**Masoud Ghodrati,**
*Neuroscience Program, Biomedicine Discovery Institute,*
*Monash University, Clayton, VIC., Australia*
E-mail: `masoud.ghodrati@monash.edu`

**Journal of Vision**
*USA*
*The editors*

Melbourne, 30/11/2018

Dear **Editor**,

The authors would like to thank the respected reviewers and editorial members who kindly participated in the review process. Please find enclosed the new version of our manuscript "Object categorization in visual periphery is modulated by a delayed foveal noise". We carefully took into account the reviewers' comments which helped us to increase the quality of the manuscript. All the suggested refinements are performed and the manuscript is revised accordingly. We hope that the new version of the manuscript is suitable for publication in *Journal of Vision*.

Yours sincerely,
**Masoud Ghodrati**, PhD

Masoud Ghodrati

---

## Reviewer #1

This paper reports on two experiments that investigated object recognition in the visual periphery, and the effects of foveal interference on such peripheral object recognition. Experiment 1 investigated the conjoint effects of retinal eccentricity and categorization level on object recognition. The results showed that superordinate object categorization was quite good from 0-24 degrees eccentricity, but that categorization was increasingly worse at greater eccentricities for the basic level, and even more so for subordinate level categorization, and RT results were similar. Experiment 2 was similar to Experiment 1, except that a foveal dynamic noise mask was presented after 100-400 ms SOA. The results showed that accuracy was worse at 300 ms SOA for the basic and subordinate level tasks, but superordinate level task performance was unaffected by foveal masking. In addition, the effects of foveal masking were eccentricity dependent, with significant effects at 24 degrees for the basic level task, and 18 degrees for the subordinate level task. Conversely, RTs were uniformly higher for all foveal masks, with longer RTs for longer SOAs for all three categorization levels. The authors discuss their results in terms of the reverse hierarchy theory.

Overall evaluation:
This was a nicely designed, carried out, and analyzed study. The discussion is OK, but seems a bit weak. But overall it is a good and interesting study. I have only a small number of minor comments for improvement. Those are mostly in terms of other related phenomena that you should consider and try to make connections with.
Comments:
p.2: "peripheral vision should coordinate with central vision for planning the next gaze location on a target object while viewing a scene (Ludwig, Davies, & Eckstein, 2014)."
I think you may also want check out the work by Golomb et al. (Golomb, Chun, & Mazer, 2008; Golomb, Nguyen-Phuc, Mazer, McCarthy, & Chun, 2010) and Rolfs et al. (Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011) who have looked at this issue in the area of visual attention. In addition, you should also consider the work on parafoveal preview in object recognition (Henderson, 1992; Henderson, Pollatsek, & Rayner, 1989). All of this work seems relevant to your study.

> **Reply**: We reviewed these papers in the introduction, page 2 line 44:
> "It is suggested that covert attention facilitates this coordination while eyes are saccading elsewhere (Golomb, Chun, & Mazer, 2008; Golomb, Nguyen-Phuc, Mazer, McCarthy, & Chun, 2010). This pre-saccadic visual attention enhances perceptual performance at the target location (Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011; Colby, Goldberg, et al., 1992; Sommer & Wurtz, 2006; Gottlieb, Kusunoki, & Goldberg, 1998). Earlier studies have also shown that the presentation of parafoveal flanker objects, which are relevant to the fixated target, can facilitate identification of the target object (Henderson, 1992; Henderson, Pollatsek, & Rayner, 1989)."

p. 2: "In particular, using artificial images, it has been shown that a $\sim 250$ ms delayed foveal noise impairs the peripheral object recognition (Fan et al., 2016)."
Note that they also showed strong effects at 50 ms SOA, but may have been due to attentional capture to the foveal stimulus. You should briefly discuss this at some point.

