixed to a particular orientation difference, in the same way the iso and ortho bins are fixed.

We thank the Reviewer for this comment. In response to the concerns raised here and by Reviewer 1, we have adopted a different approach using a continuous fit of the data, thereby avoiding any arbitrary selection of bins in the main analysis. Additionally, we now present the bin analysis as a control in the Supplementary Material, implementing a fixed mid-bin as suggested.

3. I am somewhat confused whether the meta-analysis does or does not support Bayesian integration models. I have previously looked at response variability in one of my papers (Fritsche et al., 2020), of which three studies feature in the current meta-analysis. You can find the response variability patterns in Figure 7—figure supplement 4 in the eLife paper. Two out of the three experiments showed exactly the pattern predicted by the Bayesian model (Exp. 1 and 3). That is, response variability is lowest for iso orientations, intermediate for ortho orientations and highest for middle orientations. In fact, the variability at ortho orientations was equal to the average variability across all orientations. This is exactly what the Bayesian model predicts. I replotted the panels here, together with the Bayesian model which looks remarkably similar: https://drive.google.com/file/d/1DhORgH-Z6VKFEp44tCn16M5RrXTvAqCp/view?usp=share_link. Based on these analyses, I would have concluded that the evidence is in favor rather than against the Bayesian model.
When looking at Fig. 3A of the current study, these two experiments (#10 and #12) appear very similar to many of the other studies, which are collectively interpreted as evidence against the Bayesian model. To me this is observation is puzzling and makes me very doubtful whether or not the data supports Bayesian integration. I think we would first need to identify how the discrepancy between my previous and the current results arises before drawing strong conclusions in either direction. For instance, in my analyses I did not remove cardinal biases and serial dependence biases prior to computing response variability. It could be easily tested whether these two analysis steps have a sizable impact on the results and could explain the divergent findings. It should also be straightforward to create continuous plots of response variability as a function orientation difference for each of the 20 studies, similar to the plots that I have attached. This would allow to better understand the shapes of the response variability patterns and may clarify the differences between analysis strategies.
In my analyses, one of my experiments does not conform to the Bayesian model's prediction (Exp. 2), and yields more similar patterns to the ones reported by the authors (i.e., similar iso and ortho response variability). Curiously this is not reflected in Fig. 3A, where this studies' iso-ortho difference (#11) is in between studies #10 and #12. Is this due to an inconsistency in labelling, or due to a difference between analyses?
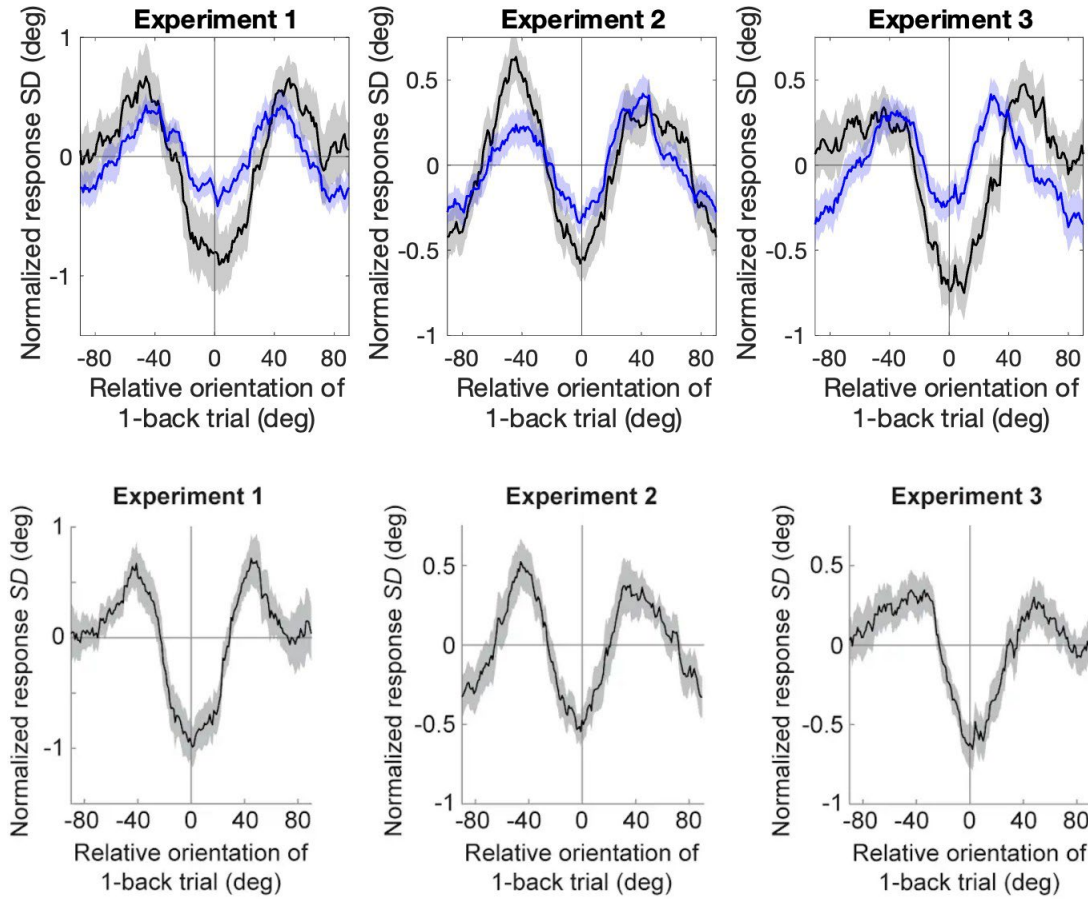
Overall, based on my observations the response variability patterns provide somewhat mixed evidence for a Bayesian model, although I found the similarity between Exp. 1 and 3 and the Bayesian observer quite remarkable. I am puzzled why the current analyses provide seemingly opposite evidence, despite using the same data.

We are very grateful to the Reviewer for motivating this further analysis and providing us with the opportunity to examine the problem in greater detail. Indeed, we agree that previous reports on patterns of error scatter provide a mixed picture—some appearing consistent with the superiority effect discussed in our manuscript, such as those mentioned by the Reviewer.

As the Reviewer correctly noted, our analysis of the same data did not reveal the same patterns. To investigate the factors driving this discrepancy, we started by examining the main difference in data preprocessing—specifically, the removal of stimulus-specific biases (or 'orientation biases'), which was not applied in the original study. This led to a series of striking observations and the identification of a critical confound in estimating error scatter as a function of serial dependence effects. That is, two sources of bias—the stimulus-specific and the serial dependence bias—previously considered independent in many existing studies, interact in problematic ways. This interaction can lead to spurious estimates of error scatter, creating apparent superiority effects that are purely artifacts of this confound.

In the *Confounding Effect of Stimulus-Specific Biases* section of the Supplementary Material we provide a clear demonstration of this issue, with detailed description of our analyses and findings, which we invite the Reviewer to carefully assess.

Once we removed orientation biases, the superiority effect largely disappeared across datasets, including those from Fritsche et al. (2020). **Below, we provide a figure showing the results with the original datasets** (first row: our analysis; black curve: minimal cleaning using 3*std; blue curve: stimulus-specific bias, i.e., orientation bias, removed; second row: plots provided by the Reviewer).

Thus, thanks to this analysis, we are now more confident than before that previously reported superiority effects are merely artifacts of stimulus-specific biases. We demonstrate this through a series of targeted control analyses, and we believe these findings will be of key importance not only in supporting the conclusions of our study but also for research on serial dependence in circular feature domains more broadly. This confounding interaction has remained entirely hidden in prior studies, and our work highlights the fundamental need to account for it in future investigations focusing on effects of serial dependence on error scatter.

Minor points:

4. I was wondering about the procedure to remove cardinal biases. The fitting of the two polynomials introduces a discontinuity around orientations 0 and +- 90 (cardinals?). This discontinuity is not real, right? There should be a smooth (but steep) transition of the bias around the cardinals. Perhaps I am misunderstanding, but would this procedure not lead to an overcorrection for orientations that are very close to the cardinals?

    We thank the Reviewer for this comment. To address this concern, we have refined our method for removing stimulus-specific biases by implementing circular regression with sine and cosine basis functions. For details, please refer to the *Removal of Stimulus-Specific Biases* section in the Methods. Importantly, our results and conclusions remain unchanged, despite the inclusion of new datasets and the use of different bias removal methods across versions.

5. I am slightly confused by the "broader discussion on the concept of optimality and ideal models" (line 445 - 458) in relation to the current findings. Even if the authors would have found support for Bayes-optimal integration, this integration would still be suboptimal in the context of randomized stimulus sequences. It is only optimal when visual input is temporally correlated (and if the observer has accurate knowledge of these correlations). The suboptimality in laboratory settings does not specifically follow from the current findings, as the last sentence of the paragraph seems to suggest.

Furthermore, I don't understand the difference in logic to multisensory integration. Multisensory integration is only optimal when the signals from different modalities in fact originate from the same source, which is often the case in natural environments but not necessarily the case in laboratory settings. Thus, the degree to which the logic of optimal integration holds for temporal and multisensory signals depends on the temporal and spatial correlations of signals in the natural environment of the observer.

Overall, it seems that the paragraph is mixing a few ideas from optimal and suboptimal integration to randomized (laboratory) and natural sensory input. I think the discussion would benefit from separating these ideas more clearly.

We agree with all the points raised by the Reviewer in this comment and apologize for the lack of clarity in this section of our manuscript. We have now revised the text to improve clarity and provide a more detailed discussion. Specifically, we elaborate on why integration can be beneficial when signals in natural settings originate from the same source but may lead to interference in laboratory settings where signals are uncorrelated.

Additionally, we now highlight findings demonstrating reduced or even repulsive (rather than stronger) serial dependence in experiments that introduce temporal correlations in stimulus sequences. These results further challenge some of the dominant views on serial dependence and its functional role. Lastly, we have avoided the use of the term optimal, as it is not essential to our argument or main conclusions.

Reviewer #3

Remarks to the Author:

Using meta-analysis on serial dependence studies, the authors addressed an important question: How do past-present similarities affect the variability of present perceptual performance? With large data samples from 20 datasets, they showed that the variability of estimates in the current trials is modulated by the stimulus difference between previous and present trials. This modulation occurs not monotonically, but in an inverted u-shape. Although I think the topic is interesting and the results are sound, there is a lack of theoretical explanation for the present findings. Without digging deep into the mechanism underlying these effects, I don't find the present study suitable for publication in Nature Human Behavior. My comments in details are attached below.

We appreciate the Reviewer's general evaluation and comment. We have now revised substantially all text and presented new analyses and results, as well as a more in-depth discussion of the potential mechanisms underlying these effects. We believe that these additional analyses and the expanded discussion directly address the Reviewer's concerns and provide a stronger theoretical context for our results.

1. When people say "serial dependence," they don't always agree on its definition. Some researchers think it specifically refers to the influence of past stimuli on the present perception when the stimulus similarity between past and present is high. In this condition, it usually leads to an attractive bias. In orientation tasks, the past-present stimulus similarity for the peak bias is less than 30° (e.g., studies by Whitney, Burr, and de Lange groups). In such cases, serial dependence can be well explained by Bayesian integration model. However, "serial dependence" can also refer to the influence of past trial on present perception more generally. Indeed, there is repulsive serial dependence when past-present stimulus similarity is low (e.g., Fritsche, Mostert, & de Lange, 2017, peaked at ~75° difference). Compared to the extensive research on attractive serial dependence induced by high past-present stimulus similarity, the study on serial dependence with low past-present stimulus similarity is much less. Consequently, our understanding of serial dependence under such condition remains limited. Although the low past-present stimulus similarity induces no serial bias (or minor repulsive bias), it does not imply there is no influence from the past on present. The authors' choice to use the "ortho" condition (low stimulus similarity) as the baseline for perceptual performance and compare it with the "iso" condition seems somewhat arbitrary. They find comparable errors between two conditions and argue against the functional advantage of serial dependence under high stimulus similarity condition as predicted by Bayesian integration model. I think the authors need to rule out other possibilities before reaching this conclusion.

