Object categorization in visual periphery is modulated by delayed foveal noise

Farzad Ramezani*
Department of Computer Science, School of Mathematics, Statistics, and Computer Science, University of Tehran, Tehran, Iran

Saeed Reza Kheradpisheh*
Department of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

Simon J. Thorpe
Centre de Recherche Cerveau et Cognition (CerCo) Université Paul Sabatier, Toulouse, France

Masoud Ghodrati
Neuroscience Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

Behavioral studies in humans indicate that peripheral vision can do object recognition to some extent. Moreover, recent studies have shown that some information from brain regions retinotopic to visual periphery is somehow fed back to regions retinotopic to the fovea and disrupting this feedback impairs object recognition in human. However, it is unclear to what extent the information in visual periphery contributes to human object categorization. Here, we designed two series of rapid object categorization tasks to first investigate the performance of human peripheral vision in categorizing natural object images at different eccentricities and abstraction levels (superordinate, basic, and subordinate). Then, using a delayed foveal noise mask, we studied how modulating the foveal representation impacts peripheral object categorization at any of the abstraction levels. We found that peripheral vision can quickly and accurately accomplish superordinate categorization, while its performance in finer categorization levels dramatically drops as the object presents further in the periphery. Also, we found that a 300-ms delayed foveal noise mask can significantly disturb categorization performance in basic and subordinate levels, while it has no effect on the superordinate level. Our results suggest that human peripheral vision can easily process objects at high abstraction levels, and the information is fed back to foveal vision to prime foveal cortex for finer categorizations when a saccade is made toward the target object.

Introduction

Providing a coherent perception of the visual world requires the brain to integrate information from multiple sources and across multiple time scales. Foveal and peripheral vision work in parallel to integrate both local and global features across the visual field to provide a continuous and coherent perception of a scene. The role of central vision (i.e., foveal vision) in scene and object recognition has been extensively studied (Gauthier & Tarr, 2016). However, fewer studies have investigated the perception of complex visual objects in the peripheral visual field (Loschky, Nuthmann, Fortenbaugh, & Levi, 2017; Strasburger, Rentschler, & Jüttner, 2011). Here, we address this question, and more importantly, we study how the visual information in the peripheral field can be used by foveal vision to modulate human object categorization accuracy.

Recent studies have shown that human subjects can accurately categorize peripherally presented object and face images at highly abstract categorization levels (e.g., animal/nonanimal; Boucart et al., 2016; Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013). Experimental evidence also indicates that humans can detect animals in peripherally presented natural scenes as far as 75° eccentricity while subjects are fixating centrally (S. J. Thorpe, Gegenfurtner, Fabre-Thorpe, & BuÈlthoff, 2001). Similar results are reported for recognition of isolated objects (animal/vehicle) at far periphery (Boucart, Naili, Despretz, Defoort-Dhel-

Citation: Ramezani, F., Kheradpisheh, S. R., Thorpe, S. J., & Ghodrati, M. (2019). Object categorization in visual periphery is modulated by delayed foveal noise. *Journal of Vision*, *19*(9):1, 1–12, https://doi.org/10.1167/19.9.1.

https://doi.org/10.1167/19.9.1 Received August 21, 2018; published August 1, 2019 ISSN 1534-7362 Copyright 2019 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from intl.iovs.org on 04/26/2024

lemmes, & Fabre-Thorpe, 2010). These studies show the capacity of peripheral vision in a number of visual tasks that require detailed spatial information, although the categorization accuracy in the periphery depends on both the contrast and size of the presented image (larger scales and higher contrasts are preferred; Mäkelä, Näsänen, Rovamo, & Melmoth, 2001; Wright & Johnston, 1983).

Moreover, it has been shown that the task demand (e.g., object detection or identification) modulates the processing of objects in peripheral vision (Jebara, Pins, Despretz, & Boucart, 2009). It is important to consider the potential role of crowding in categorization as well. Behavioral studies in humans show that although fine details can be recognized in isolated peripherally presented objects and scenes, perception is impaired when those stimuli are surrounded by clutter or are internally complex. While the mechanism that leads to it is still an active topic of research, crowding is the most important limitation in peripheral vision (Rosenholtz, 2016).

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). 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 presaccadic visual attention enhances perceptual performance at the target location (Duhamel, Colby, & Goldberg, 1992; Gottlieb, Kusunoki, & Goldberg, 1998; Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011; Sommer & Wurtz, 2006;). 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).

In particular, recent neuroimaging studies have found a feedback system in the visual cortex by which the peripheral information is somehow fed back to the foveal visual cortex (Chambers, Allen, Maizey, & Williams, 2013; Fan, Wang, Shao, Kersten, & He, 2016; M. A. Williams et al., 2008). These studies revealed that the information about the peripherally presented objects can be decoded from the patterns of activation in the human foveal retinotopic cortex (Fan et al., 2016; M. A. Williams et al., 2008). Further behavioral evidence showed that presenting a delayed foveal distractor can significantly modulate the categorization of the peripheral target object (Weldon, Rich, Woolgar, & Williams, 2016; Yu & Shim, 2016). In particular, using artificial images, it has been shown that a ~250 ms delayed foveal noise impairs the peripheral object recognition (Fan et al., 2016). Such a drop in accuracy (observed in these experiments) could be due to the inconsistency between what is presented in the fovea and what has already been seen in the periphery.

In addition, it is suggested that during foveal object recognition, an initial guess about the object category is made in the prefrontal cortex based on a global gist of the image (e.g., animal; Bar et al., 2006; Chaumon, Kveraga, Barrett, & Bar, 2013; Kauffmann, Bourgin, Guyader, & Peyrin, 2015). Then, this initial guess is sent back to the inferotemporal cortex (IT) through top-down feedback connections that modulate the late detailed information processed by the ventral pathway. Many studies have tried to manipulate this flow of information at neural and perceptual levels while an object is presented foveally (Bar, 2003; Bar et al., 2006); however, it is still unclear whether this holds for peripherally presented objects. Here, we hypothesize that there might be a similar situation in the visual periphery, such that the information available to peripheral vision is sufficient to support a general guess about an object's high-level category (e.g., animal vs. nonanimal). Then, this information can guide foveal visual cortex to make a saccade toward the target object and put it in the fovea to facilitate object categorization in finer levels. Therefore, distracting this interaction between the foveal and peripheral vision should affect the categorization accuracy and response time.