> **Reply**: In (Fan et al. 2016) it is explined that the accuracy drop at 50 ms SOA is actually due to the attentional distraction caused by onset of foval noise when the peripheral object stimuli were on the screen. Notably, by reducing the size of the focal noise mask (from $7^\circ$ to $4^\circ$), the accuracy drop at 250 ms remained significant, while at 50 ms SOA, the effect was not significant anymore. We have discussed these in the third paragraph of the Discussion section page 10 line 260:
> "This time course is consistent with previous reports (Fan et al., 2016), where 250 ms SOA between noise and target was shown to have a significant drop in accuracy. Although (Fan et al., 2016) found an accuracy drop at 50 ms SOA as well, this might be related to the attentional distraction caused by the noise onset as the effect disappeared by reducing the size of the noise image."

Figure 1 caption: "Then, an object image was randomly presented on one of the nine locations" => add "at 0, or +/- 6, 12, 18, or 24 deg eccentricity"

> **Reply**: Caption refined.

p. 6: "This accuracy drop was highest in subordinate level and lowest in subordinate"
=> both the highest and lowest cannot both be the subordinate. I believe the lowest was superordinate.

**Reply**: It was a typo and corrected in the revised manuscript.

p. 8: "There was a similar, but relatively smaller, decrease in accuracy at 12° and 18° eccentricities; however, this did not reach a significant level (Figure 4C, $p > 0.05$)."
You should report all p values (except those $< 0.001$), as per APA guidelines.
Also, this could be due to a lack of power, since you only had 10 subjects. You should do a power analysis and report what N-size would be needed to find significant effects with the same effect sizes for these other eccentricities.

**Reply**: The exact p-values were added to the revised manuscript ($p = 0.37$ for 12°, $p = 0.87$ for 18° eccentricities). Moreover, we performed power analysis and added the relevant information to page 7, line 203:
"This could be due to a lack of power as our power analysis suggested that 15 subjects are required to find significant effects with the same effect size for these other eccentricities."

p. 8: "We found a significant decrease of 8.6% in accuracy at 300 ms SOA relative to no noise in subordinate level categorization only at 18° eccentricity (Figure 4D, $p < 0.01$)."
The lack of difference at 24 deg was probably due to a floor effect. That is worth mentioning.

**Reply**: We updated the paragraph and mentioned it.

Figures 4a & 5a: For the two line graphs, I suggest using somewhat lighter shades of red, green, and blue for the line graphs, so it is easier to discriminate the green and blue lines. Currently, they are so dark that it is hard to discriminate them.

**Reply**: We changed the line styles rather than color. This makes it possible to differentiate different curves/lines in case of black-and-white print.

p. 9: "However, unexpectedly, the reaction times of the superordinate level were higher than basic level and almost similar to subordinate level."
This is what one would expect from Rosch's Basic Level theory of categorization. However, her theory cannot account for the superordinate advantage for accuracy. You should mention this.

**Reply**: We extened the text at Page 7 line 214 to the following:
"In other words, reaction times in basic level were less affected by the foveal noise than the superordinate and subordinate levels. This difference in reaction time might support Rosch's Basic Level theory of categorization, where basic level is proposed to be the first and most inclusive categorization that is made during perception of the environment (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). However, it cannot explain the lower accuracy for basic level with respect to the superordinate level (Figure 4) and shorter reaction times for superordinate level in no noise condition (Figure 5)."

p. 9: "However, some recent studies have shown that the peripheral vision is able to solve complex object categorization tasks including natural object discrimination (S. J. Thorpe et al., 2001; Boucart et al., 2010; Boucart, Moroni, Szaffarczyk, & Tran, 2013) and face recognition (Hershler et al., 2010; Crouzet et al., 2010) at the very far periphery (Boucart et al., 2016)."
Peripheral vision has also been shown to be very useful for rapid scene categorization (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013; Larson & Loschky, 2009), and more so than for object recognition (Ehinger & Rosenholtz, 2016).

**Reply**: We cited these relevant papers in the revised manuscript.

p. 10: "These results are consistent with previous reports suggesting that categorization in finer levels (e.g., subordinate) relies more on higher spatial frequencies (Ashtiani et al., 2017)."
This was also shown for scene categorization by Boucart et al. (2013).