We thank the Reviewer for this comment. We agree that the term serial dependence should not be used synonymously with a specific form of the effect or interpreted as implying a particular underlying mechanism. As clarified in Pascucci et al. (2023), the term refers to a statistical definition and should remain neutral with respect to its functional interpretation. To avoid ambiguity, we now explicitly mention that the form of serial dependence examined in this study is the *attractive* form—the one central to Bayesian and cue integration models predicting superiority effects. Testing the predictions of these models and the associated theoretical views is the core focus of our study.

We would also like to clarify that the majority of studies on serial dependence (both attractive and repulsive), including all studies considered here by selection criteria, assess potential effects at both high and low levels of stimulus similarity. This is achieved by systematically testing serial dependence across the full orientation/motion direction angular space. Most of these studies show attractive serial dependence for high stimulus similarity and weakly repulsive or absent effects for low stimulus similarity. This well-established pattern has also been emphasized in the meta-analysis and review by Manassi et al. (2024). Thus, the finding that attractive serial dependence primarily manifests at high stimulus similarity is not a novel or debated aspect, nor is the limited effect at large stimulus differences an 'unexplored' phenomenon.

In the current version of the manuscript, we no longer split angular differences into bins, so the 'ortho' condition is no longer arbitrarily defined (except in control analyses reported in the Supplementary Material). However, we acknowledge the Reviewer's concern that defining a baseline in serial dependence (e.g., considering the 'ortho' condition as a true baseline) is problematic, as a true baseline would imply no influence of prior stimuli. A similar issue has been addressed in spatial processing and crowding research, where it has been shown that spatial context—assumed to induce similar beneficial effects as serial dependence—does not improve performance beyond a true baseline (i.e., performance without contextual stimuli, a more clearly defined reference condition) (Cicchini et al., 2022; Ozkirli et al., 2025). We now incorporate this converging line of research into our discussion.

Nonetheless, the core prediction derived from Bayesian and cue integration models remains unchanged: a clear difference in error scatter is expected between highly similar and highly dissimilar (e.g., iso vs. ortho) angular differences. This is the prediction we directly test, and we believe our methodological approach is well-suited for this purpose. Furthermore, we have now included rigorous control analyses (see Supplementary Material) to rule out other potential confounds.

2. Serial dependence can occur at multiple cognitive stages. In fact, recent findings suggest that there is an adaptation-like repulsive effect during sensory processing (Sheehan & Serences, 2022; Hajonides et al., 2023; Luo, Zhang, & Luo, 2024). And it has been shown that brief adaptation to orientation can enhance the identification of orthogonal orientations by sharpening neuronal selectivity (Dragoi et al., 2002, nature neuroscience). It is possible that in the "ortho" condition of serial dependence, the present orientation identification is also enhanced by previous orthogonal orientation, resulting in a small error scatter. That it, although both "iso" and "ortho" conditions exhibit lower variability in perceptual estimates, they might be accounted by difference mechanisms, through Bayesian integration and adaptation, respectively.

We agree with the Reviewer's point, and indeed, one of the authors was among the first to propose a repulsive low-level adaptation stage (Pascucci et al., 2019). However, the context here is markedly different from studies reporting beneficial effects of adaptation at orientations orthogonal to the adapter. In classic adaptation paradigms, the adapter is a strong stimulus presented for several seconds—sometimes minutes—often followed by an adaptation top-up before testing its effects on a target stimulus. In contrast, the majority of studies considered here were specifically designed to minimize adaptation effects by using brief, low-contrast stimuli, mostly followed by a noise mask. Thus, we believe that adaptation at large orientation differences is unlikely to explain the observed pattern of error scatter.

Nonetheless, as this remains a possibility, we now explicitly address it in the Discussion section. Importantly, the explanation suggested by the Reviewer would require an interaction between two opposing effects that cancel out the superiority effect. This is a rather complex hypothesis, and in the absence of unequivocal evidence, a simpler explanation is preferable (following Occam's Razor). However, even if such an interplay were present, it would not change the core conclusion: there is no evidence of superiority effects in serial dependence. Regardless of whether a single or multiple mechanisms are involved, performance does not improve for highly similar stimuli as predicted by cue integration and Bayesian models.

3. The Bayesian integration model for serial dependence works only if the past and present stimuli are of high similarity. How does the brain deal with the situation that past and present are dissimilar? It's possible that like Bayesian casual inference which is typically demonstrated in multisensory cue combination tasks (Körding et al., 2007), the brain also infers whether past and present trials belong to a same event first. If they are similar (same event), the brain would integrate past with present, inducing attractive serial dependence. If not, the brain would separate them. Moreover, it has been shown that visual perception is retrospective Bayesian inference from high level to low level (Ding et al., 2017). If the sequentially presented stimuli are different, to preserve this ordinal information, perceptual difference would be exaggerated. Could this operation result in the small error scatter in the "ortho" condition?

This is an interesting hypothesis that can be addressed in future research, aligning with our broader aim of moving beyond existing models to better understand serial dependence. Emerging alternative models, which we now reference, are indeed shifting towards framing the problem in terms of inferring whether two stimuli originate from a common or separate source (e.g., Chetverikov, 2023).

However, before adopting this perspective, a simpler explanation must be ruled out—namely, that stimuli at intermediate differences interfere with the representation of the current stimulus. Following the Reviewer's suggestion, we now discuss alternative models and the challenge of inferring common sources. Additionally, we highlight a key empirical finding: in tasks where stimulus sequences are correlated—thus promoting the inference of a common source from trial to trial—attractive serial dependence is actually reduced rather than enhanced. This suggests a completely different strategy, in which the visual system exploits temporal correlations to better discriminate consecutive stimuli (Blondé et al., 2023; Ufer & Blank, 2024).

4. In either case, there would be potential decrease in the variability of perceptual estimates for "ortho" condition. Indeed, as shown by the authors as well as Cicchini, Mikellidou, and Burr (2018), the peaked variability/"root-mean square error" is around 45° past-present difference, larger than that for the peaked bias (within 30° past-present difference). Is the peak error due to the fact it is far away from both Bayesian integration improvement for "iso" condition and another improvement mechanism for "ortho" condition?

It is unclear whether the Reviewer is referring to the peak in variability or the peak amplitude in RMSE (defined as $\sqrt{\overline{bbbbbbbss^2} + ssssbbssssssrr^2}$, Eq. 3.4 in Cicchini et al., 2018). However, upon close examination of the figures in Cicchini et al. (2018), we respectfully disagree with the Reviewer's interpretation:
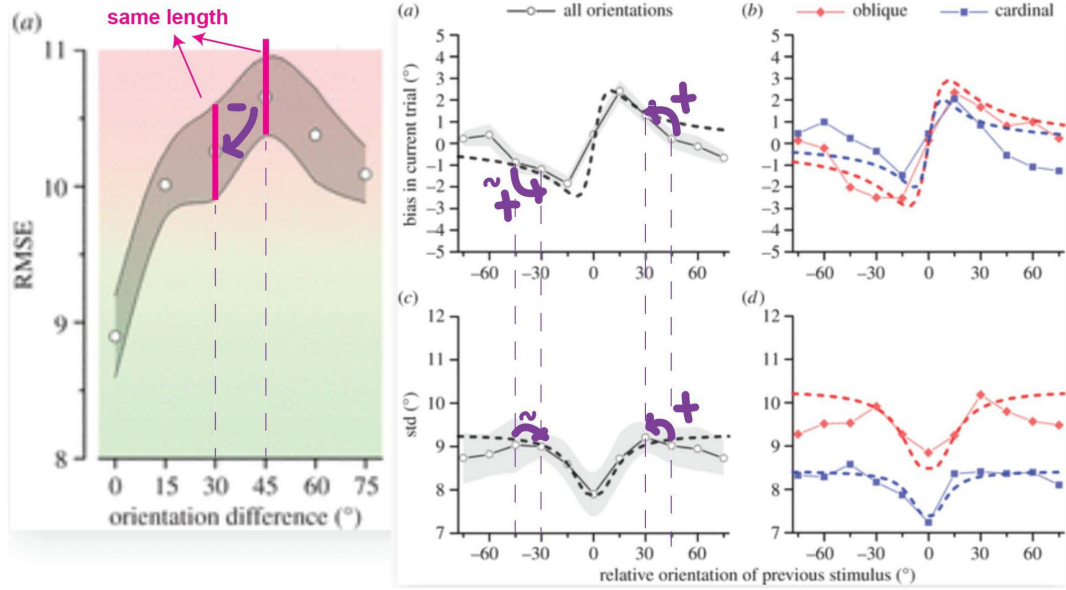- **If the reference is to RMSE peak variability:** the peak does not occur at Δ = 45°, but rather around Δ = 30°. For reference, we have attached a screenshot below with pink lines of equal length, clearly illustrating that variability at 30° exceeds that at 45°.
- **If the reference is to the RMSE peak value:** while it appears to peak at Δ = 45°, this is most likely a misreport in Cicchini et al. (2018). Both bias and scatter are greater at Δ = 30° than at 45° (see purple lines and plus signs in the figure below). Given that RMSE is computed as $\sqrt{\overline{bbbbbbss^2} + ssssbbssssssrr^2}$, a peak at 45° is not mathematically supported. Moreover, Cicchini et al. (2018) did not include data for Δ = 90°, despite its presence in their publicly available dataset.

Given these inconsistencies between the Reviewer's claim and the actual data, we suspect the perception of the situation may have been influenced by their hypothesis regarding a secondary superiority effect at orthogonal distances. Regardless, it is important to note that even Bayesian models—on which our manuscript focuses—do not inherently predict either a peak in bias and scatter at the same Δ, nor a superiority effect in the orthogonal condition.

cues. Overall the total squared error is given by their sum (**figure 2***h*):

$$ERR = BIAS^2 + VAR, \qquad 3.4$$

To the best of our ability, we do not see a principled reason why error scatter should decrease at ortho unless invoking complex and unverified explanations. We explicitly present the predictions of the models referenced by the Reviewer in the main text, for both bias and scatter. These models do not inherently predict a decrease in error scatter at ortho to the level of iso. Since the primary goal of our study is to test the predicted superiority effect (see also Ozkirli et al., 2025), we prefer not to introduce or test speculative explanations that have not been explicitly formalized or validated in prior research.

In short, "serial dependence" is a complex phenomenon, which involve interactions between different systems (memory, perception, and decision-making) and multiple computational mechanisms (efficient coding, bayesian integration, or even casual inference). I appreciated the authors effort on studying serial dependence across the whole range of past-present similarity. But I am not convinced by the present finding of comparable perceptual variability between "iso" and "ortho" conditions as evidence against Bayesian integration. A more unified framework is needed.

We partly agree with the Reviewer closing remark, particularly regarding the complexity of serial dependence and the limitations of current models—especially those assuming superiority effects. However, we do not take for granted the involvement of all suggested models until they have been properly tested and systematically compared.

Our study contributes to this effort by encouraging future research to rethink serial dependence in light of alternative frameworks. A 'more unified' framework can only be achieved after a proper mechanistic and computational understanding of the phenomenon at hand. In no way should this involve neglecting observations and patterns that deviate from current models.

Response to Reviewers

Reviewer #1

Remarks to the Author:

The authors should explain in more detail what serial dependence is, having a general audience in mind. For instance, the opening sentence of the abstract takes for granted that the reader is already familiar with these effects. This is unlikely the case for the average reader of NHB. In general, the abstract is almost impossible to understand for a researcher that has never come across this literature before. The first paragraph of the introduction doesn't offer much more information. The authors should make an effort to adjust the text to the general audience of the journal.

We would like to thank the Reviewer for this suggestion, which has considerably helped to improve our manuscript. We have now largely revised our manuscript, starting from the abstract and introduction, with the main aim of making it easily accessible to a broader range of readers, such as those of *Nature Human Behavior*.

I couldn't find anywhere information about how the authors identified the studies that they entered into the meta-analysis. The should explain the literature search strategy and the inclusion/exclusion criteria, to show, at least, that there was no inherent bias in the selection of studies.

