June 7, 2021

Dear Editor,

We thank the reviewers for their comments and thank you for your action letter.
We have addressed all the points raised by both reviewers and as a result our revised manuscript is strengthened and with a wider scope. Changes to the manuscript are marked in blue ink.

We have expanded our discussion of how this model contributes to our understanding of the different mechanisms of exogenous and endogenous attention and of their effects on visual perception. We have also demonstrated the generalizability of the model by showing that we can successfully apply it to different tasks –acuity and contrast sensitivity– for which the effects of these two types of attention have similar and differential effects, respectively. We have added the best-fitting parameters in **Table S6**.

Given that now we model in more detail the neutral condition of contrast sensitivity and add the attention conditions, we no longer include the original Supplementary Figure 1; instead we have added a new figure in the main manuscript (**Figure 9**) as well as two other supplementary figures (**Figure S6 and S7**).

We hope you find our revised manuscript ready for publication in PNAS.

Looking forward to hearing from you,

Michael Jigo, David J Heeger & Marisa Carrasco

Editor Comments:

The manuscript would benefit greatly by a clearer articulation of how the model brings additional insights to our understanding of the differential mechanisms and consequences of exogenous x endogenous attention, as well as by demonstrations of the generalizability of the model to different stimulus and task sets.

Reviewer #1:

Suitable Quality?:Yes
Sufficient General Interest?:Yes
Conclusions Justified?:Yes
Clearly Written?:Yes
Procedures Described?:Yes
Supplemental Material Warranted?:Yes

Comments on Significance Statement: It is concise, clear and easy to read.

Comments:
In the study "An image-computable model on how endogenous and exogenous attention differentially alter visual perception" Jigo et al. propose and validate a novel attentional gain model to characterize how exogenous and endogenous attentional modulation of texture target detection differs. The model is based on multiplicative modulation of inputs drive and divisive 'surround' normalization. The model is applied to a rich set of ten influential psychophysical studies varying spatial frequency and eccentricity amongst others in addition to the attentional cueing conditions.

The paper’s novel finding is that endogenous and exogenous attention effects on behavior are exceptionally well accounted for by differences in the spatial frequency tuning causing different biases in segmenting figures from backgrounds at different eccentricities.

The authors also show that the proposed novel model architecture can also be applied to other experimental variations than different texture scales, including variations with different task demands (focusing on orientations rather than texture), and multiple spatial locations.

The manuscript is well written with established model fitting mechanisms, sounds statistical analysis and clear visualization of results. The conclusions of the manuscripts are fair and a broad literature is considered to support the study. It is particularly acknowledged that the major model is evaluated statistically against multiple reduced models, and that the authors report how much variance the model explains across the multiple experiments it is fitted to.

This is a very mature manuscript with an exceptional quality of methods and convincing results. There were only a few aspects that deserve consideration.

— Thanks for your positive review and helpful comments. Addressing them has strengthened our manuscript.

- A limitation of the study is that endogenous and exogenous modulation are fitted independently, because the results do not clarify how both sources of attention interact.
This is a limitation given that it is known that symbolic and explicit cueing have different spatial profiles (also the texture based insights seem novel and very interesting). One way to address this limitation is by extending the discussion on how the different characteristics of endogenous and

exogenous modulation might interact during task processing / perception. Maybe it is useful to assume that in exogenous cueing conditions subjects have a broad/wide endogenous attentional field? Or that symbolic cues there is no exogenous attentional amplitude?

— It is unknown how interactions between endogenous and exogenous attention affect texture segmentation. We have clarified that the studies we have modeled cannot address this question and that we can only speculate on the possible effects of exogenous-endogenous interactions during texture segmentation [**p.18 para. 2**].

- The authors justify keeping the spatial spread of attention constant at 4 degrees because this reflects the largest target size used. They mention that 2 or 3 degrees constant spread would provide similar results. It is unclear why the spatial spread is not a freely varying parameter when fitting the data of individual experiments. It seems possible that the spatial spread can tune finer or wider to modulate spatial frequency based segmentation. A more explicit consideration of the role of the spatial spread in how it modifies and interacts with the improved (far) or decreased (near) performance is important. This could also help reconciling the results with previous models that explicitly fit the spatial spread.

— We note that there are no previous models of endogenous and exogenous attention effects on texture segmentation. Instead, current models that allow the spatial spread to vary (e.g., Reynolds & Heeger, 2009) capture attentional effects on low-level visual processes (e.g., contrast sensitivity). We have now clarified that that our model comparisons assessed the separate contributions of SF gain and the spatial spread of attention to texture segmentation [**p.12 para. 1**] and expanded the Discussion to clarify that the contributions of the spatial spread differ between low-level and mid-level visual tasks [**p.14 para. 3**].

