3. Eyes on crowding: crowding is preserved when responding by eye


Abstract

Peripheral vision guides recognition and selection of targets for eye movements. Crowding – a decline in recognition performance that occurs when a potential target is surrounded by other – similar – objects influences peripheral object recognition. A recent model study suggests that crowding may be due to increased uncertainty about both the identity and the location of peripheral target objects, but very few studies have assessed these properties in tandem. Eye-tracking can integrally provide information on both the perceived identity and the position of a target, and therefore could become an important approach in crowding studies. However, recent reports suggest that around the moment of saccade preparation crowding may be significantly modified. If these effects were to generalize to regular crowding tasks, it would complicate the interpretation of results obtained with eye-tracking and the comparison to results obtained using manual responses. For this reason, we assessed whether the manner by which participants responded – manually or by eye – affects their performance. We find that neither recognition performance nor response time was affected by the response type. Hence, we conclude that crowding magnitude is preserved when observers respond by eye.

3.1. Introduction

In crowding, recognition of an object is impaired when it is surrounded by other – similar – objects. A recent model assumes that crowding is due to both excessive feature pooling (e.g Greenwood, Bex, & Dakin, 2010; Pelli & Tillman, 2008; van den Berg, Roerdink, & Cornelissen, 2010) and a loss of positional information (source confusion)
(e.g. Dakin, Cass, Greenwood, & Bex, 2010; Greenwood, Bex, & Dakin, 2009; Strasburger, Harvey, & Rentschler, 1991; Strasburger & Malania, 2013). According to this model, the uncertainty about both stimulus positions and identities depends on flanker proximity (van den Berg, Johnson, Martinez Anton, Schepers, & Cornelissen, 2012). To further test this idea, crowding studies would ideally assess target position and identity in tandem, but to our knowledge hardly any studies have done so thus far (a study by Greenwood, Bex, & Dakin (2012) forms a notable exception and we will discuss it later). This could be due to the difficulty of measuring the exact positional confusion via manual responses, which is the most common response type for crowding studies. Another reason could be that feature integration theory primarily considers the effect on recognition.

A very natural way to simultaneously assess what an observer saw and where s/he saw it is through using eye-tracking. Perceived identity can be inferred from the selected object, and the saccadic localization data can be used to evaluate perceived location. However, before we can proceed with this approach, we first need to verify whether eye movements themselves influence crowding in our type of task. This is assessed in this chapter.

Comparing crowding when responding manually or by eye

The possibility that eye movements influence crowding is suggested by several recent studies that assessed the relationship between crowding and eye movements (Harrison, Mattingley, & Remington, 2013a; Harrison, Retell, Remington, & Mattingley, 2013b). These studies suggest that just prior to making an eye movement, crowding may substantially change. If these effects were to generalize to other crowding tasks (e.g. tasks that require identifying and locating targets with one's eyes), this would complicate the interpretation of results obtained with eye-tracking. In fact, it would call into question the relevance of the majority of previous crowding studies for understanding natural vision (as these mostly used manual responses and required participants to fixate).

Harrison, Mattingley, et al. (2013a), reported that about 50 ms before a saccade is initiated towards a crowded object, the magnitude of crowding was reduced. Moreover, the spatial area within which crowding occurred was approximately halved. Harrison, Mattingley, et al. (2013a), concluded that "eye movement preparation enhances object discrimination in peripheral vision at the goal of the intended saccade. These presaccadic changes may enable enhanced recognition of visual objects in the periphery during active search of visually cluttered environments". Various mechanisms may underlie the reduced crowding observed by Harrison, Mattingley, et al. (2013a). As an explanation, they proposed that extra-retinal signals during saccade preparation stop the obligatory averaging of flanker and distractor features, thereby partially releasing the target from crowding. Two alternative explanations for the phenomenon were suggested by van Koningsbruggen & Buonocore (2013). Preparing an eye movement might lead to a shift of covert attention to the saccade target (H Deubel & Schneider, 1996) thereby enhancing discrimination performance at the location of the saccade goal (Heiner Deubel, 2008). To the extend that this influence is sufficiently selective, perception of the target might be enhanced more than that of the distractors, effectively reducing crowding magnitude. However, this explanation has been disputed by (Harrison,
Mattingley, & Remington, 2013c) because "observers knew the target's position and approximate timing in the no-eye movement and eye movement conditions so that attention could be allocated in the same manner on every trial". Second, performance might have been enhanced through saccadic unmasking (De Pisapia, Kaunitz, & Melcher, 2010; Hunt & Cavanagh, 2011). In this presaccadic effect, the target and its distractors - although presented at the same physical location – are being perceived at different spatial locations. As perceived –and not physical – position determines crowding (Dakin, Carlson, & Greenwood, 2011) this would effectively release a target from the crowding influence of its distractors. In another study, Harrison, Retell, et al. (2013b) reported a different perisaccadic phenomenon, which they referred to as "remapped crowding". Harrison, Retell, et al. (2013b) found that flankers flashed at the post-saccadic location of a target but prior to the actual saccade (and thus rather distant from the target) nevertheless affected the magnitude of crowding. Authors explained these results on the basis of a predictive remapping of receptive fields prior to saccades. Effectively, this phenomenon could increase the amount of crowding around saccadic eye movements.