We thank the Reviewer for this very important point. In the revised version, we added a dedicated section explaining in detail the criteria and steps in the identification of relevant studies (*Dataset selection and exclusion criteria*). As now clarified, the datasets were selected and included based on specific criteria devised to avoid any inherent bias:

"*We conducted a search on PubMed using the keyword "serial dependence" for studies published between 2014 (from the initial study of Fischer and Whitney) and 2024. Studies were selected based on the following criteria:*
  1. *Single-trial data publicly available online,*
  2. *Adjustment tasks involving orientation or motion stimuli with circular responses (e.g., participants reproducing orientation of a Gabor patch or motion direction of dot clouds),*
  3. *Trial-by-trial orientation or motion direction changes (Δ) covering 0 to $\pm 90°$, using uniform/randomly determined trial-by-trial changes in Δ.*"

It is also unclear why they rejected studies where the serial dependence effect on bias was lower than g = 0.20. In principle, studies without any perceptible effects of serial dependence on bias would be the ideal situation to test for effects on scatter.

We thank the Reviewer for the opportunity to clarify this point. The main aim of our analysis was to assess the effects of **positive** serial dependence on scatter. Initially, we applied an additional exclusion criterion to include only datasets that exhibited positive serial dependence effects in the bias measure. This decision was based on previous studies suggesting that the effects on bias and scatter are linked to the same underlying mechanism supporting a beneficial role of serial dependence.

However, in accordance with the Reviewer's request, we have now relaxed this assumption and included all available datasets, regardless of the strength of the serial dependence bias.

If I am not mistaken, the authors only selected studies for which trial-level data was available. But, if so, why conduct a meta-analysis at all? Would it not make more sense to analyze the data with a (far

simpler) multilevel model with participant level data? It is unclear what is gained by using meta-analytic methods, in this case.

Among other advantages, a multilevel analysis would allow to enter stimuli distance as a numerical predictor instead of categorizing trials in three levels (iso, middle, ortho) and conducting pairwise comparisons, losing power and information.

Following the Reviewer's request, we have substantially revised our analysis approach and now employ a multilevel model with a maximal random effects structure (see Methods and Results). We appreciate the Reviewer's suggestion, as this approach is indeed better suited to our analysis and has improved the clarity of the results.

Reviewer #2

Remarks to the Author:
Review by Matthias Fritsche:

Ozkirli et al. report a meta-analysis of 20 orientation/motion serial dependence datasets, investigating how serial dependence affects response variability across several orientation differences between current and previous stimuli. While optimal integration models predict that response variability should be lowest for successively similar orientations, they find that the empirical estimate is similar to an orthogonal baseline, indicating no change in perceptual performance. Consequently, the authors argue that serial dependence only deteriorates perceptual performance at intermediate orientation differences, and does not improve perceptual decision-making, thus challenging optimal integration models of serial dependence.

This study investigates an important prediction of optimal integration models of serial dependence. I agree with the authors that past research has been very focused on the biasing nature of serial dependence, and has mostly neglected other influences such as effects on response variability. Given that optimal integration models make quantitively testable predictions about response variability, this is an interesting dimension to investigate. I also would like to commend the authors on their efforts to make use of the quite large collection of open-access datasets in this field, demonstrating the usefulness of open science. However, there are a few points, which make me doubt the strength of the presented conclusions, and the challenge to existing models, in particular to Bayesian models of serial dependence. I believe that the manuscript would require more detailed analyses to seriously address these points. Please find my detailed points below.

We sincerely appreciate the Reviewer's thorough feedback and the quality of his comments, which prompted us to conduct additional detailed analyses. As outlined below, this led us to identify the precise nature of the confounding effect of stimulus-specific biases in estimating error scatter—a critical aspect of this type of analysis that was previously unknown. This has significantly strengthened our confidence in the results, and we hope the Reviewer will find the analyses reported in the Supplementary Materials to be a decisive addition, not only for interpreting the current results but also for guiding future research on serial dependence in addressing similar problems.

Major points:

1. The "simple optimal integration" model by Cicchini et al. is not an appropriate model for serial dependencies in circular feature spaces. Presenting the predictions of this model as the predictions of optimal integration (Fig. 1D) and strongly refuting them in the manuscript has a bit of a strawman character. In particular, the model by Cicchini and colleagues is not developed

for a circular feature space. As a result, the model severely misestimates biases at near-orthogonal orientation differences, predicting large opposite biases for clockwise and counterclockwise orientations with a discontinuity at +-90 deg (see Supplementary Fig. 1A). This is because the model considers +89 and -89 deg orientations to be maximally distinct (in a linear space) rather than almost the same (in a circular space). The prediction of opposite biases for highly similar orientations is very implausible a priori, and indeed not observed in human data. Therefore, the model deviates from empirical bias patterns at exactly those orientations which are the main focus of the current study (performance baseline). In that sense, it is not surprising that the "simple optimal integration" model does not capture the empirical response scatter either. Since the model does not reflect optimal integration in a circular feature space, I do not think that it should be used as a reference model for predicting error scatter (Fig. 1D), and it should not be used to draw conclusions about optimal integration in general. The authors note themselves that Bayesian models (developed for circular feature spaces) predict a more complicated non-monotonic pattearn of response variability. This prediction should be shown in the main manuscript and should be tested against the empirical data.

We fully agree that the "simple optimal integration" model is limited in capturing the complex patterns in serial dependence. Therefore, we have now included the Bayesian model as a reference model upfront in the main text. Since the cue integration model remains a well-recognized reference in the field, we have decided to retain it alongside the Bayesian model as the two primary reference models. As the Reviewer will see, in the revised version of the manuscript, the main conclusions remain unchanged when comparing the observed pattern in empirical data against either model.

Related to this point, when showing the predictions of both models in Supplementary Figure 1A and B, I am wondering why they authors plot model predictions for response scatter up to 50 deg when the average empirical response scatter is around ~10 deg and does not exceed 23.34 deg? For the Bayesian model, the examples with unusually high response scatter overemphasize the monotonic increase in response scatter with increasing orientation differences, which would not be expected under more realistic conditions (at empirically observed levels of response scatter). This comes at the expense of clearly visualizing the more nuanced non-monotonic relationship visible at lower levels of response scatter.

Following the Reviewer's comments, this aspect of the study has been substantially revised, both in the main text and Supplementary Material. About this specific point, we now present simulations in the main text, obtained using σ values ranging from 5° to 20°, in 1° increments. To ensure a clear visualization of superiority effects across different σ values, we normalize the error scatter by dividing all values across Δ by the corresponding simulated σ value (see Figure 3A).

2. I feel very uneasy about the analysis choice to select the middle bin based on the peak from the individual error scatter. This procedure systematically biases the middle bin's estimates towards high values, as one always selects the bin with the highest value. Using this procedure on noisy data, I would expect that even in the complete absence of any true variation in response scatter, one would find that the middle bin exhibits larger response scatter than iso and ortho bins, simply due to this biased selection. In the current context, I think this procedure will exaggerate the genuine iso-middle and ortho-middle differences, and generally undermines the authors' statistical conclusions. When choosing the middle bin for a model-free analysis, it should be fixed to a particular orientation difference, in the same way the iso and ortho bins are fixed.

We thank the Reviewer for this comment. In response to the concerns raised here and by Reviewer 1, we have adopted a different approach using a continuous fit of the data, thereby avoiding any arbitrary selection of bins in the main analysis. Additionally, we now present the bin analysis as a control in the Supplementary Material, implementing a fixed mid-bin as suggested.

3. I am somewhat confused whether the meta-analysis does or does not support Bayesian integration models. I have previously looked at response variability in one of my papers (Fritsche et al., 2020), of which three studies feature in the current meta-analysis. You can find the response variability patterns in Figure 7—figure supplement 4 in the eLife paper. Two out of the three experiments showed exactly the pattern predicted by the Bayesian model (Exp. 1 and 3). That is, response variability is lowest for iso orientations, intermediate for ortho orientations and highest for middle orientations. In fact, the variability at ortho orientations was equal to the average variability across all orientations. This is exactly what the Bayesian model predicts. I replotted the panels here, together with the Bayesian model which looks remarkably similar: https://drive.google.com/file/d/1DhORgH-Z6VKFEp44tCn16M5RrXTvAqCp/view?usp=share_link. Based on these analyses, I would have concluded that the evidence is in favor rather than against the Bayesian model.

When looking at Fig. 3A of the current study, these two experiments (#10 and #12) appear very similar to many of the other studies, which are collectively interpreted as evidence against the Bayesian model. To me this is observation is puzzling and makes me very doubtful whether or not the data supports Bayesian integration. I think we would first need to identify how the discrepancy between my previous and the current results arises before drawing strong conclusions in either direction. For instance, in my analyses I did not remove cardinal biases and serial dependence biases prior to computing response variability. It could be easily tested whether these two analysis steps have a sizable impact on the results and could explain the divergent findings. It should also be straightforward to create continuous plots of response variability as a function orientation difference for each of the 20 studies, similar to the plots that I have attached. This would allow to better understand the shapes of the response variability patterns and may clarify the differences between analysis strategies.

In my analyses, one of my experiments does not conform to the Bayesian model's prediction (Exp. 2), and yields more similar patterns to the ones reported by the authors (i.e., similar iso and ortho response variability). Curiously this is not reflected in Fig. 3A, where this studies' iso-ortho difference (#11) is in between studies #10 and #12. Is this due to an inconsistency in labelling, or due to a difference between analyses?

Overall, based on my observations the response variability patterns provide somewhat mixed evidence for a Bayesian model, although I found the similarity between Exp. 1 and 3 and the Bayesian observer quite remarkable. I am puzzled why the current analyses provide seemingly opposite evidence, despite using the same data.
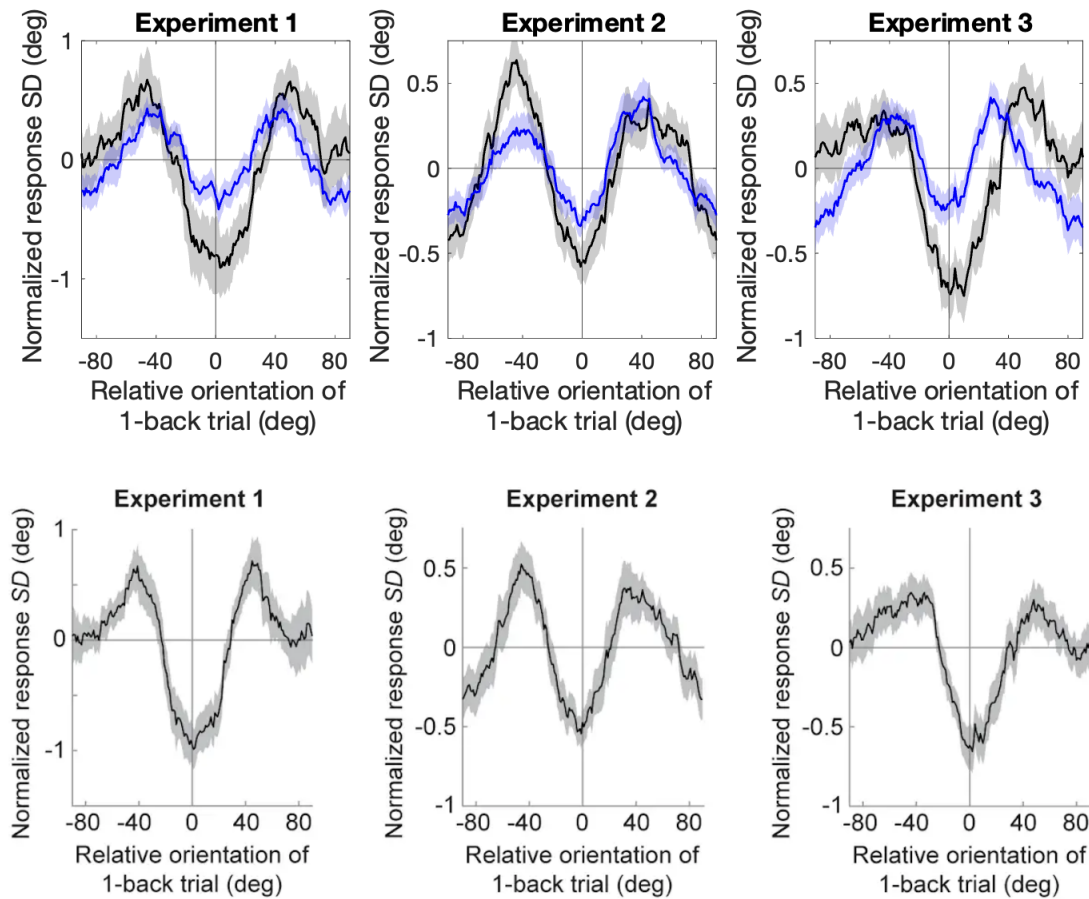
We are very grateful to the Reviewer for motivating this further analysis and providing us with the opportunity to examine the problem in greater detail. Indeed, we agree that previous reports on patterns of error scatter provide a mixed picture—some appearing consistent with the superiority effect discussed in our manuscript, such as those mentioned by the Reviewer.

