Visual crowding in driving

Ye Xia Department of Psychology, University of California, Berkeley, CA, USA

Mauro Manassi Department of Psychology, University of California, Berkeley, CA, USA; School of Psychology, University of Aberdeen, Kings College, Aberdeen, Scotland, UK

Ken Nakayama Department of Psychology, University of California, Berkeley, CA, USA

Karl Zipser Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

David Whitney Department of Psychology, University of California, Berkeley, CA, USA; Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA; Vision Science Group, University of California, Berkeley, CA, USA

Visual crowding—the deleterious influence of nearby objects on object recognition—is considered to be a major bottleneck for object recognition in cluttered environments. Although crowding has been studied for decades with static and artificial stimuli, it is still unclear how crowding operates when viewing natural dynamic scenes in real-life situations. For example, driving is a frequent and potentially fatal real-life situation where crowding may play a critical role. In order to investigate the role of crowding in this kind of situation, we presented observers with naturalistic driving videos and recorded their eye movements while they performed a simulated driving task. We found that the saccade localization on pedestrians was impacted by visual clutter, in a manner consistent with the diagnostic criteria of crowding (Bouma's rule of thumb, flanker similarity tuning, and the radial-tangential anisotropy). In order to further confirm that altered saccadic localization is a behavioral consequence of crowding, we also showed that crowding occurs in the recognition of cluttered pedestrians in a more conventional crowding paradigm. We asked participants to discriminate the gender of pedestrians in static video frames and found that the altered saccadic localization correlated with the degree of crowding of the saccade targets. Taken together, our results provide strong evidence that crowding impacts both recognition and goal-directed actions in natural driving situations.

Introduction

We live in a constantly cluttered visual world: from letters in text, and products on the shelves of supermarkets, to the crowds of cars and pedestrians on busy streets. The natural crowdedness of our visual input confronts us with a fundamental limitation of our visual system, known as visual crowding: objects that can be easily identified in isolation seem jumbled and indistinct within clutter (Levi, 2008; Pelli & Tillman, 2008; Whitney & Levi, 2011). Visual crowding operates over a wide part of our visual field, particularly in peripheral vision (Malania, Herzog, & Westheimer, 2007; Manassi, Sayim, & Herzog, 2012; Sayim, Westheimer, & Herzog, 2010; Toet & Levi, 1992), and it is considered to be a major bottleneck in recognizing objects in clutter (Levi, 2008; Manassi & Whitney, 2018; Pelli & Tillman, 2008; Strasburger, Rentschler, & Juttner, 2011; Whitney & Levi, 2011).

Citation: Xia, Y., Manassi, M., Nakayama, K., Zipser, K., & Whitney, D. (2020). Visual crowding in driving. *Journal of Vision*, *20*(6):1, 1–17, https://doi.org/10.1167/jov.20.6.1.

https://doi.org/10.1167/jov.20.6.1 Received September 9, 2019; published June 3, 2020 ISSN 1534-7362

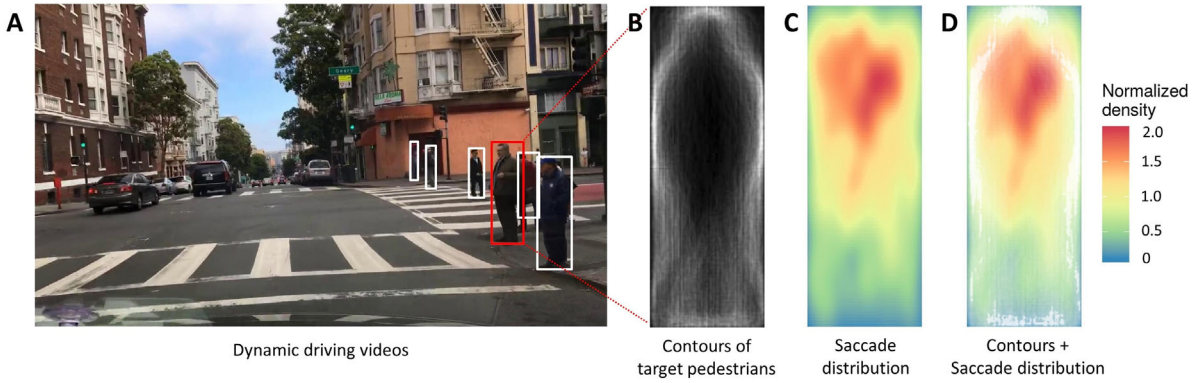

Figure 1. Participants watched crowd-sourced driving videos in a simulated driving environment while we recorded eye movements. Pedestrians were detected by a state-of-the-art object detection algorithm, Mask R-CNN (He et al., 2017). The saccades that landed on pedestrians were then identified. (A) One example frame of the crowd-sourced driving videos and the bounding boxes of the detected pedestrian given by Mask R-CNN. The red bounding box highlights the pedestrian that was targeted by one of the participants' saccades. The white bounding boxes show the other detected pedestrians. (B) Overlaid contours of the pedestrians on which participants' saccades landed, such as the one highlighted in red in Panel A. (C) Distribution of the landing points of the pedestrian-targeted saccades within the bounding box of a pedestrian. (D) The overlay of Panels A and B.

Given its ubiquity and significance, the impact of crowding on object recognition has been studied for decades (Flom, Heath, & Takahashi, 1963; Levi, Hariharan, & Klein, 2002; Levi, Klein, & Hariharan, 2002; Pelli, Palomares, & Majaj, 2004; Strasburger et al., 2011; Westheimer & Hauske, 1975). However, the vast majority of studies in the field have been mainly restricted to experiments in psychophysical laboratories. Stimuli typically used are usually static, artificial, and relatively simple, such as oriented gratings, shapes, letters, symbols, and even faces; only a few experiments have studied the impact of crowding in static natural scenes or of natural textures (Gong, Xuan, Smart, & Olzak, 2018; Wallis & Bex, 2012). Furthermore, most crowding studies rely on participants' explicit responses made during psychophysical tasks (e.g. pressing buttons or clicking the mouse to explicitly report a perceptual decision). These response methods, along with the laboratory settings and relatively artificial stimuli, raise an important question: how relevant is crowding to daily life? Specifically, how does crowding operate when viewing natural dynamic scenes in real-life situations?

Driving is a real-world behavior in which moment-to-moment decisions can have life or death consequences. We ask the question: does visual crowding play a critical role in driving? Crowding may occur in driving because objects, obstacles, and pedestrians frequently appear in clutter in the visual periphery (see Figure 1A for an example). However, on the other hand, many types of information (e.g. object configurations, facial expressions, textures, and even scene gist), are known to get through the bottleneck of crowding (for a review see Manassi & Whitney, 2018). Some of this information could guide behavior, at least in principle. Therefore, whether crowding limits behavioral performance in the context of driving is a crucial and open question.

A major challenge when studying crowding in driving (and natural scenes in general) is the availability and analysis of realistic stimuli. Recently, however, in the field of computer vision and autonomous driving, diverse large-scale driving video datasets have been created (Cordts et al., 2016; Huang et al., 2018; Maddern, Pascoe, Linegar, & Newman, 2017; Xu, Gao, Yu, & Darrell, 2017; Yu et al., 2018), and powerful object detection algorithms using deep learning have been generated (Chen et al., 2019; He, Gkioxari, Dollar, & Girshick, 2017; Liu, Qi, Qin, Shi, & Jia, 2018). Here, we adopted driving videos from one of the new datasets, and we used a state-of-the-art object detection algorithm to analyze object-level information in the videos.

In the present study, we first investigated the impact of visual clutter in dynamic driving scenes on a fundamental kind of behavior in driving: eye movements. The question is whether crowding alters saccadic localization of peripheral targets in dynamic natural scenes. Second, we further tested whether visual crowding occurs in the recognition of peripheral flanked objects (pedestrians), using a more conventional psychophysical paradigm. To foreshadow our results, we found that both saccadic localization and pedestrians recognition were impacted in manners that were consistent with the well-established diagnostic criteria of crowding (Whitney & Levi, 2011): Bouma's rule-of-thumb (Bouma, 1970; Pelli & Tillman, 2008), target-flanker similarity tuning (Kooi, Toet, Tripathy, & Levi, 1994; see Levi, 2008 for review), and radial-tangential anisotropy (Toet & Levi, 1992). Importantly,

altered saccadic localization was associated with the degree of crowding of the saccade targets. These results provide strong evidence that crowding impacts both recognition and goal-directed actions in natural driving situations, with important implications for driving safety.

Experiment 1

In Experiment 1, we recorded eye movements while observers watched natural driving videos, and we analyzed the saccades that landed on pedestrians (pedestrian-targeted saccades). We tested whether corrections in saccadic localization occurred in a manner consistent with the diagnostic criteria of crowding (Bouma's rule of thumb, flanker similarity tuning, and radial-tangential anisotropy). A positive result would suggest that crowding of pedestrians could potentially exist in driving and be associated with altered saccadic localization.

Methods

Participants

Eight naïve participants (6 women) participated in the experiment for course credits. All of the participants had driven for more than one year and had normal or corrected normal vision. All experimental procedures were approved by and conducted in accordance with the guidelines and regulations of the University of California, Berkeley Institutional Review Board. Participants were affiliates of University of California, Berkeley and provided informed consent in accordance with the Institutional Review Board guidelines of the University of California, Berkeley.