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 involves 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 the superordinate level was not affected by the foveal noise, while it significantly dropped at the 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

Downloaded from intl.iovs.org on 04/26/2024

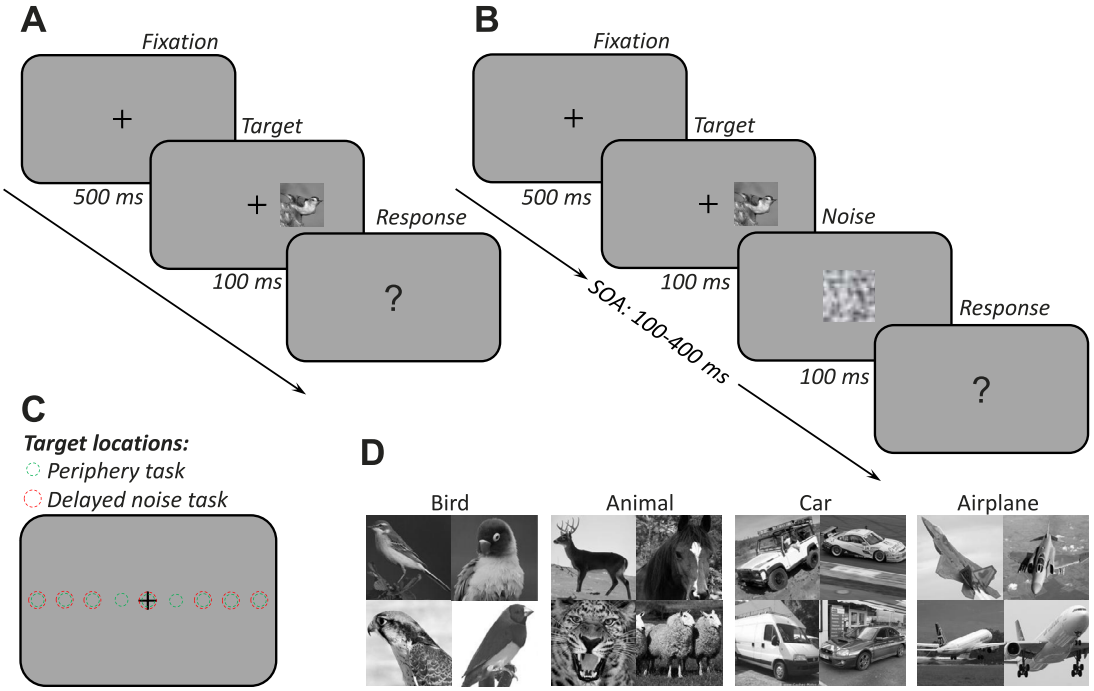

Figure 1. Stimulus presentation paradigm. (A) Peripheral object categorization. Images were presented on nine different locations on screen (C) and subjects were asked to categorize the object images. Each trial started by presenting a gray blank screen with a black fixation point (1 × 1 visual degree) at the center for 500 ms. Then, an object image was randomly presented on one of the nine locations at 0°, ±6°, ±12°, ±18°, or ±24° eccentricity for a 100-ms time window. Finally, a gray blank screen with a black question mark (1 × 1 visual degree) at the center was presented, asking the subject's decision on the category of the presented object. Images covered an area of 5×5 degrees of visual angle. (B) Peripheral object categorization with foveal delayed noise mask. As in (A) but a 100-ms dynamic 1/f noise mask was also presented after the object image with different SOAs. The SOA in each trial was randomly chosen from four values of 100, 200, 400, and 300 ms. Finally, a black question mark (1 × 1 visual degree) on a gray blank screen was presented, asking the subject's decision on the category of the presented object. (D) Sample images from different categories. Note that to analyze the results, we pooled the data on the left and right spatial locations to have five (peripheral task) or four (delayed noise task) eccentricities.

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 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 periphery, as it was shown by disrupting the consistency between the initial coarse peripheral information and detailed foveal information, using foveal noise.

Materials and methods

Image set

We used images from different categories including both animal and nonanimal categories. The animal category contains a variety of images of birds, mammals, and reptiles, while the nonanimal category consists of various object images including houses, cars, kitchen tools, and other man-made objects. Depending on the experiment, the images can be grouped into three categorical levels including superordinate, basic, and subordinate levels. All images were taken from the publicly available ImageNet dataset. 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.[1] Figure 1 demonstrates several sample images from categories. All images were grayscaled and cropped to be of size 150 × 150 pixels in a way that the object covered the largest possible portion of the image.

We had two categorization tasks in the superordinate level: (a) animal/nonanimal; and (b) vehicle/nonvehicle. Accordingly, we had four categorization tasks in the basic level including: (a) bird/nonbird; (b) reptile/nonreptile; (c) car/noncar; and (d) airplane/nonairplane. In every basic-level task, images of the nontarget category were selected from objects of the same superordinate category. For instance, in the bird/

Downloaded from intl.iovs.org on 04/26/2024

nonbird task, the nonbird images were taken from different animals but birds. Finally, considering all combinations, we had eight subordinate tasks including two bird tasks: (a) pigeon/nonpigeon; (b) duck/nonduck; two reptile tasks: (c) lizard/nonlizard; (d) crocodile/noncrocodile; two car tasks: (e) van/nonvan; (f) racer/nonracer; and two airplane tasks: (g) jumbo/nonjumbo; (h) fighter/nonfighter. Here again, the nontarget images in each subordinate task were selected from the same basic-level category as target images. For instance, in pigeon/nonpigeon task, the nonpigeon category contains images of other birds such as ducks, eagles, swans, and so on.