As the Reviewer correctly noted, our analysis of the same data did not reveal the same patterns. To investigate the factors driving this discrepancy, we started by examining the main difference in data preprocessing—specifically, the removal of stimulus-specific biases (or 'orientation biases'), which was not applied in the original study. This led to a series of striking observations

4

and the identification of a critical confound in estimating error scatter as a function of serial dependence effects. That is, two sources of bias—the stimulus-specific and the serial dependence bias—previously considered independent in many existing studies, interact in problematic ways. This interaction can lead to spurious estimates of error scatter, creating apparent superiority effects that are purely artifacts of this confound.

In the *Confounding Effect of Stimulus-Specific Biases* section of the Supplementary Material we provide a clear demonstration of this issue, with detailed description of our analyses and findings, which we invite the Reviewer to carefully assess.

Once we removed orientation biases, the superiority effect largely disappeared across datasets, including those from Fritsche et al. (2020). **Below, we provide a figure showing the results with the original datasets** (first row: our analysis; black curve: minimal cleaning using 3*std; blue curve: stimulus-specific bias, i.e., orientation bias, removed; second row: plots provided by the Reviewer).

Thus, thanks to this analysis, we are now more confident than before that previously reported superiority effects are merely artifacts of stimulus-specific biases. We demonstrate this through a series of targeted control analyses, and we believe these findings will be of key importance not only in supporting the conclusions of our study but also for research on serial dependence in circular feature domains more broadly. This confounding interaction has remained entirely hidden in prior studies, and our work highlights the fundamental need to account for it in future investigations focusing on effects of serial dependence on error scatter.

Minor points:

4. I was wondering about the procedure to remove cardinal biases. The fitting of the two polynomials introduces a discontinuity around orientations 0 and +- 90 (cardinals?). This discontinuity is not real, right? There should be a smooth (but steep) transition of the bias around the cardinals. Perhaps I am misunderstanding, but would this procedure not lead to an overcorrection for orientations that are very close to the cardinals?

    We thank the Reviewer for this comment. To address this concern, we have refined our method for removing stimulus-specific biases by implementing circular regression with sine and cosine basis functions. For details, please refer to the Removal of Stimulus-Specific Biases *section in the Methods. Importantly, our results and conclusions remain unchanged, despite the inclusion of new datasets and the use of different bias removal methods across versions.*

5. I am slightly confused by the "broader discussion on the concept of optimality and ideal models" (line 445 - 458) in relation to the current findings. Even if the authors would have found support for Bayes-optimal integration, this integration would still be suboptimal in the context of randomized stimulus sequences. It is only optimal when visual input is temporally correlated (and if the observer has accurate knowledge of these correlations). The suboptimality in laboratory settings does not specifically follow from the current findings, as the last sentence of the paragraph seems to suggest.
    Furthermore, I don't understand the difference in logic to multisensory integration. Multisensory integration is only optimal when the signals from different modalities in fact originate from the same source, which is often the case in natural environments but not necessarily the case in laboratory settings. Thus, the degree to which the logic of optimal integration holds for temporal and multisensory signals depends on the temporal and spatial correlations of signals in the natural environment of the observer.
    Overall, it seems that the paragraph is mixing a few ideas from optimal and suboptimal integration to randomized (laboratory) and natural sensory input. I think the discussion would benefit from separating these ideas more clearly.

    We agree with all the points raised by the Reviewer in this comment and apologize for the lack of clarity in this section of our manuscript. We have now revised the text to improve clarity and provide a more detailed discussion. Specifically, we elaborate on why integration can be beneficial when signals in natural settings originate from the same source but may lead to interference in laboratory settings where signals are uncorrelated.

    Additionally, we now highlight findings demonstrating reduced or even repulsive (rather than stronger) serial dependence in experiments that introduce temporal correlations in stimulus sequences. These results further challenge some of the dominant views on serial dependence and its functional role. Lastly, we have avoided the use of the term optimal, as it is not essential to our argument or main conclusions.

Reviewer #3

Remarks to the Author:

Using meta-analysis on serial dependence studies, the authors addressed an important question: How do past-present similarities affect the variability of present perceptual performance? With large data

samples from 20 datasets, they showed that the variability of estimates in the current trials is modulated by the stimulus difference between previous and present trials. This modulation occurs not monotonically, but in an inverted u-shape. Although I think the topic is interesting and the results are sound, there is a lack of theoretical explanation for the present findings. Without digging deep into the mechanism underlying these effects, I don't find the present study suitable for publication in Nature Human Behavior. My comments in details are attached below.

We appreciate the Reviewer's general evaluation and comment. We have now revised substantially all text and presented new analyses and results, as well as a more in-depth discussion of the potential mechanisms underlying these effects. We believe that these additional analyses and the expanded discussion directly address the Reviewer's concerns and provide a stronger theoretical context for our results.

1. When people say "serial dependence," they don't always agree on its definition. Some researchers think it specifically refers to the influence of past stimuli on the present perception when the stimulus similarity between past and present is high. In this condition, it usually leads to an attractive bias. In orientation tasks, the past-present stimulus similarity for the peak bias is less than 30° (e.g., studies by Whitney, Burr, and de Lange groups). In such cases, serial dependence can be well explained by Bayesian integration model. However, "serial dependence" can also refer to the influence of past trial on present perception more generally. Indeed, there is repulsive serial dependence when past-present stimulus similarity is low (e.g., Fritsche, Mostert, & de Lange, 2017, peaked at ~75° difference). Compared to the extensive research on attractive serial dependence induced by high past-present stimulus similarity, the study on serial dependence with low past-present stimulus similarity is much less. Consequently, our understanding of serial dependence under such condition remains limited. Although the low past-present stimulus similarity induces no serial bias (or minor repulsive bias), it does not imply there is no influence from the past on present. The authors' choice to use the "ortho" condition (low stimulus similarity) as the baseline for perceptual performance and compare it with the "iso" condition seems somewhat arbitrary. They find comparable errors between two conditions and argue against the functional advantage of serial dependence under high stimulus similarity condition as predicted by Bayesian integration model. I think the authors need to rule out other possibilities before reaching this conclusion.

   We thank the Reviewer for this comment. We agree that the term serial dependence should not be used synonymously with a specific form of the effect or interpreted as implying a particular underlying mechanism. As clarified in Pascucci et al. (2023), the term refers to a statistical definition and should remain neutral with respect to its functional interpretation. To avoid ambiguity, we now explicitly mention that the form of serial dependence examined in this study is the *attractive* form—the one central to Bayesian and cue integration models predicting superiority effects. Testing the predictions of these models and the associated theoretical views is the core focus of our study.

   We would also like to clarify that the majority of studies on serial dependence (both attractive and repulsive), including all studies considered here by selection criteria, assess potential effects at both high and low levels of stimulus similarity. This is achieved by systematically testing serial dependence across the full orientation/motion direction angular space. Most of these studies show attractive serial dependence for high stimulus similarity and weakly repulsive or absent effects for low stimulus similarity. This well-established pattern has also been emphasized in the meta-analysis and review by Manassi et al. (2024). Thus, the finding that attractive serial dependence primarily manifests at high stimulus similarity is not a novel

or debated aspect, nor is the limited effect at large stimulus differences an 'unexplored' phenomenon.

In the current version of the manuscript, we no longer split angular differences into bins, so the 'ortho' condition is no longer arbitrarily defined (except in control analyses reported in the Supplementary Material). However, we acknowledge the Reviewer's concern that defining a baseline in serial dependence (e.g., considering the 'ortho' condition as a true baseline) is problematic, as a true baseline would imply no influence of prior stimuli. A similar issue has been addressed in spatial processing and crowding research, where it has been shown that spatial context—assumed to induce similar beneficial effects as serial dependence—does not improve performance beyond a true baseline (i.e., performance without contextual stimuli, a more clearly defined reference condition) (Cicchini et al., 2022; Ozkirli et al., 2025). We now incorporate this converging line of research into our discussion.

Nonetheless, the core prediction derived from Bayesian and cue integration models remains unchanged: a clear difference in error scatter is expected between highly similar and highly dissimilar (e.g., iso vs. ortho) angular differences. This is the prediction we directly test, and we believe our methodological approach is well-suited for this purpose. Furthermore, we have now included rigorous control analyses (see Supplementary Material) to rule out other potential confounds.

2. Serial dependence can occur at multiple cognitive stages. In fact, recent findings suggest that there is an adaptation-like repulsive effect during sensory processing (Sheehan & Serences, 2022; Hajonides et al., 2023; Luo, Zhang, & Luo, 2024). And it has been shown that brief adaptation to orientation can enhance the identification of orthogonal orientations by sharpening neuronal selectivity (Dragoi et al., 2002, nature neuroscience). It is possible that in the "ortho" condition of serial dependence, the present orientation identification is also enhanced by previous orthogonal orientation, resulting in a small error scatter. That it, although both "iso" and "ortho" conditions exhibit lower variability in perceptual estimates, they might be accounted by difference mechanisms, through Bayesian integration and adaptation, respectively.

We agree with the Reviewer's point, and indeed, one of the authors was among the first to propose a repulsive low-level adaptation stage (Pascucci et al., 2019). However, the context here is markedly different from studies reporting beneficial effects of adaptation at orientations orthogonal to the adapter. In classic adaptation paradigms, the adapter is a strong stimulus presented for several seconds—sometimes minutes—often followed by an adaptation top-up before testing its effects on a target stimulus. In contrast, the majority of studies considered here were specifically designed to minimize adaptation effects by using brief, low-contrast stimuli, mostly followed by a noise mask. Thus, we believe that adaptation at large orientation differences is unlikely to explain the observed pattern of error scatter.

Nonetheless, as this remains a possibility, we now explicitly address it in the Discussion section. Importantly, the explanation suggested by the Reviewer would require an interaction between two opposing effects that cancel out the superiority effect. This is a rather complex hypothesis, and in the absence of unequivocal evidence, a simpler explanation is preferable (following Occam's Razor). However, even if such an interplay were present, it would not change the core conclusion: there is no evidence of superiority effects in serial dependence. Regardless of whether a single or multiple mechanisms are involved, performance does not improve for highly similar stimuli as predicted by cue integration and Bayesian models.

3. The Bayesian integration model for serial dependence works only if the past and present stimuli are of high similarity. How does the brain deal with the situation that past and present are dissimilar? It's possible that like Bayesian casual inference which is typically demonstrated in multisensory cue combination tasks (Körding et al., 2007), the brain also infers whether past and present trials belong to a same event first. If they are similar (same event), the brain would integrate past with present, inducing attractive serial dependence. If not, the brain would separate them. Moreover, it has been shown that visual perception is retrospective Bayesian inference from high level to low level (Ding et al., 2017). If the sequentially presented stimuli are different, to preserve this ordinal information, perceptual difference would be exaggerated. Could this operation result in the small error scatter in the "ortho" condition?

    This is an interesting hypothesis that can be addressed in future research, aligning with our broader aim of moving beyond existing models to better understand serial dependence. Emerging alternative models, which we now reference, are indeed shifting towards framing the problem in terms of inferring whether two stimuli originate from a common or separate source (e.g., Chetverikov, 2023).