Stimuli and equipment

We used 519 videos from Berkeley DeepDrive Attention (BDD-A) dataset (Xia et al., 2018). BDD-A contains crowd-sourced driving videos recorded by vehicle-mounted dashboard cameras in cities under various weather and lighting conditions (Xia et al., 2018). The videos are mostly 10 seconds long and contain diverse driving activities (e.g. lane following, turning, switching lanes, and braking).

Stimuli were displayed in full screen on a CRT monitor (display area size 34 cm × 23 cm). Display resolution was set to 1,024 × 768 and refresh rate to 60 Hz. Participants viewed the stimuli binocularly in a darkened experimental booth, and head position was stabilized with a chinrest at a viewing distance of 57 cm. At this distance, 30 pixels subtended approximately 1° of visual angle.-

Eye tracking

Eye movements were recorded at 1,000 Hz monocularly with an EyeLink 1000 desktop mounted infrared eye tracker (SR Research Ltd., Mississauga, Ontario, Canada) used in conjunction with the Eyelink Toolbox scripts for Matlab. Participants were calibrated with a standard nine-point calibration procedure before completing each block (average error < 0.5°). Saccades were parsed out by the EyeLink Online Parser with the default high-sensitivity configuration (velocity threshold = 22°/s, acceleration threshold = 4,000°/s^2, and motion threshold = 0°).

Procedure

Participants performed a driver instructor task adopted from Xia et al. (2018). They watched the driving videos after they were informed that they were driving instructors sitting in the copilot seat. They were asked to press the space key whenever they felt it necessary to correct or warn the student driver of potential dangers. Their eye movements during the task were recorded. It was previously shown that the gaze maps collected by this method are considered as reasonable driver attention maps by independent human viewers (Xia et al., 2018), and can be used to improve autonomous driving models (Xia et al., 2020). Therefore, we theorized that the driving instructor task can simulate an engaging driving environment for the participants while allowing them to make natural eye movements throughout the scene. The conclusions drawn from these eye movements could then be presumably generalized to drivers' eye movements during actual driving.

Each participant watched 200 driving videos in random order while performing the driving instructor task. Before each driving video, a yellow bullseye was displayed at the center of the screen on top of a uniform gray background. Participants were asked to gaze at the yellow bullseye and press the enter key to start the next video. The yellow bullseye disappeared when the video started. The whole experiment was about one hour long.

Results and discussion

Saccades that landed on pedestrians

To identify the saccades that landed on pedestrians (pedestrian-targeted saccades), we extracted the video frame at the ending point of each saccade. We then applied an object detection model to those extracted

video frames using a state-of-the-art deep learning object detection algorithm, Mask R-CNN (He et al., 2017), and acquired bounding boxes around the detected objects (Figure 1A). Note that it would take more than one thousand hours to manually label those objects (600K in total) and to register their bounding boxes. We identified the pedestrian-targeted saccades by looking for saccades with landing points within the bounding box of a detected pedestrian. Saccades with landing points beyond 15° away from the starting points were considered outliers and excluded (1.8% of total). In total, 2,067 pedestrian-targeted saccades were identified and recorded.

We extracted the contours of the target pedestrians (the ones on which saccades landed), scaled them to the median height-to-width ratio of all the target pedestrians (the median height-to-width ratio = 2.8), and overlaid them together. The overlaid contours showed the shape of a standing/walking pedestrian (Figure 1B). The contour of each pedestrian was extracted in the following two steps. First, the Mask R-CNN (He et al., 2017) object detection algorithm output a binary pixel map for each detected pedestrian besides the bounding box. The binary pixel map shows which pixels in the image belong to the pedestrian. Second, we applied the Canny edge detection filter (Canny, 1986) to the binary pixel map of the target pedestrian to get its contour.

We also calculated the distribution of the landing points of the pedestrian-targeted saccades within the bounding box of the standardized pedestrian. The distribution showed that the participants directed their saccades mostly to the upper part of the pedestrian, presumably because the participants wanted to look at the pedestrians' faces (Figure 1C,D; Boucart et al., 2016; Crouzet, Kirchner, & Thorpe, 2010). This result suggests that the pedestrian-targeted saccades were directed to specific regions of the pedestrians. If the localization of one saccade was inaccurate (i.e. the saccade was not going toward the desired location), a correction might be made in the landing stage of the saccade, which marks a corrected/altered saccade.

Identification of altered saccades

In order to identify altered saccades (i.e. pedestrian-targeted saccades that contained a correction in the landing stage), we analyzed the speed-time curves of the saccades. For most of the pedestrian-targeted saccades, speed increased monotonically to the peak velocity and then monotonically decreased. One example is shown in Figure 2A. For some saccades, after reaching the peak values, speed decreased to a low value below the speed threshold for saccade detection, it increased again to above threshold, and then finally ended under threshold (one example is shown in Figure 2B). The intermediate low-speed stages typically occurred approximately 13 ms prior to the ends of the saccades (Figure 2D) and were usually accompanied with a direction change in the saccade trajectory. We used the presence of the intermediate low-speed stage as a mark of saccade landing correction (i.e. altered saccadic localization). We used 30°/s as a speed threshold (which is also equal to the conservative speed threshold for saccade parsing suggested by the Eyelink Online Parser) and defined the intermediate low-speed stages as the stages where at least three consecutive speed measurements (i.e. longer than 3 ms) were below the speed threshold prior to the landing stage (i.e. the last consecutive speed measurements that were below the speed threshold). Saccades that contained intermediate low-speed stages were defined as altered saccades. Three hundred ninety-six of the 2,067 pedestrian-targeted saccades were defined as altered saccades (more examples are shown in Figure 2C), and the rest were defined as direct saccades. The altered saccades were also associated with delays in saccadic localization. We conducted a linear regression that predicts saccade duration by distance between the saccade landing point and starting point in visual angle and whether it was an altered saccade or direct saccade. The result showed that the duration of the altered saccades was on average 12 ms longer than the direct saccades given the same landing-starting distance (permutation test $p < 0.001$). Note that the 95% percentile of the duration of the direct saccades was just 64 ms, which suggests that a delay of 12 ms is a significant amount.

Altered saccades were more frequent when the target pedestrians were flanked

In order to test whether altered saccades were due to crowding, we investigated whether the proportion of altered saccades differed for flanked versus unflanked pedestrians. In a preliminary control analysis, we first checked whether the retinal size (i.e. the subtended retinal angle) of the target pedestrian influenced the proportion of altered saccades. The correlation coefficient between the target retinal size and the proportion of altered saccades was −0.001 and the p value was 0.96. Therefore, we ruled out target retinal size as a confounder and did not include it in the following analyses of Experiment 1.

We divided the target pedestrians (i.e. the pedestrians on which saccades landed) into flanked and unflanked targets according to whether there were other pedestrians (flanking pedestrians) in the 2.5° vicinity around them (the circular area with a radius of 2.5° viewing angle centered at the centers of the target pedestrians). The previous examples in Figure 2A,B are also examples of flanked and unflanked targets, respectively. Only pedestrians were considered as flankers in this analysis, and we will consider other potential flanking objects in the next section. There