Participants

In total, 100 subjects participated in the experiments (50 females, aged between 20–26). All participants had normal or corrected-to-normal vision. Subjects were students from Faculty of Psychology and Educational Science at the University of Tehran. All subjects voluntarily participated to the experiments and gave their written consent prior to participation. All experimental protocols were approved by the ethical committee of the University of Tehran. All experiments were carried out in accordance with the guidelines of the Declaration of Helsinki.

Psychophysical experiments

We performed two main behavioral experiments—each consisted of several experiments—to study the accuracy and reaction time of human peripheral vision in different levels of object categorization relative to foveal vision. The main difference between the two experiment types was that in the first set of experiments (peripheral object categorization), there was no foveal delayed noise mask image, while in the second we also presented a mask (peripheral object categorization with delayed noise mask). The mask was added to investigate how modulating the foveal representation can influence visual discrimination in periphery (Fan et al., 2016; M. A. Williams et al., 2008). Object images in both tasks were presented in nine different equally spaced locations on the screen between −24° to +24° of eccentricity on a hypothetical horizontal line halving the screen (Figure 1, 0° is at foveal vision). We pooled the data on the symmetric left and right spatial locations to have five eccentricities (0°, 6°, 12°, 18°, and 24°) for the analysis. In each experiment type, there were three levels of categorization in separate blocks: (a) superordinate, (b) basic, and (c) subordinate, and in each level, we defined different categorization tasks (e.g., car/noncar and bird/nonbird in basic level categorization). Images covered an area of 5 × 5 degrees of visual angle.

Peripheral object categorization

In the first set of experiments, we presented object images on different locations on the screen and asked subjects to categorize the object images. In total, subjects performed 14 categorization tasks at three levels of categorization. This included two superordinate, four basic, and eight subordinate tasks (see image set). In each task, we recorded the categorization accuracy and reaction time of 10 human subjects. Note that in each recording session, subjects only performed one of the tasks in one of the levels. In total, we recorded the responses of 70 subjects. Each task comprised of 400 trials (200 images per category) randomly divided into four blocks. Each trial started by presenting a gray blank screen with a black fixation point (1 × 1 visual degree) at the center for 500 ms. Then, an object image was randomly presented in one of the nine locations for a 100-ms time window. Finally, a gray blank screen with a black question mark (1 × 1 visual degree) at the center was presented, asking the subject's decision on the category of the presented object (Figure 1). We controlled to have on average 45 images at each location (i.e., in a range of 44 to 46 images). To familiarize the subjects with the task prior to the main experiment, we had a training block of 50 trials and the auditory feedback indicating the outcome of their decision.

Peripheral object categorization with foveal delayed noise mask

In the second set of experiments, we also presented a 100-ms dynamic (continuously changing) 1/f noise mask after the object image with different stimulus onset asynchronies (SOA). However, we decreased nine peripheral locations to seven and only had one categorization task in each level: (a) animal/nonanimal in the superordinate level, (b) bird/nonbird in the basic level, and (c) pigeon/nonpigeon in the subordinate level. Each task contained 560 trials (280 per category). Trials started with a 500-ms blank screen with a black fixation point (1 × 1 visual degree), followed by an object image randomly presented at one of the seven locations for a 100-ms time window. Then, a dynamic 1/f noise mask of size 7 × 7 visual degrees was presented at the center of the screen for 100 ms. The SOA in each trial was randomly chosen from four values of 100, 200, 300, and 400 ms. Finally, a black question mark (1 × 1 visual degree) on a gray blank screen was shown until the subjects reported their decision on the object category. We controlled to have 140 images (70 per category) at each SOA and 20 images (10 per category)

Downloaded from intl.iovs.org on 04/26/2024

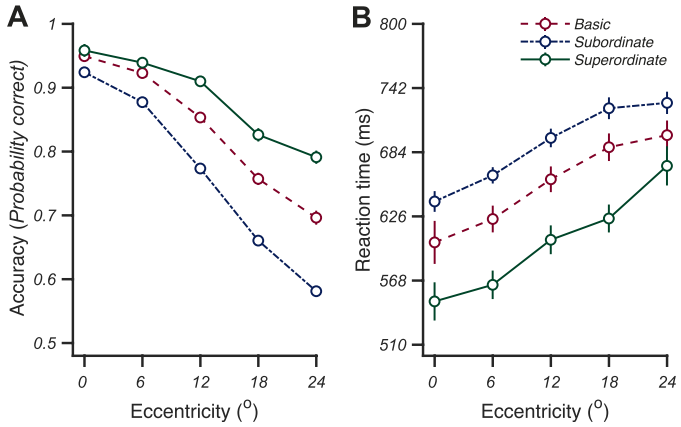

Figure 2. Accuracy and reaction time in categorizing images at different levels (i.e., superordinate, basic, and subordinate) presented at central and peripheral visual fields. (A) Categorization accuracy at superordinate (green), basic (red), and subordinate (blue) levels presented at different eccentricities. (B) The median reaction time in categorizing images at different levels and eccentricities. Error bars are standard error of means.

at each location and SOA. Subjects became familiar with the task during a training block of 50 trials and feedback indicating the outcome of their decision. In each task, we recorded the categorization accuracy and reaction time of 10 human subjects. In total, we recorded the responses of 30 subjects.

Setup and eye tracking

During the experiment, subjects sat on a chair and had a 50-cm distance from a 23-in. monitor (Tobii TX300 eye tracker device) with 1920 × 1080 resolution. We asked subjects to keep their gaze on the fixation point and the eye position was monitored using the eye tracker while any saccades were detected by the camera. The eye tracker device had a sampling frequency of 300 Hz. 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 Experiments 1 and 2, respectively).