**Reply**: We added the citation to Boucart et al. (2013).

p. 10: "In further analysis, we realized that this reliance on high spatial"
[Sentence fragment] => finish thought

**Reply**: It was a sentence that were supposed to be completely removed, but seems that some parts of it were mistakenly remained. So, we completely removed this sentence.

p. 10: "Such primary activations before saccades have been found in other cases; for instance, neurons in some brain areas become responsive to the visual stimuli that will be brought into their receptive field by an imminent saccade (Walker, Fitzgibbon, & Goldberg, 1995; Melcher, 2007; Cavanagh, Hunt, Afraz, & Rolfs, 2010)."
This is related to the well-established finding of parafoveal preview (Henderson, 1992; Henderson et al., 1989).
You should also discuss the relationship between 1) this finding of the effects of disruption of foveal processing on disruption of peripheral processing, and 2) the well-studied effect of foveal load on peripheral processing (Ball, Beard, Roenker, Miller, & Griggs, 1988; Ikeda & Takeuchi, 1975; Mackworth, 1965; Ringer, Throneburg, Johnson, Kramer, & Loschky, 2016; Williams, 1985).

**Reply**: We added and discussed the related papers. The corresponding paragraph in the main text was updated accordingly (page 11, line 284):
"This is consistent with the well-established finding of parafoveal preview (Henderson et al., 1989; Henderson, 1992), where parafoveal presentation of visually related flankers to the target object facilitates identification of the target. Moreover, physiological measurements in several studies have shown pre-saccadic activations in several brain areas where neurons become responsive to the visual stimuli that will be brought into their receptive field by an imminent saccade (Walker, Fitzgibbon, & Goldberg, 1995; Melcher, 2007; Cavanagh, Hunt, Afraz, & Rolfs, 2010). Some recent studies have also shown that pre-saccadic spatial attention (i. e., presenting a prior spatial cue at the location of peripheral target stimuli) can facilitate recognition of crowded objects (Harrison, Mattingley, & Remington, 2013; Harrison, Retell, Remington, & Mattingley, 2013; Wolfe & Whitney, 2014).
Note that the accuracy drop in our peripheral object categorization task with foveal delayed noise cannot be due to the foveal load, a phenomenon also known as tunnel vision (Ball, Beard, Roenker, Miller, & Griggs, 1988; Ikeda & Takeuchi, 1975; Ringer, Throneburg, Johnson, Kramer, & Loschky, 2016; L. J. Williams, 1985). The tunnel vision effect happens when subjects perform a sufficiently difficult foveal task that has a high priority in which the subject needs to respond rapidly (Ringer et al., 2016). In such experimental paradigms, subjects might lose their attention to the peripheral task. However, firstly, in our experiments, subjects were not asked to do any specific foveal task apart form fixation. Secondly, if there was any tunnel vision effect, it must have happened for all categorization levels while we observed that the foveal noise had no significant effect on superordinate categorization. Thirdly, the attention disruption in tunnel vision effect is

expected to be stronger on shorter SOAs while in our experiments, the accuracy drop was the highest at 300 ms SOA."

Grammar Comments:
p. 1: "at high abstract" => "at highly abstract"
p. 2: "The categorization accuracy at the visual periphery..."
=> "Categorization accuracy in the visual periphery..."
p. 2: "It is not yet clear how the peripheral vision processes object at each of these categorization levels."
=> "...the peripheral vision..." => "...peripheral vision..." [delete 'the']
=> "...processes object at..." => "...processes objects at..." [add plural -s]
p. 2: "The peripheral vision should coordinate with central vision..."
=> "Peripheral vision should coordinate with central vision..."
p. 2: "...while object is presented foveally..."
=> "...while an object is presented foveally..." [add 'an']
Figure 1 caption: "(D) Sample images from different category."
=> "(D) Sample images from different categories." [add plural -ies]
p. 6: "...accuracy drop was highest in subordinate level..."
=> "...accuracy drop was highest in the subordinate level..." [add 'the']
p. 9: "...psychophysical manipulation of these feedback information..."
=> "...psychophysical manipulation of this feedback information..." ['these' => 'this']