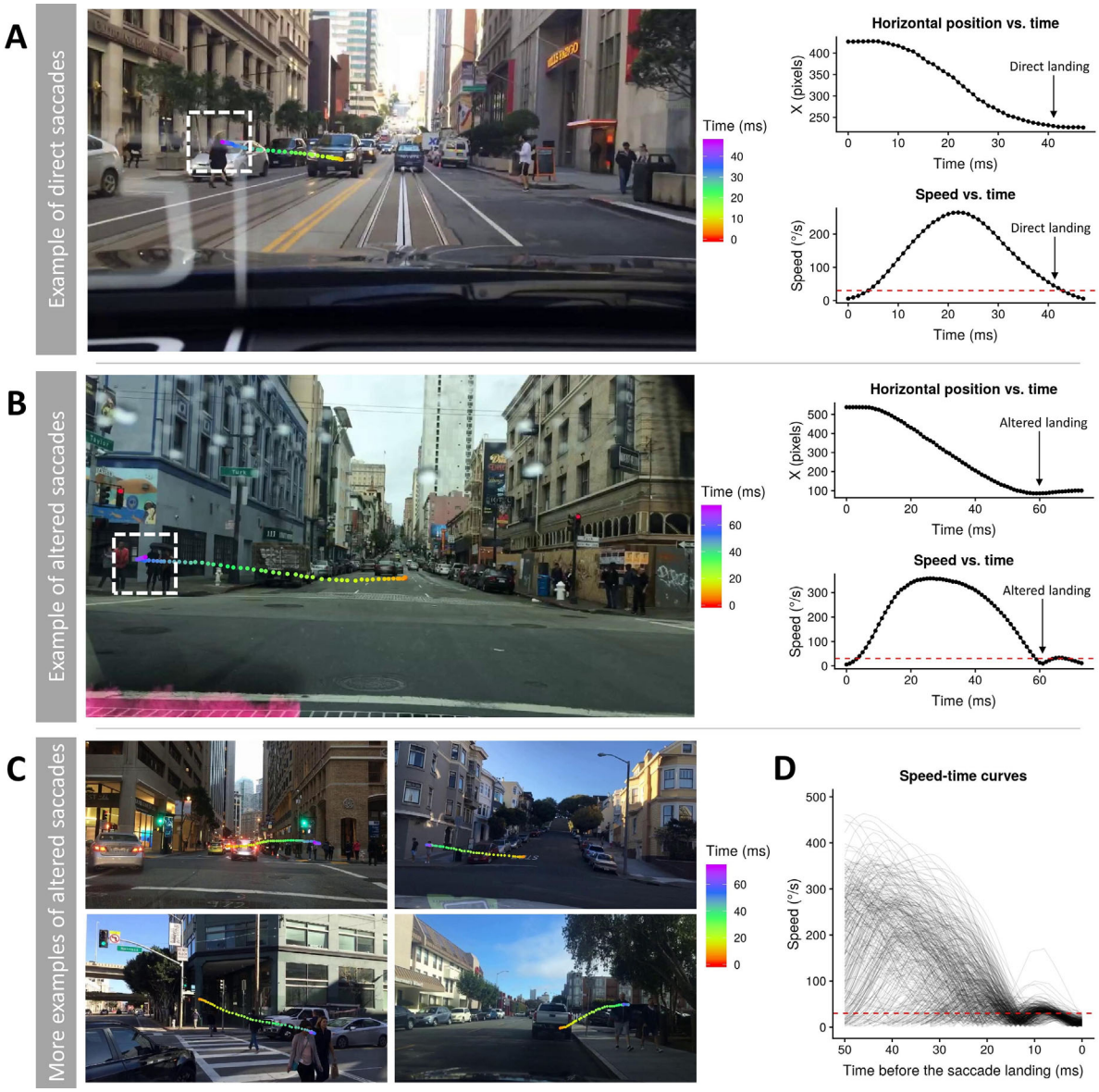

Figure 2. Direct and altered pedestrian-targeted saccades. Trajectories, horizontal pixel-time curves, and speed-time curves of one example direct saccade (A) and one example altered saccade (B). The white squares in the video frames and the arrows in the curve plots indicate the direct or altered landing of the two example saccades. The red dashed lines show the speed threshold of 30°/s. (C) The trajectories of various examples of altered pedestrian-targeted saccades. (D) Speed-time curves of all the altered pedestrian-targeted saccades over the last 50 ms prior to the transient saccade landing. The red dashed line shows the speed threshold of 30°/s.

were 1,123 flanked target pedestrians and 944 unflanked ones. For both flanked and unflanked targets, data showed that saccades became less accurate when the targets were more peripheral (i.e. the proportion of altered saccades, PAS) increased with increasing target eccentricity (the angular distance between the starting point of the saccade and the landing point of the saccade, Figure 3). If crowding occurred and led to more saccade inaccuracy, we would expect higher PAS for flanked targets than for unflanked targets given the same target eccentricity. According to Bouma's rule, crowding occurs when the target-flanker spacing is below approximately one half of the target eccentricity (Bouma, 1970). Therefore, we expected that (1) flanked and unflanked targets would show similar PAS on the eccentricity range between 0° and 5°, and (2) PAS for both flanked and unflanked targets would increase beyond 5° with increasing target eccentricity, but the increase for flanked targets should be significantly faster. To test our hypothesis, we fit the following

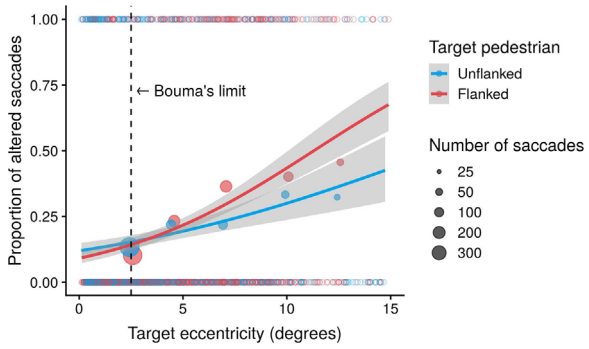

Figure 3. Proportion of altered saccades (PAS) versus target eccentricity for flanked and unflanked targets. Red represents the flanked targets (i.e. the target pedestrians with other flanking pedestrians in their 2.5° vicinity). Blue represents the unflanked targets (i.e. the target pedestrians with no other flanking pedestrian in their 2.5° vicinity). Hollow circles show the data of individual saccades, with altered saccades at the top and direct saccades at the bottom of the plot, respectively. Solid circles show the mean PAS of the eccentricity bins, and circle size indicates the number of saccades in the bin. Solid curves show the logistic regression fitting, and gray ribbons represent the 95% confidence intervals. The vertical dashed line shows 2.5° eccentricity around which PAS for flanked and unflanked targets were expected to be similar if the data followed Bouma's rule.

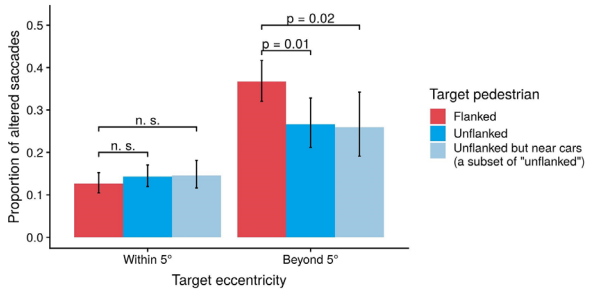

Figure 4. Mean proportion of altered saccades (PAS) for different target eccentricity ranges and different kinds of target pedestrians. Red bars represent the target pedestrians with other flanking pedestrians in their 2.5° vicinity (flanked). Dark blue bars represent the target pedestrians with no other flanking pedestrian in their 2.5° vicinity (unflanked). Light blue bars represent target pedestrians with cars but no other pedestrians in their 2.5° vicinity, which are a subset of the unflanked target pedestrians. Error bars represent 95% confidence intervals.

logistic regression model between PAS and target eccentricity:

$$\text{Model 1}: \log\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot (eccen - 2.5^{\circ})$$

where p is the PAS, and α and β are fitted parameters. The α indicates the fitted PAS at an eccentricity of 2.5° and β quantifies how fast the PAS increases with increasing eccentricity. We compared the parameters fitted for flanked targets and unflanked targets, and the results confirmed our hypothesis ($\alpha_f - \alpha_u = -0.08$, permutation test $p = 0.56$; $\beta_f - \beta_u = 0.088$, permutation test $p = 0.006$; Figure 3).

In addition to the logistic regression, we also calculated the mean PAS for the eccentricity range within 5° and the eccentricity range beyond 5°. The results (see Figure 4) showed similar PAS for flanked and unflanked targets for eccentricities within 5° (permutation test $p = 0.36$) and a significantly higher mean PAS for flanked targets than for unflanked targets for target eccentricity larger than 5° (permutation test $p = 0.01$).

More frequent altered saccades for pedestrian flankers than for car flankers

Crowding literature has previously shown that flankers similar to the target crowd more than dissimilar ones, from low-level stages of visual processing to high-level ones (Kooi et al., 1994; Reuther & Chakravarthi, 2014; for a review see Manassi & Whitney, 2018). Therefore, the flanking effect on pedestrian-targeted saccades discussed in the previous section should be specific to pedestrian flankers. Unflanked targets in the previous section were defined as target pedestrians when no other pedestrian was present in their 2.5° vicinity, so they might very well be flanked by other types of objects common to driving scenes. To this purpose, we identified the subset of unflanked target pedestrians that had cars but no other pedestrians within a 2.5° vicinity. We calculated mean PAS for two groups: target pedestrians within 5° that were flanked by cars, and target pedestrians beyond 5° eccentricity that were flanked by cars. We obtained values similar to the whole set of unflanked targets (i.e. flanked by no pedestrian). Comparison against the targets flanked by pedestrians showed that, beyond 5° eccentricity, the mean PAS associated with car flankers was significantly lower than that associated with pedestrian flankers (permutation test $p = 0.02$, see Figure 4). Within 5° eccentricity, there was no significant difference between car flankers and pedestrian flankers (permutation test $p = 0.37$, see Figure 4). These results suggest that target-similar flankers were more effective at crowding the target, consistent with the well-known similarity tuning of crowding (Andriessen & Bouma, 1976; Chung, Levi, & Legge, 2001; Kooi et al., 1994; Levi, 2008).

Consistency with Bouma's rule

A signature of crowding is that, regardless of the target and flanker size, critical spacing (i.e. the largest

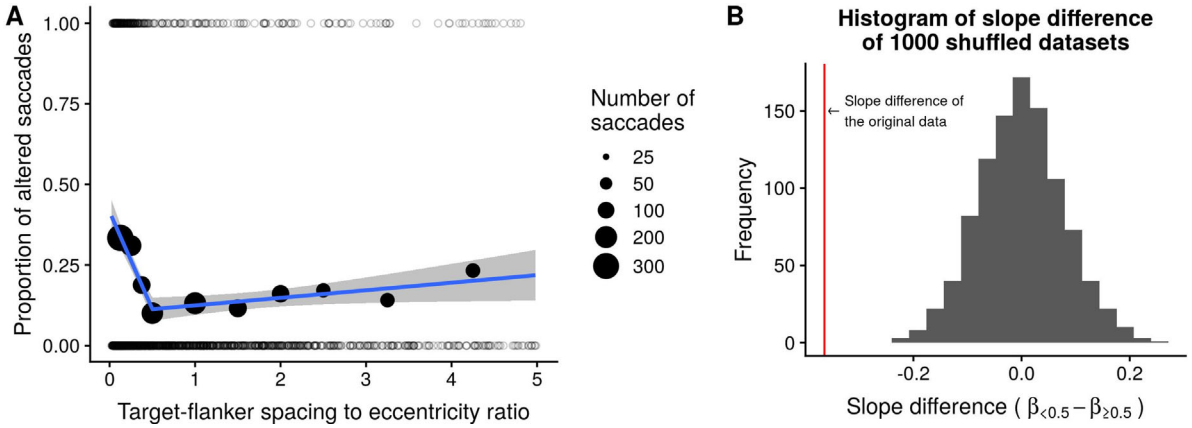

Figure 5. (A) Proportion of altered saccades (PAS) versus spacing-to-eccentricity ratio. Hollow circles show the data of individual saccades. Solid circles show the mean PAS of the ratio bins, and circle size indicates the number of saccades in the bin. Blue solid lines show the clipped line fit, and gray ribbons represent the 95% confidence intervals. (B) A permutation test was conducted to test the significance level of the difference between the two slopes of the clipped line fit ($\beta_{<0.5} - \beta_{\geq 0.5}$). The spacing-to-eccentricity ratio values of the saccades were shuffled 1,000 times. The histogram summarizes the slope differences fitted to the shuffled data. The red line shows the slope difference of the original data, which was significantly negative ($\beta_{<0.5} - \beta_{\geq 0.5} = -0.36$, permutation-test $p < 0.001$).