Results

In the first set of experiments, we measured the accuracy and reaction time of human subjects in categorization of object images in different categorization levels randomly presented at any of the five different eccentricities from foveal (0°) to peripheral field (24°). Then, in a complementary set of experiments, we studied how adding a dynamic foveal noise mask influences peripheral object categorization in different abstraction levels. In the paper, accuracy refers to the ratio of correct responses and reaction time is the median reaction time of correct responses. We used ANOVA with category level and eccentricity as factors for the statistical analysis and Bonferroni multiple comparison correction for p values.

Object categorization in the visual periphery: The superordinate advantage holds across the visual field

Here, we examined whether human subjects are able to categorize rapidly presented images across different visual eccentricities and if the well-known phenomenon of "superordinate advantage" holds at different eccentricities. We found that the average categorization accuracy was slightly, but significantly, higher in superordinate and basic than subordinate levels at central (foveal) visual field (Figure 2A; $p < 0.01$, $F = 7.1$), confirming previous results on "superordinate advantage" at the central visual field (Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Wu, Crouzet, Thorpe, & Fabre-Thorpe, 2015). The accuracy at central vision was above 90% in all three categorization levels. However, the accuracy significantly dropped as a function of eccentricity ($F = 394.27$, $p < 0.001$). This accuracy drop was highest in the subordinate level and lowest in superordinate (category level main effect: $F = 245.33$, $p < 0.001$). In particular, the accuracy in superordinate level was above 80% even at the highest eccentricity (24°), while this decreased to 60% in subordinate level (Figure 2A). As expected, we found a significant interaction between the categorization level and eccentricity ($F = 15.36$, $p < 0.001$). These together indicate that human peripheral vision is easily able to categorize object images in higher abstraction levels (e.g., animal), but is less accurate in categorizing the same image in a finer abstraction level (e.g., pigeon). Put simply, it is easier to spot an animal in the visual periphery but it is more difficult to say what the animal and its species are.

We also calculated the median reaction time for each categorization level and eccentricity. There was a significant main effect of categorization level ($F = 49.54$, $p < 0.001$). Also, subjects' reaction times at all levels significantly increased as a function of eccentricity ($F = 26.23$, $p < 0.001$). Overall, observers were fastest in the categorization of superordinate level images and slowest in the subordinate level in all tested eccentricities (Figure 2B). The superordinate categorization at the center was completed 550 ± 12 ms (M ±

Downloaded from intl.iovs.org on 04/26/2024

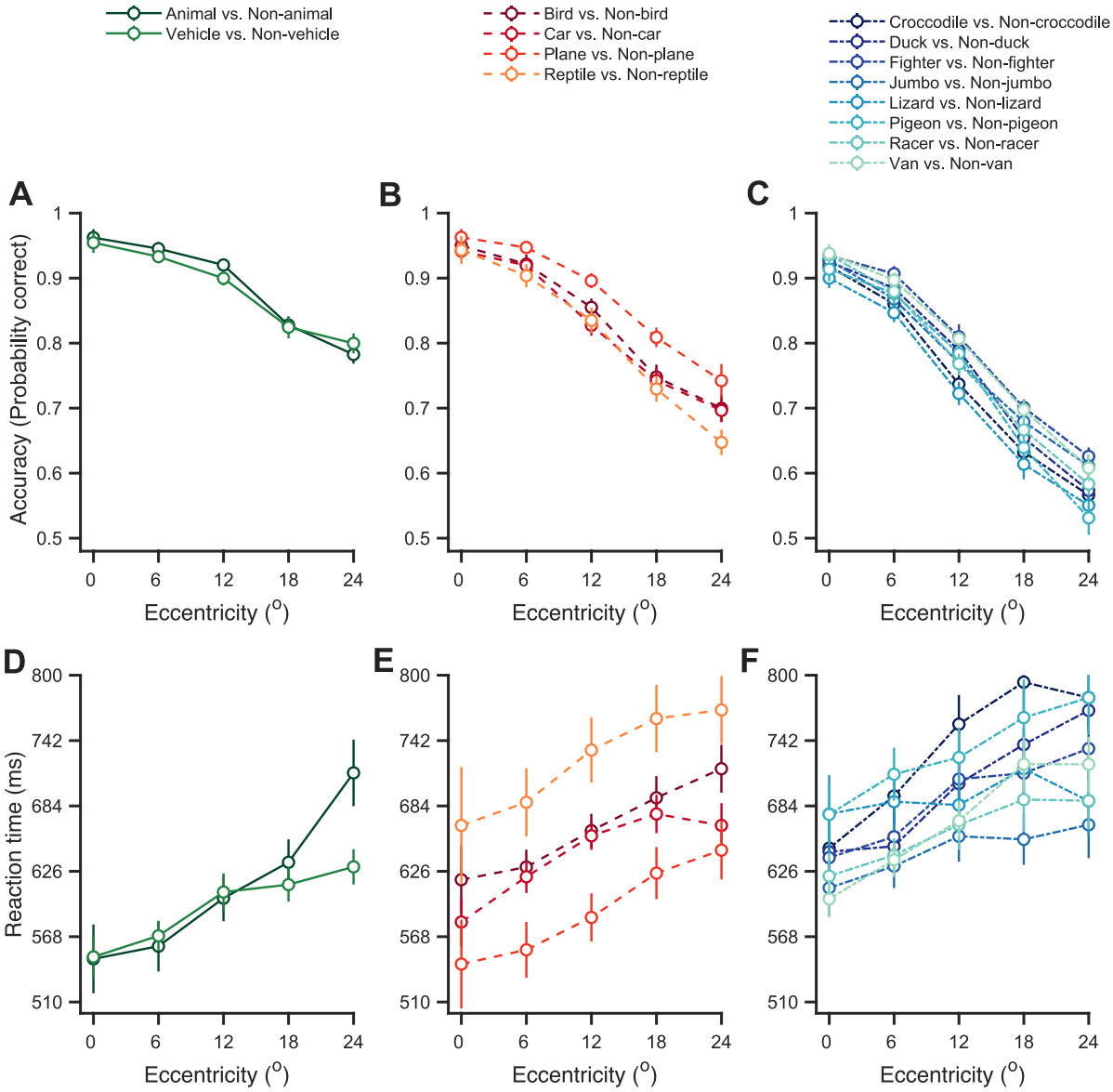

Figure 3. Accuracy and reaction time in categorizing images at different levels (i.e., superordinate, basic, and subordinate) presented at central and peripheral visual fields for individual tasks. Average accuracy in categorizing images from different classes (tasks) in superordinate (A), basic (B), and subordinate (C) levels. Images presented at different eccentricities. Each color refers to a particular task. (D–F) Median reaction time in different tasks within each level. Error bars are standard error of means.