**Reply**: We corrected them.

---

## Reviewer #2

Reviewer summary: There is evidence in the literature that brain areas later in the visual stream send feedback to earlier areas, perhaps to clarify in more precise detail what was seen. This late signal has been purported to go from regions retinotopic to the periphery to regions retinotopic to the fovea. In particular, showing noise in the fovea roughly 200-400 ms after the onset of a peripheral stimulus decreases performance on tasks which depend on the peripheral stimulus. The authors extended this idea to a categorization task with cropped natural images, where subjects had to categorize objects at subordinate, basic, and superordinate abstraction levels, e.g. pigeon vs hawk, bird vs dog, and animal vs non-animal, respectively. They found that at about 300ms, foveal noise significantly inhibited peripheral categorization ability. They argue that this supports a type of "reverse hierarchy theory" of vision.
I found the methods and scholarly aspect of this paper to be rigorous and clearly informed. The experiment was set up in a straightforward way, and the analysis was similarly straightforward. The authors answered their question about whether peripheral object categorization is impaired by foveal noise.
My main criticisms of the paper are threefold:
(1) There is inadequate justification as to how this is significantly different than the existing research on foveal interruption of peripheral vision. The authors state that they use "natural" images, while prior work uses artificial images or natural scenes; they do not justify why natural images should be any different. Is there a reason to suspect that their specific task will be more or less prone

to foveal interference? With respect to (1), the authors should cut down the introduction on very basic information, e.g. the list of things peripheral vision can and cannot do second paragraph on page 2. This can be replaced with discussion about why they might expect something different than what is already found, or why if they found the same thing, it would be important. For example, if they found that there was no foveal interference, would that disprove the "reverse hierarchy" class of models?

**Reply**:

Previous studies have addressed object categorization in peripheral vision using natural images (e.g., Boucart et al. Journal of vision, 2016; Boucart et al. Vision Research, 2013). Studies on foveal feedback, however, either used novel 3D objects (with no abstract category levels) that mostly involve higher level visual areas (Williams et al. Nature Neuroscience, 2008; Chambers et al. Cortex, 2013), or only manipulated basic dimensions of the peripheral objects, such as orientation of the object (Yu and Shim. Journal of vision, 2016; Fan et al. PNAS, 2016), to study the foveal feedback effect for the processing of high/low-level visual features. Our study is the first that systematically measures the performance of human peripheral vision (across different eccentricities) in object categorization in three different abstraction levels, and re-measures this performance while manipulating the foveal presentation to study the effect of feedback. Therefore, we have two main findings in one study: We found that higher category levels 1) are easier to categorize in visual periphery and 2) are less impacted by the delayed foveal noise. Our findings suggest that higher category levels are more likely to be explained by bottom-up responses/attention (without the need for feedback information from periphery) while foveal feedback contributes to discriminating more complex objects (i.e., basic and subordinate) in periphery, and that's why the noise mask had disrupting effects while categorizing these images.

As suggested by the reviewer, we removed the second paragraph of Introduction and replaced it with some content related to visual crowding. Moreover, we rewrote the last paragraph of our Introduction to better state the contributions of our study:

"Previous studies have used novel three-dimensional artificial object images to investigate how manipulation of foveal information can affect categorization of peripherally presented objects (Chambers et al., 2013; M. A. Williams et al., 2008). Other studies addressed this issue by manipulating basic dimensions of the peripheral objects, such as orientation of the object (Fan et al., 2016; Yu & Shim, 2016). As categorization of these objects mostly involve higher level visual features and higher areas in the brain, it remains unclear how manipulation of the foveal feedback information can affect the processing of other features. Here, we measure the performance of human peripheral vision in natural object image categorization while manipulating subjects' foveal presentation to study the effect of feedback information. In particular, we first performed a series of psychophysics tasks to measure humans' categorization accuracy and reaction time at the central and peripheral fields in three categorization levels (i.e., superordinate, basic, and subordinate). We then ran further psychophysics experiments to investigate how modulating subjects' foveal representation can affect their peripheral categorization of object at different abstraction levels. We found that higher category levels are easier to be categorized in periphery and are less impacted by the delayed foveal noise. Particularly, the accuracy of peripheral object categorization at superordinate level was not affected by the foveal noise, while it significantly dropped at subordinate level. This result suggests that peripheral visual information is sufficient for high level categorization (e.g., spotting that it is an animal); however, a finer categorization (e.g., detecting it is a pigeon) should be done by making a saccade toward the target object.