target-flanker spacing at which target recognition is affected) is roughly half target eccentricity (Bouma's rule-of-thumb; Bouma, 1970). Bouma's rule-of-thumb has been reported to be fairly consistent across a wide range of stimuli (e.g. oriented gratings, shapes, letters, faces, etc.), although the exact value can be strongly affected depending on the task, stimulus, attentional demands, etc. (Strasburger et al., 2011; Whitney & Levi, 2011). Nevertheless, the half-eccentricity rule of thumb is a reasonable a priori estimate of the average critical spacing at which crowding would often be expected to occur.

In order to test whether the influence of flanking pedestrians on saccade landing accuracy follows this rule, we collected the saccades that landed on pedestrians regardless of the flanker size, target size, and target eccentricity. We plotted the PAS against the spacing-to-eccentricity ratio (the ratio between the target-flanker spacing and the target eccentricity, Figure 5A). We then performed a clipped line fitting to the data (i.e. we fit two straight lines for the spacing-to-eccentricity ratio range below 0.5 and the ratio range above 0.5, respectively, with a shared intercept at the ratio equal to 0.5). In other words, the fitting included three free parameters: the two slopes of the two lines and the shared intercept at the ratio equal to 0.5. Data points over the ratio range above 5 were sparse and were, therefore, excluded from the fitting for robustness (11% of the data, exclusions of which had no effect on the qualitative results or significance). The fitting result (see Figure 5A) showed a steep negative slope over the ratio range below 0.5 ($\beta_{<0.5} = -0.33$), and a relatively flat slope over the ratio range above 0.5 ($\beta_{\geq 0.5} = 0.03$). Hence, the difference between the two slopes was consistent with Bouma's rule-of-thumb. However, target eccentricity could still be a confounder because larger target eccentricity leads to higher PAS (as seen in Figure 3) and the saccades of different spacing-to-eccentricity ratios have different mean target eccentricities. To determine the significance level of the difference in slopes ($\beta_{<0.5} - \beta_{\geq 0.5}$) after accounting for target eccentricity, we added target eccentricity as a regressor into the clipped line fitting model and conducted a permutation test where we shuffled the spacing-to-eccentricity ratio values of the saccades. Results showed a significant slope change ($\beta_{<0.5} - \beta_{\geq 0.5} = -0.36$, permutation-test $p < 0.001$; see Figure 5B), thus further confirming consistency with Bouma's rule.

Radial-tangential anisotropy

Another signature of crowding is a radial-tangential anisotropy: flankers aligned along the radial direction cause stronger crowding than flankers aligned along the tangential direction (Toet & Levi, 1992). To test whether the influence of flanking pedestrians on saccade landing accuracy follows the radial-tangential anisotropy, for each flanked target we calculated the angle between the line connecting the target and the flanker and the line connecting the starting and landing points of the saccade (α, $\alpha \in [0°, 90°]$). If the angle α was smaller than 30°, the flanker was identified as a radial flanker. If the angle α was greater than 60°, the flanker was identified as a tangential flanker.

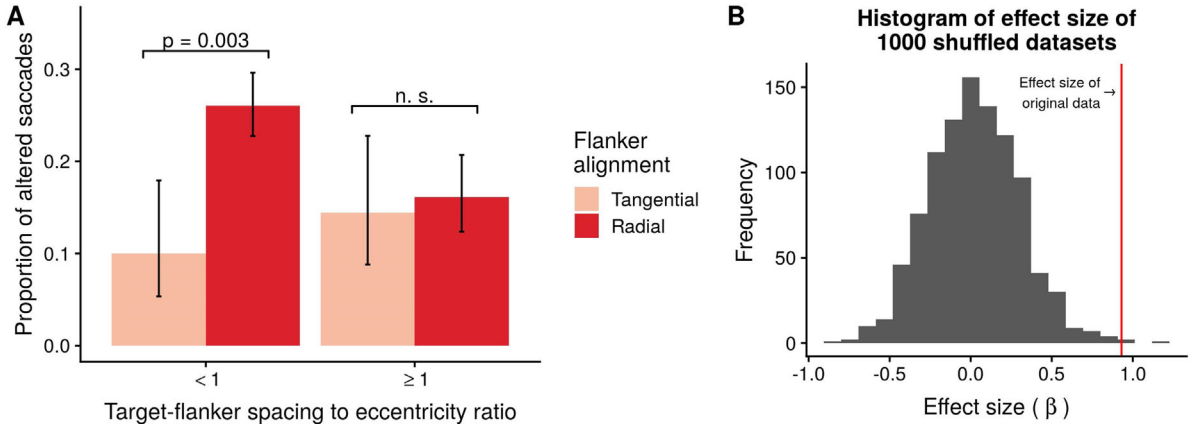

Figure 6. (A) Mean proportion of altered saccades (PAS) for tangential and radial flankers. Pink bars and red bars represent the tangential and radial flankers, respectively. Error bars represent 95% confidence intervals. (B) A permutation test was conducted to determine the significance level of the influence of radial versus tangential flanker alignment on PAS after accounting for target eccentricity, saccade direction and target-flanker depth difference (β in model 2). For the saccades with spacing-to-eccentricity ratio smaller than 1, the spacing-to-eccentricity ratio values of the saccades were shuffled 1,000 times. The histogram summarizes the β values fitted to the shuffled data. The red line shows the β value of the original data, which was significantly positive ($\beta = 0.93$, permutation-test $p = 0.003$).

First, among the saccades with a spacing-to-eccentricity ratio below 1 (where crowding might occur), we calculated the mean PAS separately for the saccades with radial flankers and the ones with tangential flankers. The mean PAS for radial flankers was higher than the mean PAS for tangential flankers (Figure 6A). This positive difference in mean PAS between radial and tangential flankers is consistent with the radial-tangential anisotropy of crowding.

However, in real-life driving scenes, the radial versus tangential flanker alignment naturally correlates with target-flanker 3D depth difference, saccade direction (e.g. horizontal or vertical) and target eccentricity. The potential confounding effects of these variables need to be tested. To quantify target-flanker depth difference, we used pictorial size difference between target and flanker pedestrians as an approximation based on the fact that pedestrians have relatively constant height. More specifically, we used logarithmic pictorial size difference (LPSD) as a variable to quantify target-flanker depth difference:

$$LPSD = \left| \log_2 \frac{flanker\ pictorial\ size}{target\ pictorial\ size} \right|$$

Greater LPSD indicates greater target-flanker depth difference.

To determine the significance level of the difference in mean PAS between radial and tangential flankers after accounting for the potential confounding factors, we fit the following model:

$$\text{Model 2} : \log\left(\frac{p}{1-p}\right) = \alpha + \beta \cdot X_{radial} + \gamma \cdot eccen + \sigma \cdot X_{horizontal} + \tau \cdot LPSD$$

where X_{radial} is a dummy variable that is equal to 1 when the flanker alignment of the saccade is radial and 0 when the flanker alignment is tangential, $X_{horizontal}$ is a dummy variable that is equal to 1 when the saccade is horizontal and 0 otherwise, and $LPSD$ is logarithmic pictorial size difference. β quantifies the independent influence of radial versus tangential flanker alignment on PAS after accounting for the influence of target eccentricity, saccade direction, and target-flanker depth difference. The fitting showed that $\beta = 0.93$, permutation test $p = 0.003$ (Figure 6B). The effect was significant even after accounting for target eccentricity.

We applied the same calculations to the saccades with a spacing-to-eccentricity ratio above 1 (i.e. where crowding was unlikely to occur). Results showed that there was no significant difference in mean PAS between the saccades with radial and tangential flankers after accounting for target eccentricity, saccade direction, and target-flanker depth difference (see Figure 6A, permutation test $p = 0.16$). Overall, the result suggested that the influence of flanking pedestrians on saccade landing accuracy is consistent with the radial-tangential anisotropy.

To summarize the results of Experiment 1, we found that visual clutter around the target pedestrians

are associated with altered saccadic localization in a manner that is consistent with the diagnostic criteria of crowding (i.e. Bouma's rule of thumb, flanker similarity tuning, and the radial-tangential anisotropy).

Experiment 2

In order to further confirm that the saccade landing inaccuracy shown above is a behavioral consequence of visual crowding, we conducted a more conventional crowding experiment with a pedestrian gender discrimination task. We identified the pedestrian-targeted saccades from Experiment 1 and used the video frames at the ending points of those saccades as static stimulus images for Experiment 2. Participants were asked to fixate at the starting point of the saccades and to identify the gender of the pedestrian on which the saccade landed. Gender discrimination tasks have been used before to successfully measure crowding effects in psychophysical experiments with cropped faces (Farzin, Rivera, & Whitney, 2009). In natural scenes, there is contextual information (e.g. flanker's gender, cloth color, physical size, etc.), which may allow participants to discriminate the target's gender even if the target is crowded. However, we can still use this task to test a lower bound of crowding effect. We hypothesized that the magnitude of crowding on gender recognition of the saccadic targets measured in Experiment 2 should correlate with the frequency of altered saccadic localization measured in Experiment 1, thus providing strong evidence that altered saccadic localization is a behavioral consequence of visual crowding.