SEM) after image onset while the basic and subordinate levels needed 602 ± 15 ms and 640 ± 7 ms, respectively. Similar differences in reaction time were observed at other eccentricities (Figure 2B). There was no interaction between the categorization level and eccentricity ($F = 0.45$, $p = 0.8$).

Note that the accuracy and reaction time in the above analysis were collapsed across different tasks (e.g., animal/nonanimal and vehicle/nonvehicle in superordinate level). We also calculated the accuracy and reaction time for every individual task within each categorization level. We did not observe any significant difference between the accuracies in two superordinate tasks at any eccentricity (Figure 3A; $p > 0.05$). However, there was a significant difference between the reaction times of animal/nonanimal and vehicle/nonvehicle tasks only at 24° eccentricity (Figure 3D; $p < 0.05$). For the basic-level tasks, we observed differences in the accuracies and reaction times of vehicle and animal tasks. For example, subjects were 142 ms faster and 5.9% more accurate in categorizing plane images as compared to reptile images (Figure 3B and E; $p < 0.001$ for both reaction time and accuracy).

Downloaded from intl.iovs.org on 04/26/2024

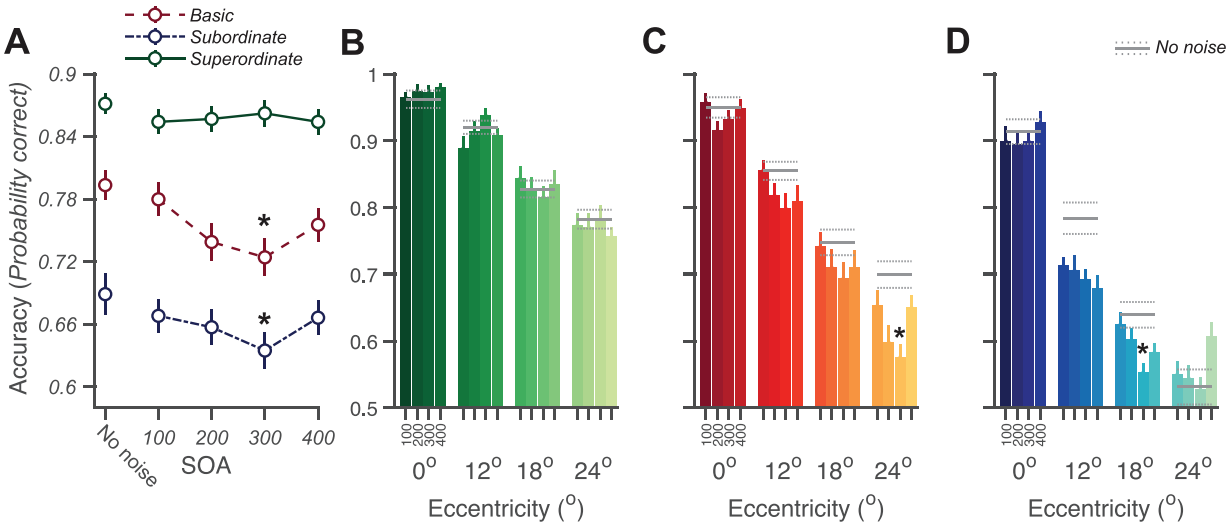

Figure 4. Average accuracy in different levels and under four SOA conditions. (A) Average accuracy in superordinate (green), basic (red), and subordinate (blue) levels as a function of SOAs. Accuracies averaged over all eccentricities. Error bars are standard error of means. Average accuracy presented for every SOA and eccentricity in superordinate (B), basic (C), and subordinate (D) levels. The gray horizontal lines refer to accuracy in no-noise condition.

Delayed foveal noise impairs peripheral object categorization in basic and subordinate levels

Although a few studies have demonstrated that delayed foveally presented noise disrupts peripheral object discrimination for artificial images, it is unclear how these findings hold when subjects categorize natural images in different levels. Here, we investigated whether and with what delay the foveally presented noise could disrupt discrimination of peripherally presented natural images. Also, we examined where in the visual periphery this disruption could have been observed. To do this, we presented the noise image with four different SOAs relative to the object image.

We found a significant main effect of SOA on accuracy ($F = 3.1$, $p < 0.01$). Our analysis showed that delayed foveally presented noise can significantly affect subjects' accuracy in categorizing peripherally presented natural images in basic and subordinate levels but not in the superordinate level (Figure 4). In particular, subjects' accuracy in basic level categorization significantly decreased by 6.9% at 300 ms SOA relative to the no-noise condition (Figure 4A; $p < 0.001$) and by 5.41% in subordinate level (Figure 4A; $p < 0.001$). This is consistent with previous studies where the highest decline in categorization accuracy of artificial images was observed for SOA = 250 ms (Fan et al., 2016).

We examined this at four different eccentricities (i.e., 0°, 12°, 18°, and 24°). Our analysis revealed that delayed foveal noise did not affect subjects' accuracy in categorizing peripherally presented objects in the superordinate level at any tested eccentricities compared to the accuracy in the no-noise condition (Figure 4B; $p > 0.05$ in all eccentricities). However, in the basic level categorization, we observed a significant decrease of 12.2% in accuracy at 300 ms SOA and 24° eccentricity relative to the no-noise condition (Figure 4C; $p < 0.001$). 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; $p = 0.37$ for 12°, $p = 0.87$ for 18° eccentricities). 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. We found a significant decrease of 8.6% in accuracy at 300 ms SOA relative to no noise in the subordinate level categorization only at 18° eccentricity (Figure 4D; $p < 0.01$). The lack of accuracy difference at 24° may be due to a floor effect.