Our findings also suggest that higher categorization levels are more likely to be explained by the bottom-up responses (i.e., without the need for feedback information from periphery), whereas foveal feedback contributes to categorization of more complex objects (i.e., basic and subordinate) in the periphery, as it was shown by disrupting the consistency between the initial coarse peripheral information and detailed foveal information, which was forced by the foveal noise."

(2) They do not mention crowding anywhere, despite the fact that it is the biggest limitation in peripheral vision; in fact, they misrepresent a paper (Rosenholtz 2016) as claiming "These studies show the capability of the peripheral vision in a number of visual tasks... result of lower acuity and contrast sensitivity in peripheral vision." On the contrary, Rosenholtz 2016's main argument is that acuity is a minor loss when compared to crowding. It is important to note here that crowding occurs within a stimulus, given that it is complex enough; it is not enough to suggest that since these are single objects crowding should not happen.
For (2), the discussion in the introduction needs to include crowding, and if not, they need to justify why crowding is not relevant. I don't think that crowding-specific experiments need to be added, rather they need to acknowledge its effect in peripheral vision. Importantly, they need to re-examine and correct how they cite Rosenholtz 2016.

**Reply**: We corrected the citation to Rosenholtz 2016. Moreover, we added the crowding to the introduction (page 2, line 36):
"The behavioral studies in humans show that although fine details can be recognized in isolated peripherally presented objects, perception is impaired when those objects are surrounded by clutter (Rosenholtz, 2016; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). This "crowding" phenomenon may arise from several mechanisms such as spatial pooling of peripheral information (Pelli, 2008), poor resolution of attention in the peripheral field (He, Cavanagh, & Intriligator, 1996), and substitution between target and flankers' features (Ester, Klee, & Awh, 2014; Huckauf & Heller, 2002). However, it is still unclear which stimulus features are not represented in the brain and therefore lost in peripheral vision. In fact crowding is the biggest limitation in peripheral vision when compared with lower peripheral acuity (Rosenholtz, 2016)."

Please note that although crowding can be an important factor in any peripheral task, here we did not aim to measure the effect of crowding on subjects' performance. It is also important to point out that making a direct relationship between crowding and our results is a bit complicated as we presented the same object images in different categorization tasks – i.e., the same objects could accurately be categorized in superordinate level but subjects were less accurate in categorizing the same images in subordinate level.

(3) They repeatedly claim that higher-spatial frequency content is what makes subordinate and basic categorization harder than superordinate in the periphery, but do not test this explicitly. Given that it is relatively trivial to filter out spatial frequency bands from images, it is not unreasonable to expect them to actually test their claim directly. Given that crowding is the main limitation in peripheral vision, I doubt it will pan out. On the other hand, if this is true, then the paper amounts to re-discovering that some high spatial frequencies are lost in peripheral vision which would be interesting, given crowding limitations. For (3), there either needs to be explicit modeling of spatial frequency information, a separate experiment where they selectively remove high spatial frequencies and show that subordinate and basic category performance decreases to superordinate, or a strong justification as to why they do not find this necessary to do so. Alternatively, they could leave out the spatial frequency discussion, but that seems unsatisfactory. It is not adequate simply to

speculate that spatial frequency is driving their findings, without explicitly testing it. If the other parts of this paper were very novel, it might justify not doing this, but in its current state that is not the case.

**Reply**: Previous studies on foveal vision have shown that lower spatial frequencies are sufficient for superordinate categorization, while higher frequencies are necessary to basic and subordinate categorization (Collin et al. 2005, Perception & Psychophysics). Here, our aim was not to study how spatial frequency affects peripheral object categorization and we have directly addressed this question in our previous study (Ashtiani et al. 2017, Frontiers in Psychology) and showed the contribution of different frequency bands on different categorization levels in fovea. In the manuscript, we only mentioned this topic in the discussion where we talk about the contribution of different spatial frequency bands on categorization levels in fovea.