Methods

Participants

Ten participants (5 women) participated in this experiment. They had normal or corrected normal vision. All participants except one were naïve to the purpose of the experiment. All experimental procedures were approved by and conducted in accordance with the guidelines and regulations of the University of California, Berkeley Institutional Review Board. Participants were affiliates of UC Berkeley and provided informed consent in accordance with the Institutional Review Board guidelines of the University of California, Berkeley.

Equipment and stimuli

Display setup was the same as Experiment 1. Six hundred two pedestrian-targeted saccades identified in Experiment 1 were used to make the stimuli for Experiment 2. The video frames at the ending points of those saccades were extracted as static stimulus images. The average retinal size of the target pedestrians was 4.2° and the standard deviation was 2.6°.

Procedure

On each trial, participants viewed one stimulus image in full screen and were asked to report the perceived gender of the target pedestrian by a keypress. Each participant was presented with 140 stimulus images in total. It was ensured that the 140 stimulus images were extracted from all different videos so that the target pedestrian in one image would not appear in another image. Each participant viewed and responded to their 140 stimulus images three times in three blocks of trials, with self-paced pauses in between (i.e. 140 trials in each block). The stimulus images were presented in a different random order in each block. Participants viewed the stimulus images with required fixation in the first two blocks and freely in the third block. The responses made in the third block were used as the subjective within-subject ground-truth of the gender of the target pedestrians to determine whether their response made in the first two blocks were correct. Using other participants' responses made in the third block as (between subject) ground-truth did not change the results qualitatively.

In the first two blocks, participants' eye movements were tracked using an Eyelink 1000 at 1,000 Hz for fixation monitoring. On each trial, first a pre-cue image was displayed. The pre-cue image consisted of a gray background, a white fixation cross showing the required fixation point (the starting point of the pedestrian-targeted saccade in Experiment 1), and a red bounding box showing the location of the target pedestrian (the pedestrian on which the pedestrian-targeted saccade in Experiment 1 landed). The pre-cue image was displayed for at least one second and until the participant fixated on the fixation cross. The stimulus image was then displayed with the fixation cross superimposed and two red bars right above and below the target pedestrian. The two red bars were placed to help the participant appreciate which pedestrian was the target pedestrian. Fixation was monitored in real-time, and the target was presented gaze-contingently. If the participant's gaze was more than 50 pixels (1.5° to approximately 1.7° in viewing angle) away from the required fixation point, the stimulus image was masked by the pre-cue image. The stimulus was displayed again once the fixation was restored. This process continued until the response was made, but no longer than two seconds after the initial onset of the stimulus image. The participant reported the perceived gender by key press (1 for male and 0 for female) and then the next trial started.

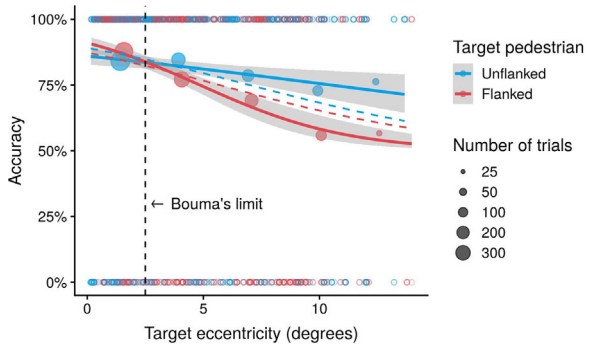

Figure 7. Gender discrimination accuracy versus target eccentricity for flanked and unflanked targets. Hollow circles show individual trial data. Solid circles show the mean accuracies of the eccentricity bins, and circle size indicates the number of trials in the bin. Solid curves show the logistic regression fit, and the gray ribbons represent the 95% confidence intervals. Dashed lines show the baseline accuracy-eccentricity curves for flanked and unflanked targets, under the null hypothesis that the accuracy depends on target eccentricity and target retinal size but not on whether the target pedestrian is flanked or not. The vertical dashed line shows 2.5° eccentricity around which the accuracy for flanked and unflanked targets are expected to be similar if the data follows Bouma's rule.

In the third block, there was no pre-cue image. In each trial, the stimulus image was displayed with the fixation cross and the red bounding box around the target pedestrian on top. The participant viewed the stimulus image freely with unlimited time until they made a response by key press. The next trial started right after the response was made.

Results and discussion

Reduced recognition when the target pedestrians were flanked

Following the selection criterion in Experiment 1, we divided target pedestrians into flanked and unflanked categories, depending on whether there were other flanking pedestrians within the 2.5° vicinity of the target pedestrians. For both flanked and unflanked targets, data showed that gender discrimination accuracy dropped with increasing target eccentricity (Figure 7). Furthermore, we would expect lower accuracy for flanked targets than for unflanked targets given the same target eccentricity. According to Bouma's rule (Bouma, 1970), we expected that flanked and unflanked targets would show similar accuracy on the eccentricity range between 0° and 5°; beyond 5°, accuracy for both flanked and unflanked targets would decrease with increasing target eccentricity, but the decrease for flanked targets should be significantly faster.

Therefore, to test our hypothesis, we fit the following logistic regression model between gender discrimination accuracy and target eccentricity:

$$\text{Model 3}: \log\left(\frac{accur-0.5}{1-accur}\right) = \alpha + \beta\cdot(eccen-2.5^{\circ})$$

where *accur* is the gender discrimination accuracy, *eccen* is the eccentricity of the target, and α and β are fitted parameters. α indicates the fitted accuracy at an eccentricity of 2.5° and β quantifies how fast the accuracy decreases with increasing eccentricity. We compared the parameters fitted for flanked targets and unflanked targets, and the results followed our expectation ($\alpha_f - \alpha_u = 0.016$, permutation test $p = 0.92$; $\beta_f - \beta_u = -0.22$, permutation test $p < 0.001$; see Figure 7).

It might be argued that retinal size of the target pedestrian is a possible confounder, because data showed that flanked pedestrians on average had smaller retinal sizes, and that low gender discrimination accuracy was correlated with small target retinal size. To determine the baseline accuracy-eccentricity curves for flanked and unflanked targets under the null hypothesis that the accuracy was influenced by both the eccentricity and target retinal size but not whether the target was flanked or not, we added target retinal size as a regressor into model 3 and fit the following new model to all the data (i.e. the union of flanked and unflanked targets):

$$\text{Model 4}: \log\left(\frac{accur-0.5}{1-accur}\right) = \alpha + \beta\cdot(eccen-2.5^{\circ}) + \mu\cdot size$$

where *size* is the retinal size of the target pedestrian. This model captured how the accuracy varied based on target eccentricity and target retinal size, regardless of whether the target was flanked or not. For each trial, we then simulated the outcome of the trial (i.e. correct or wrong) based on a binomial distribution with the probability of correctness equal to the accuracy predicted by model 4. Simulated data showed how data would distribute under the null hypothesis. We then fit model 3 to the simulated data separately for flanked targets and unflanked targets to obtain the baseline accuracy-eccentricity curves under the null hypothesis. Baseline curves are plotted as dashed curves in Figure 7. To calculate the significance level of the difference of how fast accuracy decreased with eccentricity between flanked and unflanked targets after discounting the effect of target retinal size, we fit model 4 separately to the flanked data and unflanked data. This time $\beta_f - \beta_u = -0.24$. We ran a permutation test where we shuffled the flanked/unflanked labels of all the trials to get a null distribution of $\beta_f - \beta_u$. The permutation test showed

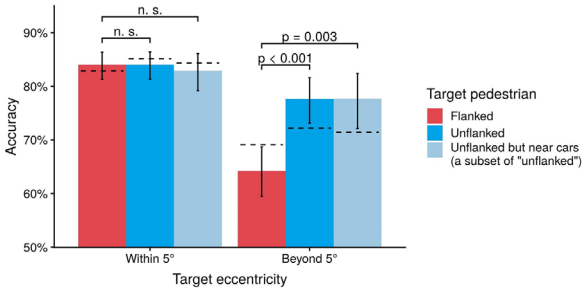

Figure 8. Mean gender discrimination accuracy for different target eccentricity ranges and different kinds of target pedestrians. Dashed lines show the baseline values under the null hypothesis that within 5° eccentricity or beyond 5° eccentricity the gender discrimination accuracy only depends on target retinal size but not the types of the target pedestrians. Error bars represent 95% confidence intervals.

that $p < 0.001$. Hence, we confirmed that even after accounting for the effect of target retinal size, gender discrimination accuracy dropped with increasing eccentricity significantly faster for flanked targets than for unflanked targets. Similarly, we confirmed that after discounting the effect of target retinal size, there was no significant difference in the accuracy at 2.5° eccentricity between flanked and unflanked targets ($\alpha_f - \alpha_u = 0.23$, permutation test $p = 0.49$).

Besides the logistic regression, we also calculated mean gender discrimination accuracy for the eccentricity range below 5° and the eccentricity range beyond 5°. Results (Figure 8) showed similar accuracies for flanked and unflanked targets for target eccentricities less than 5° ($accur_f - accur_u = -0.05\%$), and a lower mean accuracy for flanked targets than for unflanked targets for target eccentricities beyond 5° ($accur_f - accur_u = -13.4\%$).