    However, before adopting this perspective, a simpler explanation must be ruled out—namely, that stimuli at intermediate differences interfere with the representation of the current stimulus. Following the Reviewer's suggestion, we now discuss alternative models and the challenge of inferring common sources. Additionally, we highlight a key empirical finding: in tasks where stimulus sequences are correlated—thus promoting the inference of a common source from trial to trial—attractive serial dependence is actually reduced rather than enhanced. This suggests a completely different strategy, in which the visual system exploits temporal correlations to better discriminate consecutive stimuli (Blondé et al., 2023; Ufer & Blank, 2024).

4. In either case, there would be potential decrease in the variability of perceptual estimates for "ortho" condition. Indeed, as shown by the authors as well as Cicchini, Mikellidou, and Burr (2018), the peaked variability/"root-mean square error" is around 45° past-present difference, larger than that for the peaked bias (within 30° past-present difference). Is the peak error due to the fact it is far away from both Bayesian integration improvement for "iso" condition and another improvement mechanism for "ortho" condition?
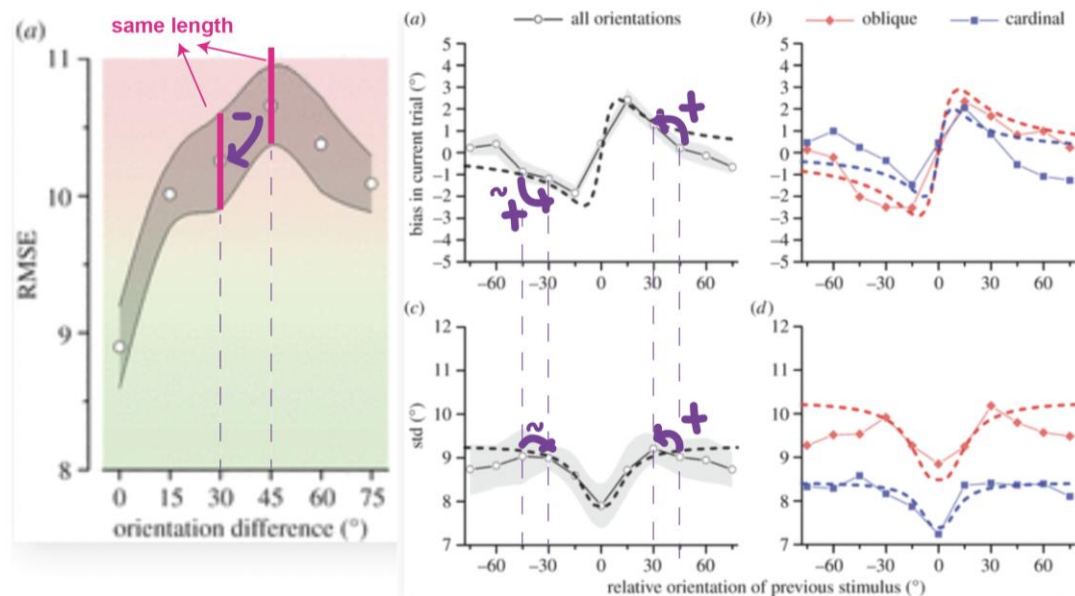
    It is unclear whether the Reviewer is referring to the peak in variability or the peak amplitude in RMSE (defined as $\sqrt{bias^2 + scatter^2}$, Eq. 3.4 in Cicchini et al., 2018). However, upon close examination of the figures in Cicchini et al. (2018), we respectfully disagree with the Reviewer's interpretation:
    -   **If the reference is to RMSE peak variability:** the peak does not occur at $\Delta = 45°$, but rather around $\Delta = 30°$. For reference, we have attached a screenshot below with pink lines of equal length, clearly illustrating that variability at 30° exceeds that at 45°.
    -   **If the reference is to the RMSE peak value:** while it appears to peak at $\Delta = 45°$, this is most likely a misreport in Cicchini et al. (2018). Both bias and scatter are greater at $\Delta = 30°$ than at 45° (see purple lines and plus signs in the figure below). Given that RMSE is computed as $\sqrt{bias^2 + scatter^2}$, a peak at 45° is not mathematically supported. Moreover, Cicchini et al. (2018) did not include data for $\Delta = 90°$, despite its presence in their publicly available dataset.

Given these inconsistencies between the Reviewer's claim and the actual data, we suspect the perception of the situation may have been influenced by their hypothesis regarding a secondary superiority effect at orthogonal distances. Regardless, it is important to note that even Bayesian models—on which our manuscript focuses—do not inherently predict either a peak in bias and scatter at the same Δ, nor a superiority effect in the orthogonal condition.

cues. Overall the total squared error is given by their sum (figure 2h):

$$ERR = BIAS^2 + VAR,$$

3.4

To the best of our ability, we do not see a principled reason why error scatter should decrease at ortho unless invoking complex and unverified explanations. We explicitly present the predictions of the models referenced by the Reviewer in the main text, for both bias and scatter. These models do not inherently predict a decrease in error scatter at ortho to the level of iso. Since the primary goal of our study is to test the predicted superiority effect (see also Ozkirli et al., 2025), we prefer not to introduce or test speculative explanations that have not been explicitly formalized or validated in prior research.

In short, "serial dependence" is a complex phenomenon, which involve interactions between different systems (memory, perception, and decision-making) and multiple computational mechanisms (efficient coding, bayesian integration, or even casual inference). I appreciated the authors effort on studying serial dependence across the whole range of past-present similarity. But I am not convinced by the present finding of comparable perceptual variability between "iso" and "ortho" conditions as evidence against Bayesian integration. A more unified framework is needed.

We partly agree with the Reviewer closing remark, particularly regarding the complexity of serial dependence and the limitations of current models—especially those assuming superiority effects. However, we do not take for granted the involvement of all suggested models until they have been properly tested and systematically compared.

Our study contributes to this effort by encouraging future research to rethink serial dependence in light of alternative frameworks. A 'more unified' framework can only be achieved after a proper mechanistic and computational understanding of the phenomenon at hand. In no way should this involve neglecting observations and patterns that deviate from current models.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have done a good job at addressing my comments. The present version is far more accessible for non-expert readers (like me). I also think that the decision to replace the meta-analysis with a trial-level analysis simplifies the text and analyses substantially.

We would like to thank the Reviewer for their feedback. The manuscript has indeed improved substantially as a result of the suggested changes.

Just a few minor comments:

- Title: Consider presenting your work as a "mega-analysis" instead of a "large-scale analysis"

We have changed the title as suggested.

- Figure 1: This is the first time your reader is presented with an experimental paradigm for the study of sequential effects. Please, mention explicitly that the current and previous orientation refer to consecutive trials (not to consecutive stimuli within a trial). This is explained in the main text only afterwards.

We have now clarified this point in the caption:

> "Single-trial errors are plotted against the difference between the target feature (e.g., orientation) in two consecutive trials (Δ: past-present, in degrees)"

- Line 164: why use a standardized effect size (Hedges' g) instead of the raw number of degrees? Given that the experimental paradigms and dependent variables are quite homogeneous, I think it makes sense to report exclusively degrees.

We have now included the mean and median bias across datasets:

> "Despite variations in paradigms, stimuli, and conditions, most datasets showed consistent positive serial dependence: adjustment errors were systematically biased toward the previous trial feature, with a median positive effect across all datasets (**mean =1.06°; median = 0.99°;** Hedge's g = 1.47)."

- any evidence of heterogeneity across studies or participants? Although the analysis is no longer presented as a meta-analysis, heterogeneity is still relevant. It can be easily assessed with the random slopes of the LMMs.

We agree with the Reviewer that heterogeneity remains an important aspect to consider. We now quantify it by reporting the proportion of total variance due to heterogeneity using the I² metric, computed as the ratio between the variance associated with a given random effect (τ^2) and the total variance (τ^2 + MSE). For both the *obsid* and *codenum* random effects, the results indicate medium to large heterogeneity for the random intercept, reflecting variability in error scatter at the iso level.

Specifically, for *obsid*, the intercept variance was $\tau^2 = 9.54$, corresponding to $I^2 = 80.98\%$, while for *codenum*, $\tau^2 = 13.23$, with $I^2 = 85.53\%$. In contrast, the slopes associated with the effects of *mid* and *ortho* showed weak or negligible heterogeneity, suggesting consistent differences across studies and observers in how *iso* compares to these other similarity levels. For *obsid*, the *mid* and *ortho* slopes had $\tau^2 = 0.15$ and 0.42, with $I^2 = 6.33\%$ and 15.84%, respectively. For *codenum*, the *mid* slope had $\tau^2 = 0.16$, $I^2 = 6.77\%$, and the *ortho* slope had $\tau^2 = 0.06$, $I^2 = 2.44\%$. These results suggest that while baseline variability in error scatter is substantial across observers and experiments, the relative modulation by stimulus similarity (i.e., differences between *iso* and *mid*, and *iso* and *ortho*) is highly consistent.

Reviewer #2 (Remarks to the Author):

I would like to thank the authors for their thorough revisions and replies to my concerns. I particularly appreciate the detailed control analyses to understand the interactions between stimulus-specific and serial dependence biases. The authors demonstrate quite convincingly that removing versus not removing stimulus-specific biases during preprocessing explains the differences between current and previous results. I have a few follow-up questions related to this explanation and some minor suggestions.
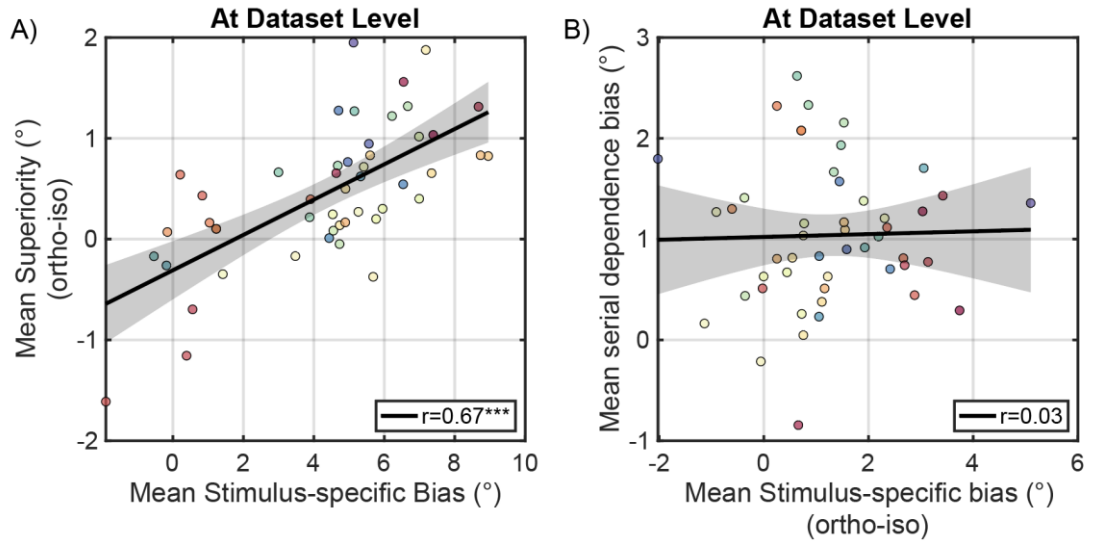
We thank the Reviewer for his comments. Together with the feedback received from the other Reviewers, these suggestions have significantly improved the manuscript. We would like to explicitly acknowledge this contribution in the final version of the manuscript. Below, we provide detailed responses to the remaining points.