There is one important reason to question whether presaccadic phenomena would be able to improve saccadic goal selection: the existence of a saccadic dead time. This is the brief period just prior to saccadic execution during which neither the execution of the saccade can be cancelled nor its goal can be changed (I. Hooge & Erkelens, 1996; I. T. Hooge, Beintema, & van den Berg, 1999). Hooge & Erkelens (1996), estimated that this saccadic dead-time starts approximately 70 ms prior to a saccade. Thus, there appears to be a substantial overlap between this estimate of saccadic dead time and the period in which the reduced crowding has been observed. De Vries, Hooge, Wiering, & Verstraten (2011) reported the presence of crowding in an eye movement-based visual search paradigm thus suggesting that crowding is still present in the presence of eye movements. However, they did not compare crowding magnitude with and without eye movements, so reduced crowding might still have been present. Note that this question has neither been answered by the studies of Harrison, Mattingley, et al. (2013a) or Harrison, Retell, et al. (2013b), as in order to measure crowding magnitude, observers made manual responses after having moved their eyes. Hence, the question of whether presaccadic phenomena can influence perception during natural tasks remains open.

Therefore, in this experiment we set out to assess the influence of responding by eye on crowding magnitude. To do so, we adapted the most common crowding paradigm – recognition of a single isolated or flanked object in peripheral vision – for use with eye movements. In our paradigm, observers are simultaneously shown an (isolated or flanked) object to the left and to the right sides of a fixation mark and respond by making an eye movement to the one perceived to be the target. Note that in essence, this paradigm mimics what the visual system has to do in most natural viewing behavior such as e.g. visual search: select a potential target – out of the usually many available in peripheral vision – for further scrutiny and plan a saccade to the its location. During manual responding, observers fixate and indicate the target location by pressing one of two buttons. Our hypothesis is that recognition performance in this crowding task does not change between eye and manual responses.
3.2. Methods

3.2.1. Overview
In the experiment, we measured the errors in discriminating between a target and a reference presented left and right of fixation. Target and reference could either be presented in isolation or be surrounded by four flankers. The target could be discerned from the reference on the basis of tilt: the observer’s task was to choose the most right tilted object. In different blocks of trials, observers indicated their response by 1) making a saccade to the perceived target or 2) pressing a left or right arrow key. Below the experiment is described in detail.

3.2.2. Observers
Eight observers (age range 20-49; 3 women) participated in the experiment. Authors FY and FWC were amongst the observers. The remaining observers were naïve as to the purpose of the experiment. All observers had normal or corrected to normal vision.

3.2.3. Materials
Observers viewed stimuli on a 22-inch CRT RGB monitor with a frame rate of 75 Hz (LaCie) from a distance of 59 cm. Stimulus presentation, eye movement recording, and response collection were programmed in Matlab (MathWorks) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and EyeLink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002). Eye movements were recorded at 250 Hz with an EyeLink 1000 (SR Research, Kanata, Ontario, Canada) infrared eye tracker. We used the EyeLink’s built-in 9-point calibration procedure. Background luminance during the experiment was 35 cd/m². We used a chin rest and a forehead rest to stabilize the observer’s head position.