In order to account for the influence of target retinal size, we calculated the baseline accuracies under the null hypothesis that either below 5° eccentricity or above 5° eccentricity the mean accuracy is only influenced by target retinal size but not whether the target is flanked or not. We first fit the following logistic regression model to the data below 5° eccentricity:

$$\text{Model 5} : \log\left(\frac{accur}{1-accur}\right) = \alpha + \mu \cdot size$$

For each trial below 5° eccentricity, we then simulated the outcome of the trial (i.e. correct or wrong) based on a binomial distribution with the probability of correctness equal to the accuracy predicted by model 5. The simulated data showed how the data would distribute under the null hypothesis. We then calculated the mean accuracies of the flanked and unflanked trials of the simulated data, which are the baseline accuracies under the null hypothesis (plotted as dashed lines in Figure 8). To determine the significance level of the difference between flanked and unflanked data in the original data, we first fit the following model to the data below 5° eccentricity:

$$\text{Model 6} : \log\left(\frac{accur}{1-accur}\right) = \alpha + \beta \cdot X_{flanked} + \mu \cdot size$$

where $X_{flanked}$ is a dummy variable that is equal to 1 when the target is flanked and 0 otherwise, and β quantifies the independent influence of being flanked versus unflanked on mean accuracy after accounting for the influence of target retinal size. The fitting showed that $\beta = 0.18$, permutation test $p = 0.20$. Hence, there was no significant effect from being flanked versus unflanked for the data below 5° eccentricity. We applied the same analysis to the data above 5° eccentricity to get the baseline mean accuracies and the significance level of the independent effect of being flanked versus unflanked. Baseline accuracies are shown as dashed lines in Figure 8; $\beta = -0.54$, permutation test $p < 0.001$. Flanked pedestrians versus unflanked pedestrians significantly decreased mean accuracy even after accounting for the influence of target retinal size.

Lower accuracy for pedestrian flankers than for car flankers

Similar to Experiment 1 (Figure 4), we selected the subset of unflanked target pedestrians that had cars but no other pedestrian in their 2.5° vicinity. We calculated mean gender discrimination accuracy for target pedestrians flanked by cars for within 5° eccentricity and beyond 5° eccentricity, and obtained values similar to the whole set of unflanked targets (i.e. flanked by no pedestrian; see Figure 8). We then compared these mean accuracies calculated with these trials with car flankers to the mean accuracies calculated with the trials with pedestrian flankers. We applied the same calculations to account for the influence of target retinal size as a confounder. Baseline mean accuracies are shown as dashed lines in Figure 8. Results showed that, beyond 5° eccentricity, having pedestrian flankers versus car flankers significantly decreased mean accuracy ($\beta = -0.53$, permutation test $p = 0.003$). Within 5° eccentricity, there was no significant independent effect from having pedestrian flankers versus car flankers ($\beta = 0.17$, permutation test $p = 0.30$). These results, again, show the target-flanker similarity tuning of crowding in a gender discrimination task.

Consistency with Bouma's rule

To test Bouma's rule, we collected all the trials for the following analysis regardless of flanker size,

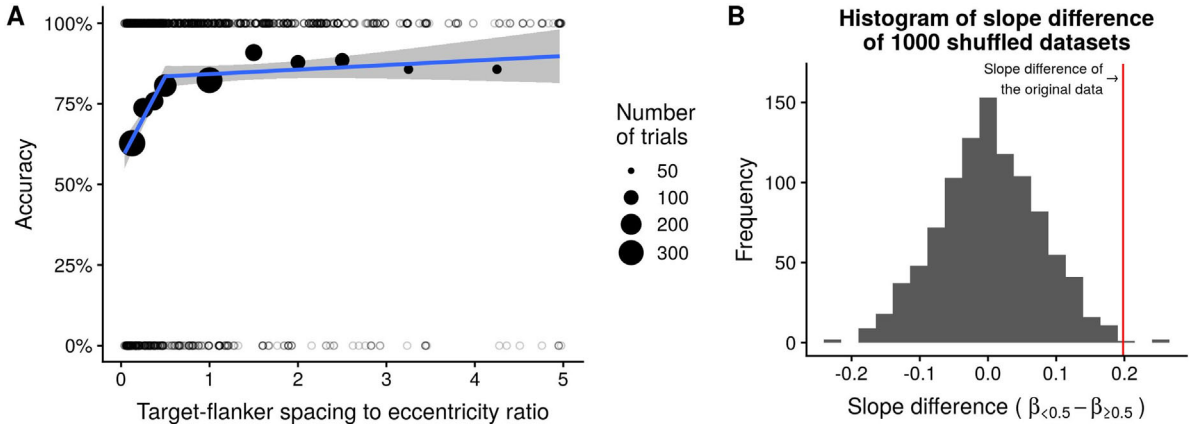

Figure 9. (A) Gender discrimination accuracy versus spacing-to-eccentricity ratio. Hollow circles show individual trial data. Solid circles show the mean accuracy of the ratio bins, and circle size indicates the number of trials in the bin. Solid curves show the clipped line fit, and gray ribbons represent the 95% confidence intervals. (B) A permutation test was conducted to test the significance level of the difference between the two slopes of the clipped line fit ($\beta_{<0.5} - \beta_{\geq 0.5}$). The spacing-to-eccentricity ratio values of the trials were shuffled 1,000 times. The histogram summarizes the slope differences fitted to the shuffled data. The red line shows the slope difference of the original data, which was significantly positive ($\beta_{<0.5} - \beta_{\geq 0.5} = 0.20$, permutation-test $p = 0.01$).

target size, and target eccentricity. We plotted gender discrimination accuracy against spacing-to-eccentricity ratio (the ratio between the target-flanker spacing and the target eccentricity, Figure 9A). We then performed the same clipped line fitting as described in Experiment 1 (see Figure 5A; i.e. we fit two straight lines for the spacing-to-eccentricity ratio range below 0.5 and the ratio range above 0.5, respectively, with a shared intercept at the ratio equal to 0.5). Data points over the ratio range above 5 were sparse and were, therefore, excluded from the fitting for robustness. The fitting result (see Figure 9A) showed a steep positive slope over the ratio range below 0.5 ($\beta_{<0.5} = 0.52$) and a relatively flat slope over the ratio range above 0.5 ($\beta_{\geq 0.5} = 0.01$). The difference between the two slopes was consistent with Bouma's rule-of-thumb. To determine the significance level of the slope difference ($\beta_{<0.5} - \beta_{\geq 0.5}$) after accounting for target eccentricity and target retinal size, we added target eccentricity and target retinal size as additional regressors into the clipped line fitting model, and conducted a permutation test where we shuffled the spacing-to-eccentricity ratio values of the trials. The results showed a significant slope change ($\beta_{<0.5} - \beta_{\geq 0.5} = 0.20$, permutation-test $p = 0.01$; Figure 9B).

Radial-tangential anisotropy

In order to check whether our gender discrimination task showed any radial-tangential anisotropy, we conducted the following analysis. Among the saccades with a spacing-to-eccentricity ratio below 1, where crowding might occur, we calculated mean gender discrimination accuracy separately for the saccades with radial flankers and the ones with tangential flankers (Figure 10A). Besides target eccentricity and target size, radial-tangential flanker alignment naturally correlates with target-flanker 3D depth difference and target meridian (i.e. target closer to horizontal or vertical meridian). To determine the significance level of the difference between radial and tangential flanker alignment after accounting for these confounding factors, similarly to model 2 in Experiment 1, we first fit the following model:

$$\text{Model 7} : \log\left(\frac{accur}{1-accur}\right) = \alpha + \beta \cdot X_{radial} + \gamma \cdot eccen + \mu \cdot size + \sigma \cdot X_{horizontal} + \tau \cdot LPSD$$

where X_{radial} is a dummy variable that is equal to 1 when the flanker alignment is radial and 0 when the flanker alignment is tangential, $X_{horizontal}$ is a dummy variable that is equal to 1 when the target is closer to horizontal meridian than vertical meridian and 0 otherwise, and $LPSD$ is logrithmic pictorial size difference between target and flanker, and β quantifies the independent influence of being radial versus tangential on mean accuracy after accounting for the influence of target eccentricity, target retinal size, target meridian, and target-flanker depth difference. Results showed that radial versus tangential flanker alignment significantly decreased mean accuracy after accounting for the confounding factors ($\beta = -0.72$, permutation test $p = 0.003$; Figure 10B). We applied the same calculations to the data with spacing-to-eccentricity ratios above 1 where crowding was unlikely to