Major points

1. I can see how response variance can be reduced by accounting for stimulus-specific biases. This explains the orientation-dependent up-down shift in response variance in Figure S3D. However, it does not by itself explain the iso-ortho difference in response scatter. To explain this, the authors assume a history dependency in the stimulus-specific bias, which becomes central to the argument: "This [iso-ortho difference] likely reflects differences in the magnitude of stimulus-specific bias after repeating the same stimulus ($\Delta = 0°$) or changing by 90° ($\Delta = \pm 90°$)." Do the authors think that this would occur due to "classical" serial dependence or would this be an additional history-dependency? Wouldn't this reduction of oblique bias for $\Delta = 0°$ constitute a "superiority effect", as responses would be more accurate (less biased) than for $\Delta = \pm 90°$? It would be helpful to explain how such a superiority effect would be similar or different to that posited by Bayesian integration. Also, to my knowledge such a history-dependency in stimulus-specific biases has not been previously shown. Therefore, it would be important to assess whether such a history dependence is present in the empirical data.

    We thank the Reviewer for raising this point, and we agree that distinguishing between the typical serial dependence bias and the observed effect on orientation bias is important. We believe that the history-dependency on orientation bias reflects a distinct form of sequential effect, separate from classical serial dependence. This interpretation is supported by three main observations:

    I.   Across datasets, the magnitude of the reduction in orientation bias shows no correlation with the strength of serial dependence (now included as Fig. S5B, also shown below for convenience). If the two effects shared a common underlying mechanism, a positive correlation would be expected.
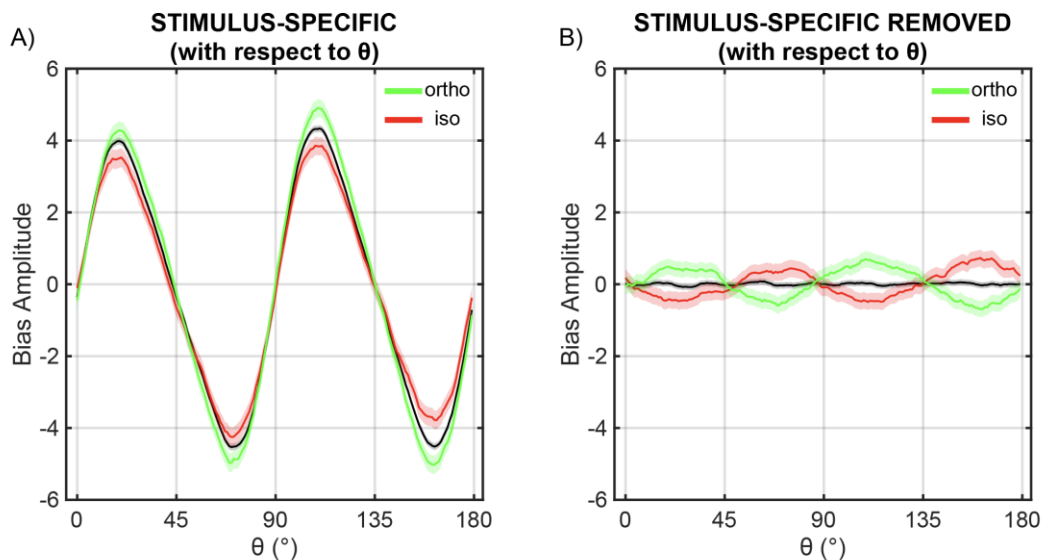
A) At Dataset Level — Mean Superiority (°) (ortho-iso) vs Mean Stimulus-specific Bias (°); r=0.67***
B) At Dataset Level — Mean serial dependence bias (°) vs Mean Stimulus-specific bias (°) (ortho-iso); r=0.03

II. The largest difference in stimulus-specific bias between *iso* and *ortho* occurs over a range of orientations/motion directions (θ) where the stimulus-specific bias peaks (dark red in S3A), while the typical serial dependence bias is minimal (Fig. S3B). In fact, as shown in Fig. S3E, we observed a strong negative correlation between the two types of biases. This further supports the interpretation that the history dependence in the stimulus-specific bias is qualitatively distinct from classical serial dependence, which, if driven by the same mechanism, would be expected to show a positive correlation.

III. While there is an overall increase in stimulus-specific bias from *iso* to *ortho* across datasets ($t(48) = 6.46$, $p < .001$, Cohen's d $= 0.93$), this finding does not support Bayesian or optimal cue integration models as explanations for serial dependence. In the classic formulations considered in our study, these models do not predict that the stimulus-specific bias decreases at $\Delta = 0°$, but rather, only a reduction in overall response variability. Moreover, while such models predict larger superiority effects (i.e., lower error scatter) when serial dependence bias is stronger (as shown in Fig. 3A), our data show the opposite pattern—Fig. S3G reveals that stronger serial dependence bias peaks are associated with smaller superiority effects, contradicting the basic predictions of these models.

We hope that these analyses answer the Reviewer's question. In sum, while the lower stimulus-specific bias for *iso* trials may be interpreted as an "improvement"—that is, a partial release from long-term stimulus-specific biases possibly reflecting environmental priors—we find no evidence that this effect arises from the same process responsible for the typical bias toward previous stimuli (i.e., classical serial dependence). We now discuss this point in the revised manuscript (page 11, lines 355-359). We agree that this is a novel finding and encourage future studies to focus on history effects on stimulus-specific biases and their interplay with other forms of sequential effects.

2. Furthermore, the authors' procedure for removing stimulus-specific biases is history-free. It removes the average bias, regardless of stimulus history. Therefore, it would not appear to be able to account for the history-dependent variation in stimulus-specific bias magnitude. I suppose the procedure would overestimate stimulus-specific biases for $\Delta = 0°$ and underestimate those for $\Delta = 90°$. Is the idea that an overestimation would inflate variance at $\Delta = 0°$ to account for the failed removal of variance at $\Delta = 90°$? It is not immediately obvious to me that this would work. It would be important to clarify the logic behind why the removal of history-*independent* stimulus-specific biases manage to provide an unbiased estimate of error scatter.

This is an important point. Below, we demonstrate how our debiasing method works on the real data (*iso* in red, *ortho* in green), followed by a simulation to show the effects of stimulus-specific bias on the error scatter analysis.
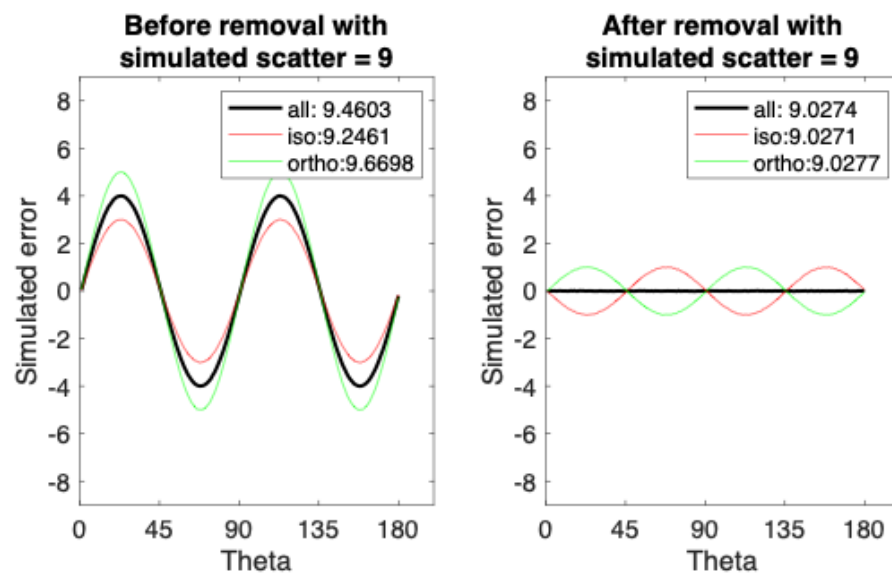

As scatter is measured via the standard deviation, larger orientation biases in one condition (left panel, for instance, a peak-to-peak amplitude of 10° in *ortho* vs. 8° in *iso*) would *artificially* increase the standard deviation—not due to increased variability per theta, but because of that across theta. After subtracting the overall bias pattern across theta (e.g., with peak-to-peak amplitude of 9°), this recenters the errors around zero (right panel), preserving the fluctuations of difference in stimulus-specific bias between *iso* and *ortho* conditions without inflating the variance much in either: The deviation from zero across theta becomes less prominent (with peak-to-peak amplitude of 1° in both *iso* and *ortho*), hence minimizes any artifacts due to stimulus-specific bias in error scatter analysis as a function of Δ.

In other words, what remains in panel B here is the residual error of the bias estimation procedure. That error is similar between *ortho* and *iso*. Hence, one cannot say that the procedure "inflates variance at $\Delta = 0°$ to account for the failed removal of variance at $\Delta = 90°$" as the Reviewer puts it (see also the simulation below). Rather, it removes the stimulus specific biases leaving only minor and comparable fluctuations between *iso* and *ortho*, which do not confound

the error scatter analyses as a function of Δ. Although the removal does not fully flatten the stimulus-specific bias curve, it helps avoid overestimation of error scatter and makes the anomalous pattern in Fig. S3D—where the levels of response variability do not align when grouped by θ and Δ—disappear after this correction. Furthermore, serial dependence bias pattern remains unchanged (see Fig. S3B vs. S4B). All of these demonstrate the effectiveness of our method.

As a proof of concept, we also demonstrate this in a simple simulation. We generate two sets of errors (2 x 180 million trials) by sampling from a normal distribution with mean 0° and standard deviation of 9°. Then we add a mean shift to both sets as a function of θ, resembling stimulus-specific orientation biases with different amplitude per set (e.g., *iso* and *ortho*). As shown in the left panel of the figure below, after adding stimulus-specific biases of different amplitude, computing the standard deviation across all data points results in an overestimation of the scatter, to a greater extent for the condition with larger stimulus-specific biases. However, by subtracting the mean stimulus-specific bias (e.g., debiasing as we do in our main analysis), this overestimation issue is largely overcome and the estimated scatter for both cases is closer to the true simulated scatter value (right panel).
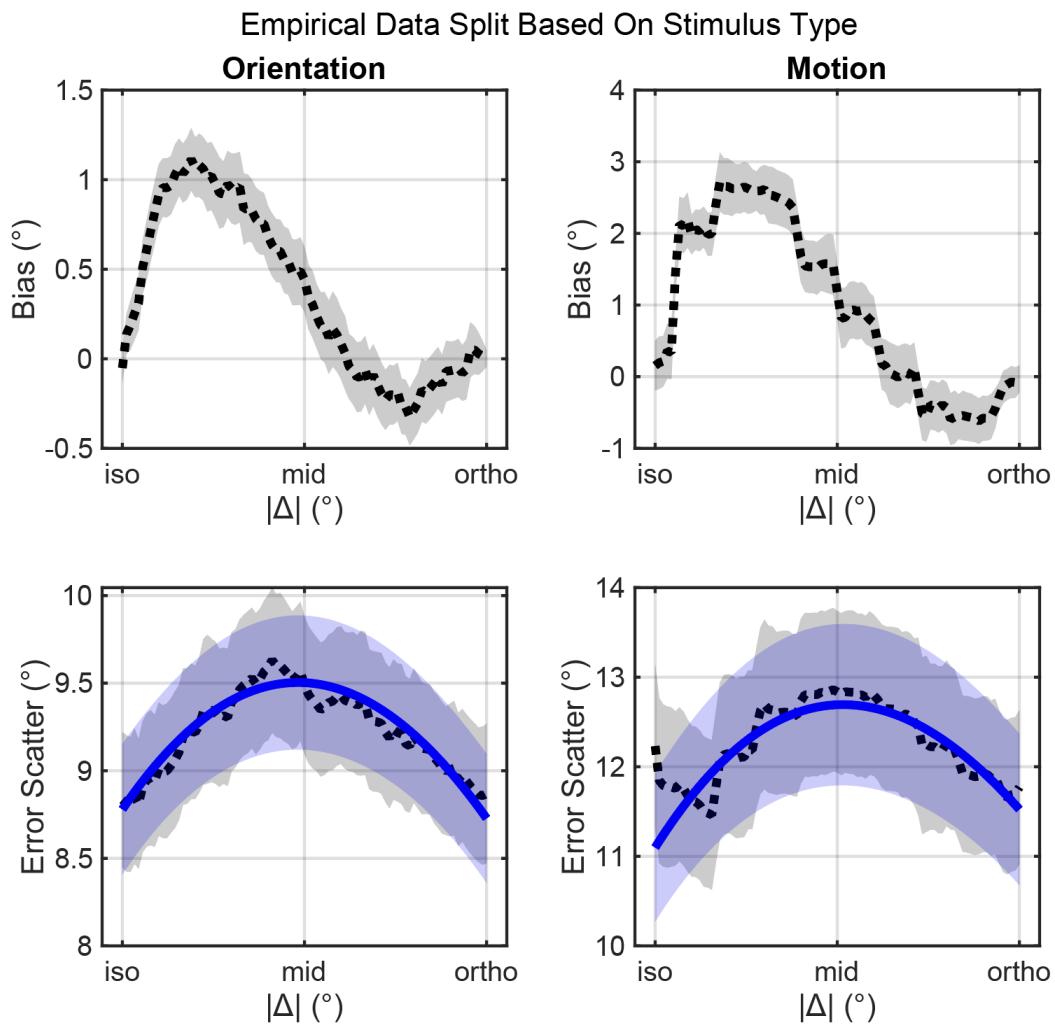

That said, we agree that an ideal approach would involve modeling an interaction between serial dependence and orientation bias within a single framework, which would allow for a more accurate removal of history-dependent orientation bias. However, fitting such a model would require substantially more data per dataset (and subject) to obtain stable estimates, particularly if applied at the subject level. We now discuss this further in the Supplementary Materials (page 4, lines 135-142).