3.2.4. Stimuli and procedure
Prior to the start of a trial, a white horizontal line (a minus sign; 0.2 deg) was presented at the center of the display. The observer pressed a key to commence the trial. Following the key press, the horizontal line changed into a fixation cross. 200 ms thereafter, the stimulus was presented for 200 ms. An example stimulus is shown in Figure 1. The stimulus consisted of a target and reference that were presented left and right of fixation, both either at 8 or at 10 deg of eccentricity, and either in isolation or surrounded by flankers. The side at which the target appeared was determined randomly. Target, reference and flankers were Gabor patches (width = 1.0 deg, spatial frequency 3.0 cycles/deg). Base target and reference tilt was set to 45°. To distinguish the target from the reference, it was tilted clockwise from base tilt by 5°, while the reference was tilted counterclockwise by 5°. In 20% of the trials, target and reference were presented in isolation. In the remaining 80% of the trials four flankers surround the target and reference. Flankers were positioned at the four corners of an invisible square with the target or reference at the center. Flanker tilt was either the same as the base tilt (45°) to create a high (difference 0°) or differed (by 45° or -45°) to create a low target-flanker similarity condition (see Figure 2 for an example). The 0° flanker tilt condition occurred twice as often as each of the 45° or -45° flanker tilt conditions.
Flankers were always presented at 25% contrast. Target and reference were presented at individually determined contrast and tilt thresholds (see below).

The observer’s task was to choose the most right-tilted target (the right one in the example of figure 1). In alternating blocks of trials, observers either indicated their response by making a saccade to the target or by pressing one of two keys. After the response, the fixation point turned either red (error) or green (correct) to provide feedback to the observer. During manual responding, observers were required to maintain steady fixation throughout the trials and their gaze was monitored. A single block of trials consisted of 260 trials. In their first session, observers first completed a 100-trial training block for each condition. Following these, observers completed two blocks of experimental trials in a row. They commenced with a block in which they responded manually, followed by a block in which they responded by eye. In each of their next three sessions, observers completed two more blocks of trials in which the order of eye and manual responding was alternated each time. All 8 observers thus completed 8 blocks of trials for a total of 1040 trials in each response mode.

![Figure 1. An example stimulus.](image)

In different blocks of trials, observers were instructed to indicate the most rightward-tilted target (central object) by either making a saccade or by pressing a button while maintaining fixation.

3.2.5. Individual contrast and tilt threshold determination

In a session preceding the experimental ones, we individually determined the tilt and contrast levels that enabled observers to achieve 80% correct recognition performance (manual responses) at each eccentricity for isolated targets and references. Observers were asked to report the more rightward tilted target. Target and reference tilt were set to \(45^\circ \pm 5^\circ\) respectively and were presented at either 8 or 10 degrees of eccentricity. In different trials, target and reference were presented at 13 different levels of contrast (1.0, 1.3, 1.7, 2.2, 2.9, 3.8, 5.0, 6.5, 8.5, 11.2, 14.6, 19.1, or 25.0 percent contrast). For the first two experimental sessions, we determined an individual contrast level for each observer that enabled them to achieve 80% correct performance. We did so by first fitting a cumulative normal distribution to a participant’s performance data as a function of contrast level. Next, using this distribution, we determined the threshold level of contrast. For the final two sessions, thresholds were determined slightly different. The contrast was fixed and the required tilt difference was determined. Contrast was set to the level to achieve 70% correct performance and next we determined the tilt level that enabled 80% correct performance. Target and references were presented at 13 different tilt values (difference from base tilt; 1.0°, 1.3°, 1.7°, 2.2°, 2.9°, 3.8°, 5.0°, 6.3°, 7.9°, 10.0°, 12.6°, 15.9°, or 20.0°). Again, a cumulative normal distribution was fitted to the performance data as a function of tilt, and this function was used to calculate the threshold tilt level. In the results section, the results for all sessions will be integrally
presented (after verifying that the type of threshold setting had no significant influence on the results).

Figure 2. Example stimuli. (a) low target-flanker similarity and (b) high target-flanker similarity.

3.2.6. Eye movement analysis
Saccades were determined using the EyeLink’s built-in analyses routines. Prior to entering the statistical analysis, eye movement responses were filtered based on saccadic amplitude, saccade latency and saccadic direction. Trials were removed in which saccades were either: i) made within 150 ms or after 3000 ms following the start of the stimulus presentation, or ii) in which saccadic direction differed more than 15 deg from horizontal (the direction of the target or reference), or iii) in which saccadic amplitude was less than 2/3 of the target or reference eccentricity. On average, this excluded about 8% of the eye-response trials. For the manual trials, responses that were accompanied by an eye movement were excluded from the analysis. On average, this excluded about 3% of the manual response trials.