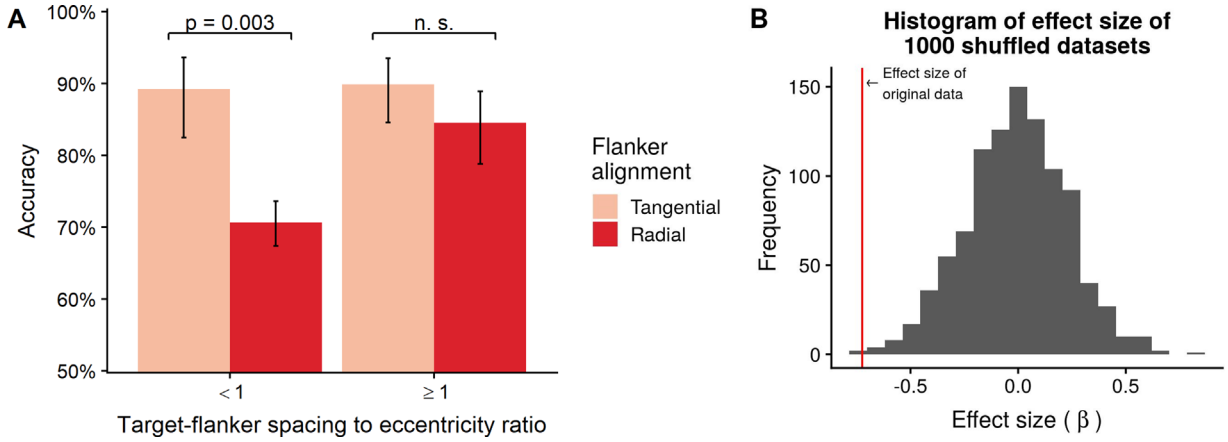

Figure 10. (A) Mean gender discrimination accuracy for tangential and radial flankers with low spacing-to-eccentricity ratios (< 1) and high spacing-to-eccentricity ratios (≥ 1). Error bars represent 95% confidence intervals. (B) A permutation test was conducted to determine the significance level of the influence of radial versus tangential flanker alignment on gender discrimination accuracy after accounting for target eccentricity, target retinal size, target meridian, and target-flanker depth difference (β in model 7). For the trials with spacing-to-eccentricity ratio smaller than 1, spacing-to-eccentricity ratio values of the trials were shuffled 1,000 times. The histogram summarizes the β values fitted to the shuffled data. The red line shows the β value of the original data, which was significantly negative ($\beta = -0.72$, permutation-test $p = 0.003$).

occur. Results showed that there was no significant independent effect of radial versus tangential flanker alignment after accounting for the confounding factors ($\beta = -0.26$, permutation test $p = 0.40$). Overall, the result confirmed the presence of a radial-tangential anisotropy in Experiment 2, consistent with the findings in Experiment 1 (see Figure 6).

Correlation between saccade landing accuracy and gender discrimination accuracy

Because the stimuli of Experiment 2 were made based on the pedestrian-targeted saccades collected in Experiment 1, we correlated gender discrimination accuracy of Experiment 2 with saccade landing accuracy of Experiment 1. If the altered saccadic localization discussed in Experiment 1 is a behavioral consequence of crowding observed in Experiment 2, trials in Experiment 2 that corresponded to the altered saccades in Experiment 1 should show a lower mean gender discrimination accuracy than the trials that corresponded to the direct saccades. The results indeed showed that the mean gender discrimination accuracy of the altered saccades was lower than the one of the direct saccades ($accur_{altered} - accur_{direct} = -8.6\%$; Figure 11A). To determine the significance level of the difference in accuracy between altered and direct saccades after accounting for target eccentricity and target retinal size, we fit the following model to the data:

$$\text{Model 8} : \log\left(\frac{accur}{1-accur}\right)$$

$$= \alpha + \beta \cdot X_{altered} + \gamma \cdot eccen + \mu \cdot size$$

where $X_{altered}$ is a dummy variable that is equal to 1 when the trial corresponded to an altered saccade in Experiment 1 and 0 otherwise, and β quantifies the independent influence corresponding to an altered saccade versus to a direct saccade on the mean gender discrimination accuracy after accounting for the confounding influence of target eccentricity and target retinal size. The results showed that altered versus direct saccades significantly decreased mean gender discrimination accuracy after accounting for target eccentricity and target retinal size ($\beta = -0.31$, permutation test $p = 0.02$; Figure 11B).

General discussion and conclusions

In Experiment 1, we used crowd-sourced natural driving videos as stimuli and recorded participants' eye movements during a simulated driving task. We identified the saccades that landed on pedestrians by using a deep learning object detection algorithm (He et al., 2017). We found that visual clutter around target pedestrians is associated with altered saccadic localization. Furthermore, this altered saccadic localization is consistent with the diagnostic criteria of crowding (Bouma's rule of thumb, flanker similarity tuning, and the radial-tangential anisotropy). In Experiment 2, we used a pedestrian gender discrimination task with a more conventional psychophysical crowding paradigm to confirm that

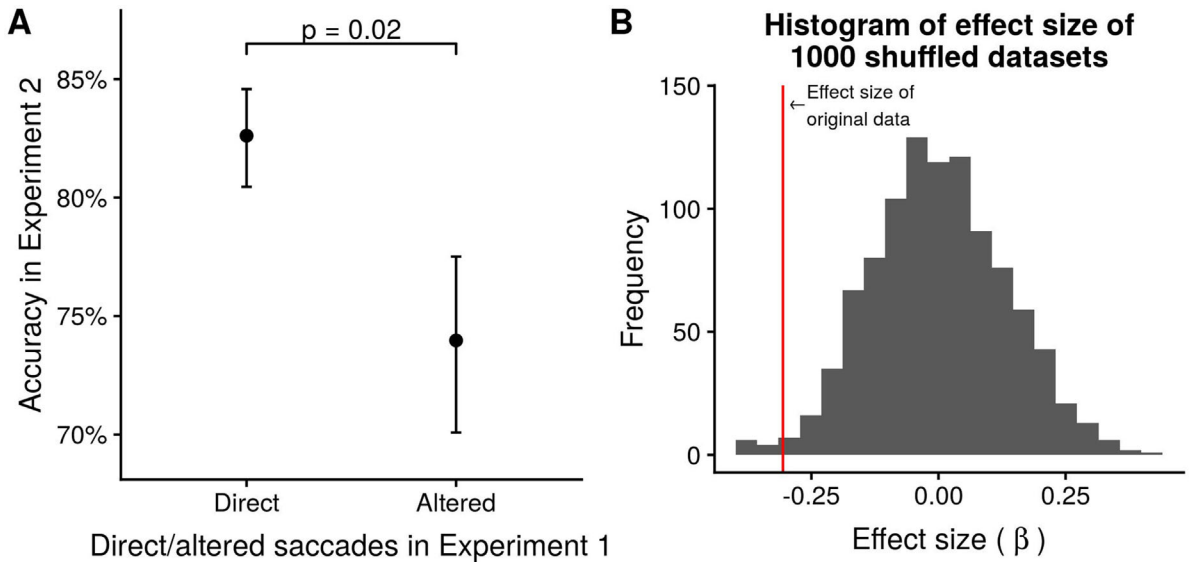

Figure 11. (A) Mean gender discrimination accuracy of the trials using stimuli from altered saccades and direct saccades in Experiment 1. The *p* value was calculated from the permutation test described in panel B. (B) A permutation test was conducted to determine the significance level of the influence of altered saccades versus direct saccades on gender discrimination accuracy after accounting for target eccentricity and retinal size (β in model 8). Altered/direct saccade labels were shuffled 1000 times. The histogram summarizes the β values fitted to the shuffled data. The red line shows the β value of the original data, which was significantly negative ($\beta = -0.31$, permutation-test $p = 0.02$). The results show that the altered saccades in Experiment 1 are significantly associated with lower gender discrimination accuracy in Experiment 2, after accounting for target eccentricity and retinal size.

visual crowding indeed occurs in the recognition of the pedestrians targeted by the saccades recorded in Experiment 1. Importantly, we showed that the altered saccadic localization observed in Experiment 1 is associated with the degree of crowding of the saccade targets measured in Experiment 2 (see Figure 11). Taken together, the results of Experiments 1 and 2 show strong evidence that visual crowding occurs in natural driving scenes, and has behavioral consequences in driving-like situations (i.e. altered saccadic localization, which is associated with delays in saccadic localization as well).

To the best of our knowledge, our work demonstrates for the first time that crowding occurs in dynamic, natural driving-related scenes. Crowding has been studied for decades from low-level features, such as oriented gratings (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), letters (Bouma, 1970; Pelli et al., 2004), and symbols (Grainger, Tydgat, & Isselé, 2010), to high level features, such as faces (Farzin et al., 2009; Louie, Bressler, & Whitney, 2007; Sun & Balas, 2015) and biological motions (Ikeda, Watanabe, & Cavanagh, 2013). One important motivation of all these studies is that crowding supposedly influences how we recognize cluttered objects in real life. However, the stimuli used in these studies are typically artificial, unnatural, and often static (Bex & Dakin, 2005; Bex, Dakin, & Simmers, 2003; Dakin, Greenwood, Carlson, & Bex, 2011; Maus, Fischer, & Whitney, 2011). Whether the rich findings of this literature can apply to real-life are questionable because of the characteristics of dynamic natural scenes. (i) Complex configurations. Recent studies show that the effects of crowding can be reduced by grouping processes in complex images. For example, crowding is diminished when flankers can be grouped together or segmented from a central target (Bex et al., 2003; Livne & Sagi, 2007, 2010; Saarela, Sayim, Westheimer, & Herzog, 2009), and crowding is weaker for the objects containing internal structures than for object silhouettes or letters (Wallace & Tjan, 2011). (ii) The variety of target-flanker differences. Crowding is tuned to target-flanker similarity (for a review see Whitney & Levi, 2011). However, in previous studies, targets and flankers typically only vary along one simple dimension (e.g. gratings with different orientations). In natural scenes, even clustered objects of the same type differ in various dimensions (e.g. pedestrians of different sizes, genders, cloth colors, etc.). (iii) Natural scene depth and perspective. In natural scenes, clustered objects may appear at different depths and previous studies suggest that depth difference between target and flankers can release crowding (Kooi et al., 1994; Sayim, Westheimer, & Herzog, 2008). Therefore, studies of crowding with dynamic natural stimuli/scenes are still needed for studying the impact of crowding in object recognition in real life.