Minor points

It would be interesting to mention separate results for orientation and motion estimation tasks, to show that the results hold for different stimulus features.

We thank the reviewer for the suggestion. Indeed, we assessed the results separately for the two types of stimuli in our initial analysis, and the findings held for both orientation and motion stimuli. We now provide the figure attached below in Supplementary Materials and S7.

There are some errors in labelling the supplemental figures both in captions and in text.

We have now fixed these errors.

Reviewer #3 (Remarks to the Author):

I appreciate that the authors have expanded their datasets and refined their analysis, which increases my confidence in their findings. As I stated in the first round of review, the authors raise an important question that has been largely overlooked in the serial dependence literature.

1. However, my primary concern remains unaddressed. Only reporting the phenomenon that variance follows an inverted U-shape as past-present similarity decreases provides limited insight without a mechanistic explanation. I believe the conclusion remains overstated. The "inverted U-shape" of variance is insufficient to support the statement that 'serial dependence deteriorates rather than improves perceptual decision-making'.

   We thank the Reviewer for inviting us to further elaborate on this point. Our main argument is that existing models cannot explain the inverted U-shape, as they systematically predict a pattern consistent with 'superiority'. Our results clearly show that 'superiority', as predicted by existing models, is not supported by the data, thus, we need alternative accounts that relax the assumption of a beneficial role. We suggest that the observed pattern may instead reflect interference and propose how this might occur (see Discussion). However, as stated, this remains a hypothesis—the primary goal of our mega-analysis is to document the phenomenon, not to propose a new model.

   While the statement 'serial dependence deteriorates rather than improves perceptual decision-making' may seem overstated based on the inverted U-shape alone, it could also be considered an understatement given that dominant Bayesian models, which currently guide the field, fail to capture this pattern. We also invite the Reviewer to consider a parallel issue in a distinct domain—spatial vision (e.g., crowding)—where similar models and notions of superiority have been proposed, but where interference provides a more plausible and better-supported explanation of perceptual biases, with unequivocal baseline levels to assess superiority (see 1e in Ozkirli et al., 2025).
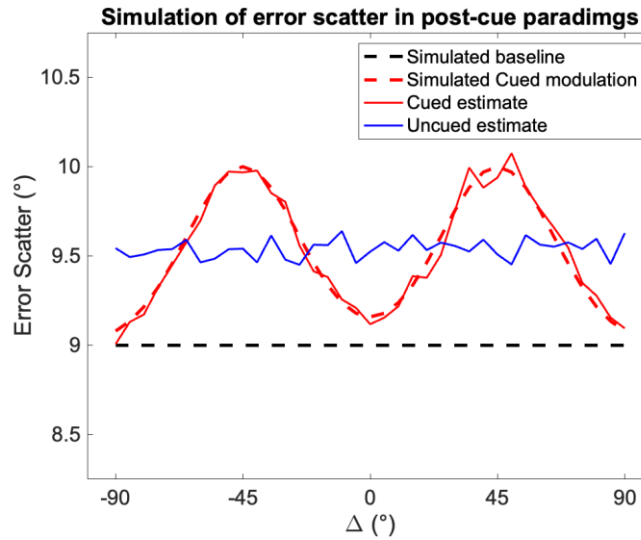
2. Given that the authors have clarified they are examining only the attractive form of serial dependence—which manifests when past and present stimuli are similar—the 'ortho' condition is not an appropriate baseline for this analysis. To properly investigate the functional role of serial dependence, the study should compare conditions where past and present stimuli are similar against conditions where no prior influence exists, rather than using orthogonal conditions that are inherently irrelevant to the attractive serial dependence phenomenon. For example, since serial dependence is context-dependent (Fischer et al., 2020), the influence of previous orientation under different contextual conditions might provide a more appropriate baseline for comparison.

   We respectfully disagree with the Reviewer, based on the points below.

   I. Since we are specifically testing superiority effects predicted by current models of *attractive* serial dependence—effects expected to arise when past and present stimuli are

similar—the orthogonal (*ortho*) condition provides the most appropriate reference. Precisely because serial dependence is not expected for highly dissimilar stimuli ("inherently irrelevant to the attractive serial dependence phenomenon"), the *ortho* condition serves as a meaningful "null" case: it allows us to evaluate the modulation of error scatter that can be attributed to attractive serial dependence. Model predictions displayed in 3A of our main text is in total agreement with this: the condition *ortho* best approximates the error statistics with no serial dependence in both models (see the convergence to no bias and baseline scatter).

II.   We agree that context-based manipulations (e.g., task relevance) seem to, in principle, offer a more concrete way to assess 'baseline' error scatter. For example, many studies have shown reduced or reversed serial dependence for task-irrelevant features (Blondé et al., 2023; Ceylan & Pascucci, 2023; Fischer et al., 2020; Fritsche & de Lange, 2019; Houborg et al., 2023). However, these studies typically present task-relevant and irrelevant features within the *same* trial (e.g., via simultaneous or sequential stimuli), which complicates the picture. Consider one of the designs used by Fischer et al. (2020), where two motion directions are presented simultaneously, and a post-cue indicates which one is task relevant. On trial *n*, the error scatter reflects the independent influence of both directions presented on trial *n–1*. Even if only the cued direction exerts a serial dependence bias effect, error scatter measured as a function of the uncued direction cannot be used as a proxy for the baseline scatter. This is because such a scatter estimate will already be confounded by the effects of cued direction, and will reflect elevated values compared to the baseline. We can easily show this in a simulation. Below, we simulated 500000 trials with two stimuli shown on each trial. The true baseline error scatter is fixed at 9° (yellow dashed horizontal line). The "Cued" stimulus modulates scatter according to an inverted U-shaped function of the orientation difference (Δ, violet dashed line). The "Uncued" stimulus, as per Reviewer's argument, has no modulatory effect. As the Reviewer can see in the figure below, when we estimate the scatter from the simulated data, the measured "Cued" condition (orange) matches the imposed U-shaped modulation. The "Uncued" condition (blue) is flat, but it does not reflect the true baseline; instead, it reflects the average modulatory effect of the cued stimulus. This is because the responses used to estimate the error scatter as a function of cued and uncued previous stimuli are identical. Thus, using such approach will not provide any better estimate of the baseline compared to the *ortho* condition, and may in fact increase the risk of misinterpretation. For these obvious reasons, the *ortho* condition, which by construction involves orientations outside the range of attractive serial dependence, remains a better reference level.

Simulation of error scatter in post-cue paradimgs

III.   Additionally, in specific designs, out of context trials would have their own sources of variability due to confounding influence of context-switching. For example, if we would have compared trials with stimuli presented in the same location vs. different ones, serial dependence would indeed be diminished in the latter case, but variability would be increased due to the reallocation of spatial attention, location-based expectations, and other similar effects. Thus, we believe that *ortho* is the best comparison point possible (also according to the prevailing models we are testing).
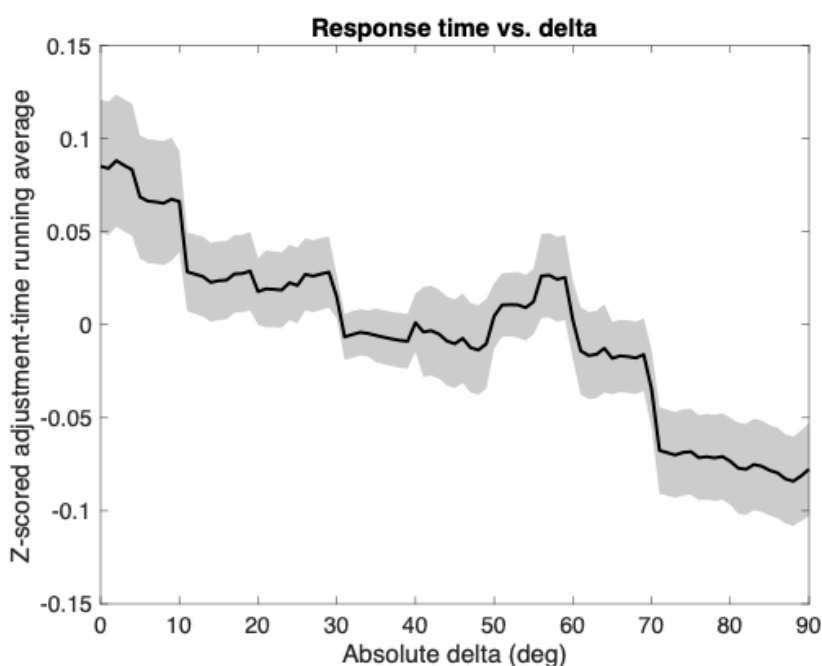
We nevertheless further clarify this logic and the goal of testing for superiority effects due to attractive serial dependence with reference to the *ortho* condition in the revised manuscript (at the end of the "Modeling the error scatter" in Methods Section).

3.   The interpretation of serial dependence as 'interference' is speculative and potentially contradictory. The authors claim that 'when consecutive stimuli share identical features, interference is minimal because their representations fully overlap.' However, overlapping representations could plausibly enhance perception by strengthening the signal rather than minimizing interference. This raises a fundamental question: if overlapping representations improve perceptual performance, how can this effect be characterized as 'minimized interference'? Supporting this alternative interpretation, Cicchini et al. (2018) demonstrated that reaction times decrease when past and present stimuli share identical features, suggesting facilitation rather than interference.

The Reviewer seems to rely on the assumption that 'overlapping representations improve perceptual performance'. But this is exactly the assumption that we break and demonstrate it is not the case across all the datasets used in our mega-analysis: overlapping representations do not improve perceptual performance as compared to non-overlapping ones (e.g., *iso* vs. *ortho*). They *could* does not mean they do.

Regarding the point on adjustment times raised by the Reviewer, when we evaluated adjustment times as a function of similarity between past and current stimuli across all datasets

12

included here, we actually found a pattern that is **completely opposite** to what was reported by Cicchini et al. (2018): adjustment times decrease as stimulus dissimilarity increases (see below). This means that, on average, participants were faster when current and past stimuli were more different. Therefore, when assessed across multiple datasets—beyond Cicchini's single report—the pattern does not support the 'alternative interpretation' of improved performance due to serial dependence, at all. Instead, it aligns well with the interpretation proposed by Bae et al. (Bae & Chen, 2025; see also Akselberg et al., 2025), who reported already this pattern and suggested that faster responses under dissimilarity reflect reduced *interference* between choice alternatives, as the current and previous ranges of possible responses become more distinct.