We reworded this part to avoid any misleading claims:
"The lower visual acuity in peripheral vision might only be sufficient to solve superordinate categorization tasks while such poor spatial details in visual periphery, which can be the result of visual crowding, results in lower categorization accuracy for the basic and subordinate levels. It is known that peripheral object recognition is impaired in case of crowding (Rosenholtz, 2016), even when a single complex object lies in periphery, the "self-crowding" effect can disrupt the recognition accuracy (Martelli, Majaj, & Pelli, 2005)."

Note that there are no line numbers in the file I received, so I will try to pinpoint within pages and paragraphs in my list of issues.

**Reply**: We added line numbers to the revised manuscript.

Technical issues that need to be addressed in the paper:
As far as I can tell, the authors make no mention of whether the images are color or grayscale. Same with the noise mask.

**Reply**: Image were grayscaled. We accordingly updated the the last sentence of the first paragraph of the "Image Set" section.

How did the authors choose their stimuli from the huge dataset? Was it randomly, or was there some selection process?

**Reply**: We first selected a large set of images that the target object covered a large portion of the image and then we randomly drawn the final image set from the initially selected image set. The information added to the revised manuscript (Page 3, line 91):
"First, we selected a larger set of images in which the target object covered a large portion of the image, we then randomly selected the final image set from the initially selected set."

On page 5, paragraph 1, the authors claim that they used 45 images per location, and 9 locations. This multiplies to 405 trials, despite their earlier claim that there are 400 trials total. Please clarify.

**Reply**: The information added to the text: We controlled to have on average 45 images at each location (i.e., in some locations 44 images were presented while in others 46).

I would like the authors to present more detail on how incorrect trials were rejected. Only dropping two trials per subject seems very low, if they reject blinks, etc. 2 degrees of slack in fixation is also quite large, given that the closest stimulus is 6 degrees; they need to justify this arbitrary parameter better.

> **Reply**: We removed trials based on the distance between the fixation point and where the subject were looking at. Although a few subjects were less accurate in keeping their gaze around the fixation point, the majority were good at the job. Note that incorrect trials were not removed as the accuracy was calculated based on the ratio of correct responses. Moreover, the rejection window for fixation was actually 1 degree and it was a typo in the previous version of the paper. The information was updated in the manuscript (Page 5, Line 148). "We removed trials in which subjects moved their eyes to any location more than 1 degree of visual angle from the fixation point or made a blink on the time of the target (or noise) presentation (3.58% and 5.6% of trials were removed from experiment 1 and 2 respectively)."

The authors use a "dynamic" noise mask, but never define what the "dynamic" part is. Is it changing during the duration?

> **Reply**: Yes. Updated the text as "dynamic (continuously changing) 1/f noise mask".

Style issues that need to be addressed in the paper:
The plots on page 6 should have more vertical markers; it is hard to tell the values with only markers at 0.5, 0.75, and 1. Furthermore, the data points do not align with their eccentricity markers on the plot, or with each other. When I draw a vertical line up from the marker for 6 degrees on Figure 2A, for example, only the red point aligns. This is similar for Fig 2B. The captions also do not use consistent spelling, using hyphens in "sub-ordinate" instead of "subordinate".

> **Reply**: We had jittered the datapoints place to prevent them masking each other in case of similar values and make them more visible. We removed it. Also, we refined the y axis of our plots.

Beginning of page 8, line 1 should be "Although a few studies". I think leaving out "a" gives the wrong connotation.

> **Reply**: Fixed.

The caption for figure 4 ends in "(SOA=0)", which is misleading, since an SOA of zero would still include noise; I would just remove this or replace it with "(leftmost data point)"

> **Reply**: Fixed.

The last sentence of the first paragraph of the discussion, "Moreover, recent findings..." claims "there is a novel feedback system in the brain...", which sounds like the brain system itself is somehow new, rather than the idea of it. Please reword.

> **Reply**: We removed it.

The first line of the abstract makes the claim "Although human peripheral vision is not specialized for fine-level object categorization..." This is too vague and unsupported by the literature; peripheral vision can do many things and is not "specialized" for any particular purpose. Please remove or qualify this statement.