In an important step toward studying crowding in natural scenes, Wallis and Bex (2012) conducted a clever

experiment where participants were asked to identify synthetic "dead leave" patches in natural scenes and found that the threshold size of "dead leave" patches scaled with eccentricity in manners consistent with crowding. However, the "dead leave" patches were added artificially and could be inconsistent with the perspective and depth of the natural scene. Gong et al. (2018) studied crowding in the recognition of the gist of natural scenes, but they did not study crowding in object recognition in natural scenes. Previous studies also showed crowding of moving targets, but limited to oriented gratings and simple shapes (Bex & Dakin, 2005; Bex et al., 2003; Dakin et al., 2011; Maus et al., 2011). Using natural scenes as stimuli allows us to test the potential impact of crowding in real-life. In real-life 3D environment, 3D depth difference between the target and flankers naturally correlates with the retinal separation between them. Although our experiment closely mimics real driving situations, future studies can try to disentangle 3D depth difference and retinal separation to closely study the influence of 3D depth on crowding in natural scenes.

Importantly, our study used saccades as a tool to study crowding in driving implicitly, while avoiding forced and unnatural explicit responses. Observers in Experiment 1 were not asked anything about pedestrians and did not know the purpose of the experiment or that pedestrians had any special place in the videos. Although this reduces the efficiency of the experiment, it avoids potentially confounding effects of an explicit task. Given that task demands can alter measured crowding thresholds (Huckauf, 2007), the use of an implicit measure, like we have developed here, may be useful in future studies of crowding and eye movements or other goal-directed action.

Our work supports and extends the known link between saccades and crowding (Greenwood, Szinte, Sayim, & Cavanagh, 2017; Harrison, Mattingley, & Remington, 2013; Wolfe & Whitney, 2014; Yildirim, Meyer, & Cornelissen, 2015). Specifically, Greenwood et al. (2017) found that saccade precision and the size of the crowding zone vary across the visual field with a strong correlation. Along the same lines, Yildirim et al. (2015) demonstrated that saccadic target localization is tuned to target-flanker similarity. Our results are consistent with these findings.

Crowding on pedestrian recognition is also consistent with previous studies on crowding between high-level features and objects. It has been shown that crowding occurs in the recognition of faces (Farzin et al., 2009; Louie et al., 2007), and that both local features and global configuration of flanking faces contribute to crowding (Sun & Balas, 2015). In addition, Ikeda et al. demonstrated crowding of biological motions using moving dots with configurations of walkers (Ikeda et al., 2013), thus providing further support for the idea that crowding can occur between dynamic representations, similar to moving pedestrians. Our study shows that pedestrian flankers impose stronger crowding on target pedestrians than car flankers. This result is, thus, consistent with the idea that crowding occurs at different levels (Manassi & Whitney, 2018).

Our study focuses on driving because it is presumably the most important real-life situation that may involve crowding, given its frequency and potentially fatal risks. Sanocki et al. (2015) conducted an experiment where observers looked for pedestrians in briefly presented traffic scenes and found higher miss rates associated with cluttered scenes. However, previous studies show that crowding has little effect on target detection (Levi, Hariharan, et al., 2002; Levi, Klein, et al., 2002; Pelli et al., 2004). Moreover, Sanocki et al. did not test the hallmarks of crowding, such as Bouma's rule-of-thumb. Therefore, it is unclear whether the effect they found was due to crowding or other phenomena such as visual masking (Pelli et al., 2004). We believe that our study is the first one to clearly demonstrate the behavioral consequences of crowding in dynamic driving-like situations (i.e. altered saccadic localization). Even though visual crowding impacts mostly peripheral vision, it has been shown that peripheral vision acquires extensive information essential for driving (Wolfe, Dobres, Rosenholtz, & Reimer, 2017) and perceptual load on the peripheral regions of the visual scene affects the ability to detect critical events initiating from the roadsides (Marciano & Yeshurun, 2015). Therefore, our finding has also significant implications for public safety. On the one hand, it raises safety concerns about visual clutter in traffic scenes; on the other hand, it suggests that the knowledge that we have gained from decades of studies of crowding can be used in traffic designs to address these concerns. For example, road signs may need to be placed with enough spacing in between, or construction workers may need to wear safety vests that differ in color from any nearby objects or signage.

Keywords: crowding, driving, contextual modulation, eye movements, saccade localization, spatial vision

Acknowledgments

This work was funded in part by NIH Grant 1R01CA236793-01.

Commercial relationships: none.
Corresponding author: Ye Xia.
Email: yexia@berkeley.edu.
Address: Department of Psychology, 2121 Berkeley Way, University of California, Berkeley, CA 94720-1650, USA.

References

Andriessen, J. J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research, 16*, 71–78.

Bex, P. J., & Dakin, S. C. (2005). Spatial interference among moving targets. *Vision Research, 45*, 1385–1398.

Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Research, 43*, 2895–2904.

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*, 10.

Bouma, H. (1970, April). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*, 679–698.

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., . . . Lin, D. (2019). Hybrid task cascade for instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 4974–4983.

Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41*, 1833–1850.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., . . . Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016 December*, 3213–3223.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*, 1–17.

Dakin, S. C., Greenwood, J. A., Carlson, T. A., & Bex, P. J. (2011). Crowding is tuned for perceived (not physical) location. *Journal of Vision, 11*, 2.

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of mooney faces. *Journal of Vision, 9*, 1–15.

Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science, 142*, 979–980.

Gong, M., Xuan, Y., Smart, L. J., & Olzak, L. A. (2018). The extraction of natural scene gist in visual crowding. *Scientific Reports, 8*, 14073.

Grainger, J., Tydgat, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 673.

Greenwood, J. A., Szinte, M., Sayim, B., & Cavanagh, P. (2017). Variations in crowding, saccadic precision, and spatial localization reveal the shared topology of spatial vision. *Proceedings of the National Academy of Sciences of the United States of America, 114*, E3573–E3582.

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

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision,* 2961–2969.

Huang, X., Wang, P., Cheng, X., Zhou, D., Geng, Q., & Yang, R. (2019). The ApolloScape open dataset for autonomous driving and its application. https://doi.org/10.1109/TPAMI.2019.2926463 . [Epub ahead of print].

Huckauf, A. (2007). Task set determines the amount of crowding. *Psychological Research, 71*, 646–652.

Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision, 13*, 20.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8*, 255–279.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*, 635–654.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision, 2*, 167–177.

Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision, 2*, 140–166.

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 8759–8768.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision, 7*, 1–12.

Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision, 10*, 1–14.

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between

configural representations of faces in crowded scenes. *Journal of Vision, 7*, 24.

Maddern, W., Pascoe, G., Linegar, C., & Newman, P. (2017). 1 year, 1000 km: The Oxford Robotcar dataset. *The International Journal of Robotics Research, 36*, 3–15.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision, 7*, 1.

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*, 13.

Manassi, M., & Whitney, D. (2018). Multi-level crowding and the paradox of object recognition in clutter. *Current Biology, 28*, R127–R133.

Marciano, H., & Yeshurun, Y. (2015). Perceptual load in different regions of the visual scene and its relevance for driving. *Human Factors, 57*, 701–716.

Maus, G. W., Fischer, J., & Whitney, D. (2011). Perceived positions determine crowding. *PLoS One, 6*, e19796.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*, 12.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*, 1129–1135.

Reuther, J., & Chakravarthi, R. (2014). Categorical membership modulates crowding: Evidence from characters. *Journal of Vision, 14*, 5.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9*, 1–11.

Sanocki, T., Islam, M., Doyon, J. K., & Lee, C. (2015). Rapid scene perception with tragic consequences: Observers miss perceiving vulnerable road users, especially in crowded traffic scenes. *Attention, Perception, & Psychophysics, 77*, 1252–1262.

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science, 21*, 641–644.

Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision, 8*, 1–9.

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*, 13.

Sun, H. -M., & Balas, B. (2015). Face features and face configurations both contribute to visual crowding. *Attention, Perception, & Psychophysics, 77*, 508–519.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*, 1349–1357.

Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision, 11*, 19.

Wallis, T. S. A., & Bex, P. J. (2012). Image correlates of crowding in natural scenes. *Journal of Vision, 12*, 6.

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research, 15*, 1137–1141.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*, 160–168.

Wolfe, B. A., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. *Applied Ergonomics, 65*, 316–325.

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

Xia, Y., Kim, J., Canny, J., Zipser, K., Canas-Bajo, T., & Whitney, D. (2020). Periphery-fovea multi-resolution driving model guided by human attention. *The IEEE Winter Conference on Applications of Computer Vision,* 1767–1775.

Xia, Y., Zhang, D., Kim, J., Nakayama, K., Zipser, K., & Whitney, D. (2018). Predicting driver attention in critical situations. *Asian Conference on Computer Vision,* 658–674. Springer, Cham.

Xu, H., Gao, Y., Yu, F., & Darrell, T. (2017). End-to-end learning of driving models from large-scale video datasets. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2174–2182.

Yildirim, F., Meyer, V., & Cornelissen, F. W. (2015). Eyes on crowding: Crowding is preserved when responding by eye and similarly affects identity and position accuracy. *Journal of Vision, 15*, 21.

Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., . . . Darrell, T. (2018). BDD100K: A diverse driving video database with scalable annotation tooling. arXiv preprint arXiv: 1805.04687.