Response time vs. delta

In fact, following the same reasoning proposed by the Reviewer, our adjustment time results would appear to support our interference framework. One might try to reverse the argument and claim that responses are faster at *ortho* because there is some additional source of facilitation at *ortho*—as the Reviewer suggests when referencing reduced noise at *ortho* in a single study with a different paradigm and scope. But this interpretation is problematic for several reasons. First, it would require positing an unnecessarily complex scenario: attractive serial dependence and enhanced performance at *iso*, repulsive adaptation and enhanced performance at *ortho*, and yet speeded responses only at *ortho*. For some reason, then, the supposed facilitation at *iso* would not affect adjustment times—or would even slow them down—which appears inconsistent and implausible compared to our simpler interference account, where response times get faster as interference reduces.

Second, interpreting adjustment times in these tasks is in general problematic. These are not speeded tasks: response was not time-constrained. As a result, the interpretation of adjustment times can be rather ambiguous. For instance, while one might assume that faster response times reflect 'improved' performance, the opposite view is also possible: longer adjustment times

might reflect a more refined and precise adjustment process, while faster responses might reflect coarse, less accurate estimations. In other words, speed cannot be directly equated with perceptual improvement in this context. For these reasons, we chose not to highlight the adjustment time results in the main text. However, we provide this analysis here to show that even if one were to interpret adjustment time as a marker of facilitation, the observed pattern would still be more consistent with reduced interference under dissimilarity, rather than the Reviewer's proposed facilitation under similarity.

We hope the Reviewer will find this additional evidence helpful in appreciating our motivation to reconsider the mechanisms behind attractive serial dependence, and to challenge widely held assumptions about its functional role.

4. Examining serial dependence across the full range of past-present differences would be more comprehensive, though this would shift the focus away from investigating the functional role of serial dependence specifically. The 'ortho' condition likely involves distinct types of past-trial influences that differ qualitatively from attractive serial dependence. As I reasoned above, the 'iso' condition likely benefits from past information. Since performance (bias and variance) is comparable between 'iso' and 'ortho' conditions, the 'ortho' condition may also benefit from past information, albeit through a different mechanism. Without empirical verification, we cannot assume this condition represents an absence of past-trial influence, and using it directly as a 'no serial dependence' baseline may confound the findings. Thus, there appears to be a gap between their investigation of variance across the full range of past-present differences and their conclusion about whether serial dependence deteriorates or improves perception.

    We thank the Reviewer for raising this point. First, we clarify that our analyses do span the full range of past–present feature differences (|0–90°|, see the polynomial fit analysis), and our primary aim is to assess whether the pattern of error scatter aligns with the predictions of models that posit beneficial/superiority effects of serial dependence. These models—whether assuming a single attractive mechanism or a combination of attractive and repulsive components (e.g., see Fritsche et al., 2020 and the related figures showing predicted scatter)—predict improvements specifically near the *iso* condition. To our knowledge, no model in this domain posits an independent mechanism of improvement that would operate selectively at the *ortho* condition.

    We do not assume that the *ortho* condition reflects an absence of past-trial influence. Rather, we treat it as a reference point where attractive serial dependence is expected to be minimal or absent, which is what current models assume. While we appreciate the Reviewer's suggestion that other mechanisms could be at play at *ortho*—perhaps even beneficial ones—we are not aware of any empirical support for this in the current literature on orientation or motion serial dependence. For this reason, and in line with the principle of parsimony, we chose to focus on the most straightforward interpretation: interference rather than two distinct improvements.

    That said, we now discuss the speculative possibility that more than one mechanism may be operating—potentially interacting in non-trivial ways across similarity space (Discussion, page

12, lines 392-398). While this view may further complicate the functional interpretation of serial dependence, we believe that the clearest contribution of this study lies in documenting a robust empirical pattern that challenges existing models in their current form, which should guide future work.

5. Both behavioral evidence, including work by one of the authors (Pascucci, PLoS Biol, 2019), and recent neural findings (Luo et al., PLoS Biol, 2025; Shan, Hajonides, & Myers, bioRxiv, 2024) indicate that attractive serial dependence is linked to decision-making processes. The decision-making system likely employs flexible strategies to guide current decisions under varying conditions. As noted in my initial review, Bayesian causal inference models provide a standard framework for determining whether two cues originate from the same or different sources (as in the 'ortho' condition here). Furthermore, Ding et al. (PNAS, 2017) demonstrated a repulsive bias in sequential orientation reproduction similar to the repulsive effect observed near the 'ortho' condition (Figure 3B, upper), which could be successfully explained through retrospective Bayesian inference from higher to lower processing levels.

The Bayesian model is a very general and powerful framework. I agree with the authors that the current Bayesian model of serial dependence is not good enough to explain the patterns of variance under conditions when the past and present are dissimilar. The authors can extend current simple model to provide more mechanistic explanations for various conditions.

We fully agree with the Reviewer's comment. Indeed, the main aim of our study is to highlight precisely this gap: the error scatter patterns that we observed across a large range of datasets do not align with the predictions of current Bayesian and cue-integration models, nor with dominant conceptual interpretations of attractive serial dependence.

As also noted in our manuscript, we encourage the development and testing of alternative models in general. However, purely Bayesian models—even those referenced by the Reviewer in their first review—do not appear to capture the error scatter pattern, as the Reviewer also acknowledges.

For example, the causal inference model cited by the Reviewer (Körding et al., 2007), while able to account for repulsive effects when two stimuli are inferred as originating from different sources, cannot explain the pattern of error scatter observed across datasets, which is the central finding of our mega-analysis. In fact, the authors explicitly state: "*The optimal visual estimate when visual and auditory sources are different is the same as when there would be only a visual signal, and likewise for the auditory estimate*". This supports our choice of *ortho* as a reference point for the baseline error scatter. This same causal inference model has also been applied in the spatial context domain to argue for optimal integration, implying similar mechanisms as in serial dependence, and yielding the same prediction of a superiority effect as cue integration models (see Cicchini et al., 2022, *Nat. Commun.*, Fig. 2b, also attached below). Thus, the causal inference model has the same prediction of a superiority effect, just as other existing accounts. Importantly, while the data in Cicchini et al. (2022) appeared consistent with the predicted superiority pattern (with *ortho* scatter larger than *iso*), this was likely driven by a between-

subject confound, as different groups of participants were tested in the ortho versus other conditions. A subsequent replication study with all conditions tested within the same group and a doubled sample size failed to reproduce the superiority effect and instead revealed the inverted U-shape (*iso = ortho*; Ozkirli et al., 2025; *Nat. Commun.*).
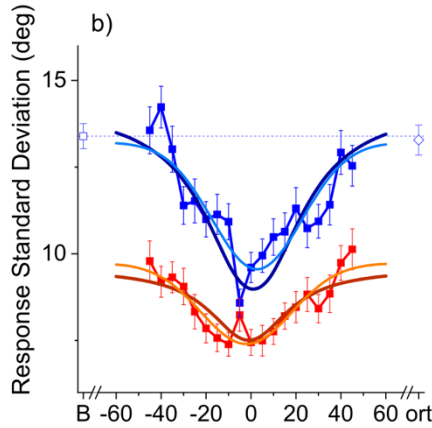
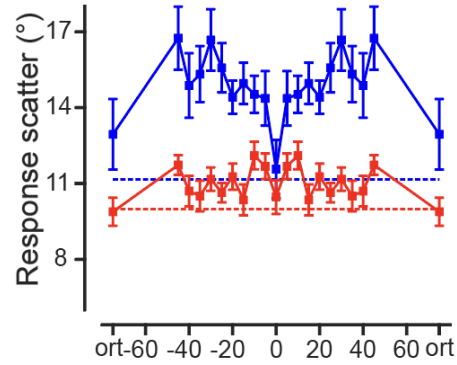

Fig. 2b from Cicchini et al., (2022)   Fig. 1e from Ozkirli et al., (2025)

Similarly, to the best of our knowledge, the retrospective Bayesian inference model (Ding et al., 2017) does not generate clear predictions about error scatter that would align with the U-shaped pattern found both in our study and in crowding. We therefore agree with the Reviewer that extending current simple models is a promising direction, but in this manuscript, we prefer to remain general and simply note that alternative models should be formally developed and tested.

Our central point is that many existing models may appear intuitively plausible but could equally well fail to capture the pattern we reported here. Until alternative frameworks are systematically formalized, evaluated through model comparison, and tested in dedicated experiments, it remains speculative, in our view, to argue that any specific model class provides a sufficient account.

Minor:

The model simulation (Figure 3) is misleading due to scale discrepancies. The simulated bias ranges from 0-10, which is substantially larger than the observed bias range of 0-2. The scales should be made comparable to enable meaningful comparison between model predictions and empirical data.

We respectfully disagree with the claim that the model simulations (Fig. 3) are misleading. The simulations are intentionally presented to highlight a scale-invariant property of the models: regardless of the specific sigma (i.e., scatter) value imposed, both models consistently predict lower variability at *iso* compared to *ortho*. This holds across the full range of parameters and is central to our argument.

Our choice to calibrate the simulations based on a range of plausible sigma values, rather than aligning the amplitude of bias, reflects the central focus of our study on scatter (sigma) rather than the precise magnitude of bias. Importantly, due to their different formulations, the Cicchini et al. model and the Bayesian model yield substantially different bias magnitudes for the same sigma. Thus, attempting to

16

directly match the empirical bias amplitude (e.g., 0–2) across models would be arbitrary and potentially misleading (still, it won't change the results and main point).

Instead, the purpose of the simulations is to demonstrate that for all values of sigma (and hence corresponding bias values), empirical evidence does not support the predictions of these models— namely, lower variability at *iso* with respect to *ortho,* which converges towards the baseline. We have now clarified this point more explicitly in the caption of the relevant figure.

Reviewer #2 (Remarks to the Author):

Major point 1: The authors have convinced me that the history dependency in stimulus-specific bias and serial dependence are likely two distinct phenomena. This is mainly because of their point I (no correlation between serial dependence and stimulus-specific bias reduction).

Major point 2: I am not convinced by the authors' reply to my comment, but I am inclined to think that this issue does not invalidate their conclusions. I will try to explain what I found unclear, and perhaps it may help to provide a more specific explanation in the paper.

I am fully on board with the authors' reasoning that their debiasing method removes most of the stimulus-specific bias, and therefore reduces artifactual error scatter in the serial dependence analysis. This is what the authors reiterated in their reply. However, my focus is on the specific question about the iso-ortho difference in error scatter in the serial dependence analysis. The authors claim that this difference occurs because the stimulus-specific bias is slightly reduced in the iso *relative* to the ortho condition, leading to a slightly smaller artifactual increase in error scatter in the iso condition. In their reply, the authors show that this *relative difference* in stimulus-specific bias between iso and ortho remains intact. When the relative difference in bias remains intact, this raises the question how the confound in iso-ortho error scatter in the SD analysis, which came about due to the relative difference in stimulus-specific bias, has been removed. I do not think that the authors have answered this question. Here is my interpretation: After debiasing, there remains a small attraction to the oblique in the ortho condition. This will lead to a small increase in error scatter in the SD analysis. In the iso condition, there now is a small *repulsion* from the obliques, which will also lead to a small increase in error scatter in the SD analysis. Ortho repulsion and iso attraction biases, although of opposite sign, have a similar magnitude and therefore will lead to similar (small) artifactual increases in error scatter. Therefore, by matching the amplitude of iso and ortho biases (but not their relative difference), the small remaining artifactual increase in error scatter cancels out.

We would like to thank the Reviewer for their comments throughout the revisions, which made the point of our manuscript clearer and stronger.

In fact, we agree with the Reviewer. The relative difference remains intact, showing that our cleaning did not artificially remove this difference. Instead, it shifted the curves in a way that overall effect of stimulus-specific bias residuals are still separable between conditions, but minimally affect serial dependence analyses (i.e., error as a function of delta). We now clarified this point in the relevant section in our supplementary materials file, by adding the following sentence:

"*Notably, the differences in stimulus-specific biases between iso and ortho conditions are preserved, but similar residual peak-to-peak amplitudes led to a fair comparison of error scatter for these conditions.*"

Reviewer #3 (Remarks to the Author):

I appreciate the authors' great revisions and clarifications. I have no further questions. Congratulations on the excellent work!

We would like to thank the Reviewer for their positive feedback and kind words of appreciation regarding our work.