- The authors report Akaike Information Criterion differences between model. Is there a reason why AIC is used? Bayesian Information Criterion is a more careful strategy (penalizing with the log of the number of observations and not with mere n of observations as AIC does). Would the best model remain the best predicting model when BIC is used?

— We have repeated our model comparisons with BIC and found identical results. We now report BIC scores in our Results [**p.9 para. 3; p.12 para. 2**] and in **Figure 4D** [**p.33**] and **Figure 7** [**p.37**].

Other aspects:
- Figure 4 legend. It is unclear what 'simultaneously' refers to here: "The model predicted performance simultaneously with the fit shown in A". Can you clarify.
— We have clarified the figure caption of **Figure 4** [**p.33**] and similar language in **Figures 5** [**p.35**], **6** [**p.36**] and **8** [**p.38**].

- Suppl. Figure C is wrongly denoted D, please correct
— We have corrected **Figure S2** [**SI Appendix, p. 10**].

- line 492 why capital THAT?
— This has been fixed.

Reviewer #2:

Suitable Quality?:No
Sufficient General Interest?:No
Conclusions Justified?:Yes
Clearly Written?:No
Procedures Described?:Yes
Supplemental Material Warranted?:Yes

Comments:
This paper reports a model of attention that differentiates between distinct effects of exogenous and endogenous attention that have been observed during texture segmentation tasks. The model is based on the prominent normalization model of attention that has been shown to explain a variety of behavioral and neural effects in the past (e.g., effects on contrast sensitivity). The present research expands this model by implementing two different attentional gain profiles for spatial frequency (narrow and broad) that are then fit to behavioral results of past research using texture segmentation tasks.

The paper addresses an interesting question in the attention literature, namely why attention effects in texture segmentation tasks are different for exogenously and endogenously cued locations. The model itself seems straightforward and fits behavioral data very well, which is great to see. However, I was not compelled that the main conclusion - namely that exogenous attention boosts fine SF while endogenous attention boosts fine and coarse SF across the visual field was that novel, as this has been strongly suggested by previous data. While it is certainly important to show this computationally as well, I was not sure about the novelty of the results and modeling itself.

— Thanks for your comments; addressing them has significantly strengthened our manuscript and broadened its scope.

— We agree that it is important to demonstrate the computational validity of suggestions raised by previous data. Without a computational framework to test, simulate and predict behavior, the mechanisms underlying visual and cognitive function would remain unclear. We now expand on how our model improves upon suggestions by previous data [**p.15 para. 1-2**].

Thus, while I don't see anything wrong with the paper, I felt it might be more fitting for a more specialized journal. Below I am outlining just a few issues that I thought made the paper a bit too narrow in its scope.

It seemed that the model's parameters were highly specific to texture segmentation. It is argued that texture segmentation here serves as a "model system" to examine differences in exogenous and endogenous attention effects. However, in many cases exogenous and endogenous attention produce the same effects on visual performance (and related neural processes), so does texture segmentation serve as a good model system in the sense that the findings here would generalize to other tasks?

— We now demonstrate the generalizability of our model to other basic visual tasks: (1) acuity for which exogenous and endogenous attention have similar effects, and (2) contrast sensitivity, for which these two types of attention have different effects [**p.4 para. 4; p.12 para. 4; p. 13; p.18 para. 1; Figure 9; Figure S6 & S7**].

Or is it rather the case that different parameters need to be implemented and adjusted for each stimulus/task? It would potentially be useful to see how variations in the SF gain parameter relate to other, non texture-segmentation tasks, for example.

— We now clarify that the model fits to texture segmentation experiments shared parameters among different stimuli and tasks [**p.8 para. 2**]. We also specify that **no** additional parameters were implemented for our model to generalize to acuity and contrast sensitivity in the Results [**p.12 para. 4**] and in **SI Appendix, section S6**.

Relatedly, it would have been neat to see predictions of the model across tasks and stimuli (maybe even more naturalistic ones that rely on texture segmentation). Instead - if I understood correctly, the model was always fit to the data in each experiment separately.

— We have clarified that the model was **not** fit to each experiment separately [**p.8 para. 2**]. The goal of the current study was to explain endogenous and exogenous attention effects in existing datasets– now, we include texture segmentation, acuity and contrast sensitivity. It is beyond the scope of this study to assess attentional effects on naturalistic textures; however, future work may do so.

Furthermore, I would have liked to see a stronger test of other proposals in the literature that potentially explain texture segmentation. For example, previous work has related exogenous effects in texture segmentation in part to limits in temporal resolution. Could these temporal limits be implemented in the model as well, and potentially explain the behavioral effects? Or put differently, how do assumptions about temporal profile relate to the SF gain parameter?

— We now mention the temporal account for the CPD in texture segmentation and how it has been criticized. Furthermore, we explain why established effects of exogenous and endogenous attention on the speed of information accrual **cannot** explain the differential effects of exogenous and endogenous attention on texture segmentation [**p.16 para. 4**].