We also explored the changes in reaction time as a function of SOA at different eccentricities. Although subjects' reaction times 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). In particular, subjects' reaction times significantly increased as a function of SOA even in the superordinate level (Figure 5A; $F = 97.26$, $p < 0.001$), and we did not observe any nonmonotonic relationship between SOA and reaction time, meaning that by increasing the SOA of the foveal noise, the reaction time always increased but accuracy drop was dependent on task and SOA. As expected, the reaction times of the basic level is higher than those of the subordinate level for all SOAs. However, unexpectedly, the reaction times of the superordinate level were higher than the basic level and almost similar to the

Downloaded from intl.iovs.org on 04/26/2024

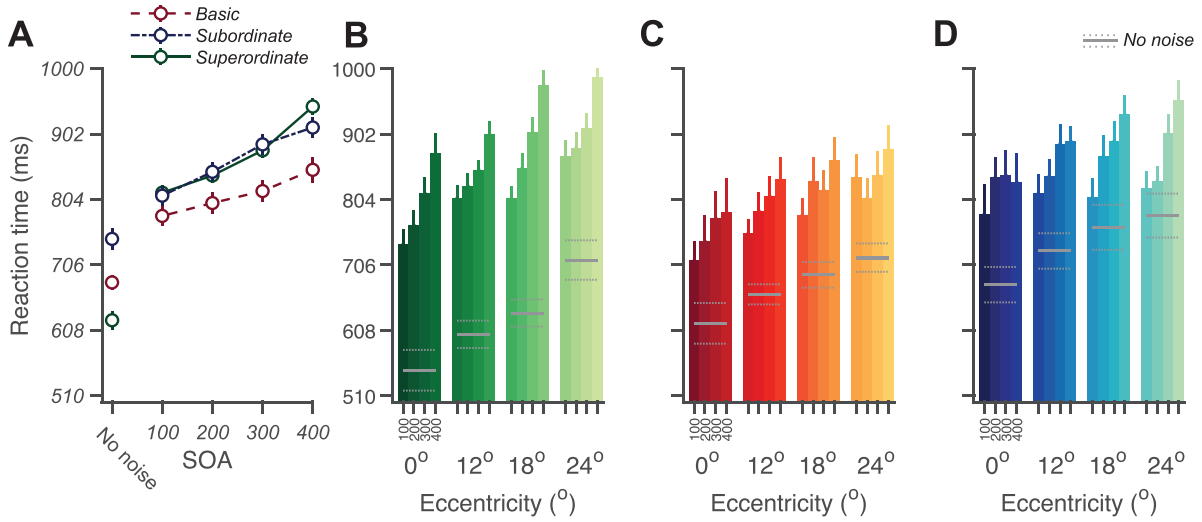

Figure 5. Median reaction times in different categorization levels and under four SOA conditions. (A) Median reaction times in superordinate (green), basic (red), and subordinate (blue) levels as a function of SOA. Reaction times collapsed across all eccentricities. Error bars are standard error of means. Median reaction times presented for every SOA and eccentricity in superordinate (B), basic (C), and subordinate (D) levels. The gray horizontal lines refer to reaction times in the no-noise condition.

subordinate level. In other words, reaction times in the 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 the basic level with respect to the superordinate level (Figure 4) and shorter reaction times for the superordinate level in the no-noise condition (Figure 5).

Discussion

Studies on peripheral vision have been mostly focused on particular visual tasks such as categorization of global scene properties (Boucart, Moroni, Thibaut, et al., 2013; Ehinger & Rosenholtz, 2016; Wang & Cottrell, 2017), visual attention (Intriligator & Cavanagh, 2001; Rosenholtz, 2016; Rosenholtz, Huang, & Ehinger, 2012), and eye movement planning (Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Wiecek, Pasquale, Fiser, Dakin, & Bex, 2012). However, some recent studies have shown that peripheral vision is able to perform rapid scene categorization tasks (Boucart, Moroni, Thibaut, et al., 2013; Larson & Loschky, 2009) and to solve complex object categorization tasks including natural object discrimination (Boucart et al., 2010; Boucart, Moroni, Szaffarczyk, & Tran, 2013; Ehinger & Rosenholtz, 2016; S. J. Thorpe et al., 2001) and face recognition (Crouzet, Kirchner, & Thorpe, 2010; Hershler, Golan, Bentin, & Hochstein, 2010) in the very far periphery (Boucart et al., 2016). Moreover, recent neuroimaging findings in humans have suggested that there is a feedback system in the brain that somehow sends the information from visual periphery to cortical regions retinotopic to fovea (Chambers et al., 2013; M. A. Williams et al., 2008), and psychophysical manipulation of this feedback information can disrupt humans' categorization performance in visual periphery (Fan et al., 2016; Weldon et al., 2016; Yu & Shim, 2016). The time course of peripheral feedback information is found to be flexible and depends on the task demand (Fan et al., 2016).

In this paper, we designed several psychophysical experiments to first measure subjects' peripheral visual discrimination performance at different eccentricities and then investigated how modulating the foveal representation could affect subjects' peripheral categorization performance. Subjects categorized a large pool of natural object images in three coarse-to-fine levels of superordinate, basic, and subordinate. Our results showed that although subjects' accuracy dropped as the object presented further from the fixation area, they could easily categorize images in the superordinate level (accuracy of ~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°).