**Reply**: We changed it as "Despite the lower visual acuity in peripheral vision, it can still do object recognition to some extent."

The statement "Therefore, the peripheral information somehow needs to be sent to the foveal parts for planning the next saccade." In the fourth paragraph on page two needs to be made precise and justified, moved later, or removed. What are the "foveal parts"? Are they in the brain, in the retina, or in some model?

**Reply**: We removed that sentence.

The authors give a potential explanation of their delayed noise finding in the second to last paragraph of the discussion. They suggest that there is some mismatch between what the fovea expects to see after the saccade and the noise, which inhibits perception; they do not mention existing work on pre-saccadic effects in foveal vision, such as Wolfe & Whitney (2014) "Facilitating recognition of crowded faces with presaccadic attention", and other Whitney lab studies. The introduction of that paper has several references I think should be mentioned in the discussion.

**Reply**: Thanks for introducing such a related paper. We added it to the Discussion:
"Some recent studies have also shown that pre-saccadic spatial attention (i. e., presenting a prior spatial cue at the location of peripheral target stimuli) can facilitate recognition of crowded objects (Harrison, Mattingley, & Remington, 2013; Harrison, Retell, Remington, & Mattingley, 2013; Wolfe & Whitney, 2014)."

Typographic or grammar errors:

"main" is misspelled as "maian" on line 5 of page 6.

**Reply**: Fixed.

The second sentence on the 4th paragraph of page 8, "Although the amount of noise SOA significantly affected..." has awkward grammar and is difficult to understand. Please reword.

**Reply**: We rewrote the whole sentence as:
"Although subjects' reaction time in all categorization levels and eccentricities significantly increased as a function of SOA, this effect did not follow the same trend as accuracy where we observed accuracy drops at 300 ms SOA (see Figure 4)."

Language: the authors regularly incorrectly use articles; a common one is "the foveal vision" or "the peripheral vision." "the" should be left out.

**Reply**: Corrected.

"superordinate" is used twice in second to last line of page 4, it should be "subordinate, basic, and superordinate".

**Reply**: Corrected.

Page 7, line 3 says "Figures 3B and 3D" when it should say "Figures 3B and 3E"

**Reply**: Corrected.

My questions that don't necessarily have to be answered in the text: The last line of page 9 claims that "they could still robustly categorize the images in superordinate level, even at 24 degrees of eccentricity". While true, the subjects were still above chance in the subordinate level at this eccentricity, so the statement is misleading; the subjects are better than the authors make it sound.

> **Reply**: We slightly changed the wording to: "Our results showed that although subjects' accuracy dropped as the object presented further from the fixation area, they could easily categorize images in superordinate level (accuracy of $\sim$80%), even at 24° of eccentricity. However, subjects found it more difficult to solve basic level categorization tasks, and the accuracy was not much above the chance level in subordinate object categorization at the furthest tested eccentricity (24°)."

There is no mention of crowding in the discussion as a potential explanation of experiment 1 results.

> **Reply**: As mentioned above, we now have discusssed the crowding effects on object recognition periphery both in Introduction and Discussion. However, we explained that it seems not to be the main factor in our experiments as subject could solve the superordinate tasks with similar images as the basic and subordinate levels accurately.

The introduction does not have much flow, it seems like a disjointed list of facts about which tasks can be performed in the periphery.

> **Reply**: The revised introduction has substantially changed by removing reviews on basic facts about visual periphery, discussing more related topics such as crowding and relations between foval and peripheral vision, and providing more explainations on the main ideas and findings of our research.

There is a new paper from Rosenholtz & Wijntjes (2018) titled "Context mitigates crowding: Peripheral object recognition in real-world images"; while it is too recent to require citation here, I think the authors might find it relevant to their work. It is a study of crowding in peripheral object recognition in natural scenes, often with single object stimuli.

> **Reply**: Thanks for introducing such an interesting paper. They have found that contextual information is really important for peripheral vision and it can reduce the impacts of crowding in a clutter scene. However they only studied the impacts of congruent contexts, it would interesting to know what happens if we present the peripheral target object in an incongruent context. Anyway, we reviewed this paper in the Discussion section.