The performance drop in the periphery is likely due to an increase in crowding and decrease in spatial acuity, which respectively jumble and reduce the spatial details necessary for categorization. It is important to note that self-crowding can occur with a single object

Downloaded from intl.iovs.org on 04/26/2024

(Martelli, Majaj, & Pelli, 2005), which may be at play with our natural image stimuli. A recent study has shown that the impact of crowding on peripheral object recognition in natural real-world scenes can be reduced by the presence of more contextual information (Wijntjes & Rosenholtz, 2018), meaning that subjects' recognition performance impairs when presented with objects segmented from their background compared to viewing the objects in their real-world congruent background. Although in our experiments all objects were presented in their real-world congruent background and we cropped the image in a way that the target object covers a large portion of the image, crowding still remains one of the possible reasons for the accuracy drop in periphery. Albeit crowding and low visual acuity of the periphery are the main factors limiting peripheral object recognition at finer abstraction levels, the contribution of each factor needs to be more carefully investigated in future studies. One way to address this can be performing similar experiments to ours but using band-pass filtered natural (cluttered) images in different spatial frequencies (Ashtiani, Kheradpisheh, Masquelier, & Ganjtabesh, 2017).

In the second set of psychophysical experiments, we measured subjects' performance in categorization of peripherally presented object images at different eccentricities while modulating their foveal representation. Using this paradigm, we tested the spatial and temporal extent of the feedback information from the peripheral to foveal vision in a set of coarse-to-fine categorization experiments. To this end, we presented a foveal noise mask with four different SOAs (100 to 400 ms) relative to the peripherally presented stimuli. We showed that the foveal noise does not affect the categorization accuracy of the coarse discrimination task (superordinate level) at any eccentricity and SOA. However, the categorization accuracy of the basic and subordinate levels significantly dropped as a result of foveally presented noise when presented at 300 ms SOA (300 ms after stimulus onset). 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. Our results also suggest that the effect of foveal noise is dependent on the task demand and only occurs for categorization tasks that require detailed spatial information. Moreover, it seems that this accuracy impairment is eccentricity dependent—for example, in basic level categorization, the significant accuracy drop was observed at 24° of eccentricity, while it happened at 18° of eccentricity for the subordinate level.

Our findings are consistent with the "reverse hierarchy theory" (Ahissar & Hochstein, 2004) and the "top-down facilitation theory" (Bar et al., 2006). Based on the reverse hierarchy theory, the initial feedforward processing through the visual cortex provides a gist about the input image to the high-level areas (e.g., IT cortex), then the fine-detailed information is queried through feedback interactions with lower areas. In the top-down facilitation theory (Bar et al., 2006), an immediate high-level guess about the presented object is first made based on the low spatial frequency information that is quickly transferred to the prefrontal cortex. Subsequently, top-down feedback connections modulate the processing of late feedforward high spatial frequency information in the IT cortex that eventually facilitates object categorization in finer levels.

We found that coarse information available to the visual periphery was sufficient for superordinate categorization and the accuracy was not affected by the foveal noise, while in basic and subordinate levels, the accuracy significantly dropped by a foveal noise at 300 ms SOA. Although this result should be interpreted with caution, it nevertheless suggests that this time window (~300 ms) may be important for resolving peripheral representations, perhaps prior to a goal-directed saccade that would eventually foveate the target object (about 150 ms for the initial object processing (S. Thorpe, Fize, & Marlot, 1996) and 150–200 ms for making the saccade (Fischer, 1987). Hence, this accuracy drop could probably be due to the inconsistency among the peripheral top-down feedback (of the target object) and the later fine-detailed feedforward foveal information (of the noise mask). When these two pieces of incongruent information simultaneously arrive at the foveal cortex, the top-down peripheral object representation is impaired by the noisy foveal information, which consequently decreases the categorization accuracy in basic and subordinate levels. In other words, our results suggest that an overall guess about the object category can be made by peripheral vision, but to do a fine-level categorization a saccade toward the object is needed. This might be a built-in mechanism in our visual system that a prior peripheral feedback primes the foveal cortex to facilitate the fine categorization of the upcoming foveal object. This is consistent with the well-established finding of parafoveal preview (Henderson, 1992; Henderson et al., 1989), where it is shown that by attending to a peripheral target before making a saccade, the parafoveally processed visual features of that target can facilitate recognition of the target after the eyes land on it. Moreover, physiological measurements in several studies have shown presaccadic activations in several brain areas where neurons become responsive to the visual stimuli that will be

Downloaded from intl.iovs.org on 04/26/2024

brought into their receptive field by an imminent saccade (Cavanagh, Hunt, Afraz, & Rolfs, 2010; Melcher, 2007; Walker, Fitzgibbon, & Goldberg, 1995). Some recent studies have also shown that presaccadic 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, first, in our experiments, subjects were not asked to do any specific foveal task apart from fixation. Second, 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. Third, 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.

Taken together, our findings suggest that the foveal feedback system plays an important role in coordination between peripheral and foveal visual perception, and can contribute to finer object categorization in periphery. Future studies should reveal how brain areas get involved in these processes, what kind of visual information is fed back to the foveal cortex, and the temporal dynamic of these feedback interactions.

Keywords: visual periphery, object recognition, foveal feedback

Acknowledgments

We would like to thank Javad Hatami for providing us the set up in the cognitive psychology laboratory at Faculty of Psychology and Education in University of Tehran. We also thank Yasamin Mokri for proofreading and editing the manuscript. This work has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n.323711 (M4 project).

*FR and SRK contributed equally to this work.

Commercial relationships: none.
Corresponding author: Masoud Ghodrati.
Email: ghodrati.masoud@gmail.com.
Address: Neuroscience Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia.

Footnote

[1] Images are available at https://github.com/Masoud-Ghodrati/vis_peri

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Ashtiani, M. N., Kheradpisheh, S. R., Masquelier, T., & Ganjtabesh, M. (2017). Object categorization in finer levels relies more on higher spatial frequencies and takes longer. *Frontiers in Psychology*, *8*, 1261.

Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *Journal of the Optical Society of America A*, *5*(12), 2210–2219.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, USA*, *103*(2), 449–454.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision*, *16*(2):10, 1–13, https://doi.org/10.1167/16.2.10. [PubMed] [Article]

Boucart, M., Moroni, C., Szaffarczyk, S., & Tran, T. H. C. (2013). Implicit processing of scene context in macular degeneration. *Investigative Ophthalmology & Visual Science*, *54*(3), 1950–1957.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35–42.

Boucart, M., Naili, F., Despretz, P., Defoort-Dhellemmes, S., & Fabre-Thorpe, M. (2010). Implicit and explicit object recognition at very large visual eccentricities: No improvement after loss of central vision. *Visual Cognition*, *18*(6), 839–858.

Downloaded from intl.iovs.org on 04/26/2024

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153.

Chambers, C. D., Allen, C. P., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex*, *49*(1), 327–335.

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2013). Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex*, *24*(11), 2899–2907.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4):16, 1–17, https://doi.org/10.1167/10.4.16. [PubMed] [Article]

Duhamel, J. R., Colby, C., & Goldberg, M. (1992, January 3). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90–92.

Ehinger, K. A., & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, *16*(2): 13, 1–19, https://doi.org/10.1167/16.2.13. [PubMed] [Article]

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, *113*(41), 11627–11632.

Fischer, B. (1987). The preparation of visually guided saccades. In *Reviews of physiology, biochemistry and pharmacology* (*vol. 106*, pp. 1–35). Berlin, Heidelberg: Springer.

Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science*, *2*, 377–396.

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662.

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., & Chun, M. M. (2010). Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. *Journal of Neuroscience*, *30*(31), 10493–10506.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998, January 29). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484.

Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, *33*(7), 2927–2933.

Harrison, W. J., Retell, J. D., Remington, R. W., & Mattingley, J. B. (2013). Visual crowding at a distance during predictive remapping. *Current Biology*, *23*(9), 793–798.

Henderson, J. M. (1992). Identifying objects across saccades: Effects of extrafoveal preview and flanker object context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 521.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics*, *45*(3), 196–208.

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision*, *10*(10):21, 1–14, https://doi.org/10.1167/10.10.21. [PubMed] [Article]

Ikeda, M., & Takeuchi, T. (1975). Influence of foveal load on the functional visual field. *Perception & Psychophysics*, *18*(4), 255–260.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216.

Jebara, N., Pins, D., Despretz, P., & Boucart, M. (2009). Face or building superiority in peripheral vision reversed by task requirements. *Advances in Cognitive Psychology*, *5*, 42.

Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The neural bases of the semantic interference of spatial frequency-based information in scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394–2405.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10):6, 1–16, https://doi.org/10.1167/9.10.6. [PubMed] [Article]

Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception from central to peripheral vision. *Journal of Vision*, *17*(1):6, 1–5, https://doi.org/10.1167/17.1.6. [PubMed] [Article]

Ludwig, C. J., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences*, *111*(2), E291–E299.

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One*, *4*(6), e5927.

Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D.

Downloaded from intl.iovs.org on 04/26/2024

(2001). Identification of facial images in peripheral vision. *Vision Research*, *41*(5), 599–610.

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*(1):6, 58–70, https://doi.org/10.1167/5.1.6. [PubMed] [Article]

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, *10*(7), 903.

Ringer, R. V., Throneburg, Z., Johnson, A. P., Kramer, A. F., & Loschky, L. C. (2016). Impairing the useful field of view in natural scenes: Tunnel vision versus general interference. *Journal of Vision*, *16*(2):7, 1–25, https://doi.org/10.1167/16.2.7. [PubMed] [Article]

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), 252.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, *2*, 437–457.

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, *3*, 13.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4):14, 1–17, https://doi.org/10.1167/12.4.14. [PubMed] [Article]

Sommer, M. A., & Wurtz, R. H. (2006, November 16). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*, *444*(7117), 374–377.

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5):13, 1–82, https://doi.org/10.1167/11.5.13. [PubMed] [Article]

Thorpe, S., Fize, D., & Marlot, C. (1996, June 6). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & BuÈlthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, *14*(5), 869–876.

Walker, M. F., Fitzgibbon, E. J., & Goldberg, M. E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, *73*(5), 1988–2003.

Wang, P., & Cottrell, G. W. (2017). Central and peripheral vision for scene recognition: A neurocomputational modeling exploration. *Journal of Vision*, *17*(4):9, 1–22, https://doi.org/10.1167/17.4.9. [PubMed] [Article]

Weldon, K. B., Rich, A. N., Woolgar, A., & Williams, M. A. (2016). Disruption of foveal space impairs discrimination of peripheral objects. *Frontiers in Psychology*, *7*, 699.

Wiecek, E. W., Pasquale, L. R., Fiser, J., Dakin, S., & Bex, P. J. (2012). Effects of peripheral visual field loss on eye movements during visual search. *Frontiers in Psychology*, *3*, 472.

Wijntjes, M. W., & Rosenholtz, R. (2018). Context mitigates crowding: Peripheral object recognition in real-world images. *Cognition*, *180*, 158–164.

Williams, L. J. (1985). Tunnel vision induced by a foveal load manipulation. *Human Factors*, *27*(2), 221–227.

Williams, M. A., Baker, C. I., De Beeck, H. P. O., Shim, W. M., Dang, S., Triantafyllou, C., . . . Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439.

Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience*, *8*, 103.

Wright, M., & Johnston, A. (1983). Spatiotemporal contrast sensitivity and visual field locus. *Vision Research*, *23*(10), 983–989.

Wu, C.-T., Crouzet, S. M., Thorpe, S. J., & Fabre-Thorpe, M. (2015). At 120 msec you can spot the animal but you don't yet know it's a dog. *Journal of Cognitive Neuroscience*, *27*(1), 141–149.

Yu, Q., & Shim, W. M. (2016). Modulating foveal representation can influence visual discrimination in the periphery. *Journal of Vision*, *16*(3):15, 1–12, https://doi.org/10.1167/16.3.15. [PubMed] [Article]

Downloaded from intl.iovs.org on 04/26/2024