3.2.7. Statistical analysis
We used repeated-measures ANOVA to determine whether differences in performance were statistically significant. Within-observer factors were response type (eye, manual), flanker mode (isolated target, low target-flanker similarity, high target-flanker similarity) and threshold (contrast, tilt). To test for differences in response time between manual and eye responses, first, a Gaussian was fit to the individual distributions, and the mean and standard deviations were determined. Paired t-tests were used to test for differences in these fitted parameters.

3.3. Results
As we used two slightly different ways to set the individual thresholds, we first determined whether the factor threshold results in significant performance differences. Repeated measures ANOVA indicated that this was not the case (F(1, 7)=1.1, p>.05). Interactions of this factor with the other factors were not significant either (response type F(1, 7)=2.7, p>.05; flanker mode F(2, 14)=.14, p>.05). For this reason, in the remainder of the analysis, this factor is not further considered.

Figure 3a-c shows the results for the individual observers, with different panels for the isolated target and the low and high target-flanker similarity conditions. Figure 3d shows the results averaged over observers. In each panel results are split by response mode (eye or manual). Not surprisingly, observers performed best when the target and
reference were presented in isolation (Figure 3a/3d). As expected, adding flankers decreased performance (Figure 3b-c, 3d). This decrease in performance was largest when target-flanker similarity was highest (Figure 3c, 3d). Overall, the small difference between the two different response modes was not significant (F(1, 7)=0.7, p>.05) and it showed no significant interaction with target-flanker similarity either (F(2, 14)=0.5, p>.05).

To examine the crowding effect in more detail, we calculated crowding magnitude by subtracting performance in the two flanker conditions from performance in the isolated target condition. This is a standard procedure to calculate the difference in performances for isolated and flanked conditions. The results are shown in figure 4. It is obvious that the results for the manual and the eye responses are nearly identical.

Figure 3: Discrimination performance of individual observers for manual and eye responses. Panel a shows the results for the isolated target condition, panel b for the low target-flanker similarity condition, and panel c for the high target-flanker similarity condition. Panel d shows the results averaged over observers. Bars in panel d indicate the standard error of the mean over observers.
Finally, we compared the time that it took observers to respond when making manual or eye responses. Figure 5 shows the response time histograms for both manual and eye responses. Response times and saccadic latencies were binned in 50 ms bins. As figure 5 shows, the histograms of response times for the manual and eye responses are very similar. Neither the means nor the standard deviations of individually fitted Gaussians differed between eye and manual responses (means: manual: 845 ms and eye: 892 ms; t(7)=-1.153, p>.05; standard deviations: manual: 68 ms and eye: 63 ms; t(7)=2.203, p>.05).
3.4. Discussion

In normal visual behavior, information from peripheral vision is used in saccade goal selection. In the light of the task at hand, the visual system uses saliency and other metrics of relevance to determine the priority with which peripheral locations need to be scrutinized by foveal vision (Fecteu & Munoz, 2006; Hayhoe & Ballard, 2005; Itti & Koch, 2001; Tatler, Hayhoe, Land, & Ballard, 2011; Yanulevskaya, Marsman, Cornelissen, & Geusebroek, 2011). Peripheral vision is also the part of the visual field that is most strongly affected by crowding (Bouma, 1970; Korte, 1923; Stuart & Burian, 1962; Toet & Levi, 1992).

If eye movement preparation would effectively change peripheral discrimination – and thereby saccade goal selection– this would profoundly affect our understanding of the relationship between peripheral vision and crowding in natural viewing. In our view, it thus seemed worthwhile to compare crowding magnitude when observers responded either manually, or used their eyes to respond. The main finding of this experiment is that crowding magnitude is preserved when observers use their eyes to respond in a crowding task, thus confirming our hypothesis. This implies that previously observed changes in crowding as a result of eye movement preparation (Harrison et al., 2013a,b) do not appear to change crowded object discrimination as measured in our – relatively standard – crowding paradigm. Therefore, we also conclude that eye tracking is a valid tool to assess crowding.

An important reason for not finding a difference was already suggested in the introduction. The temporal window of the observed enhancement and integration effects is such that they show a substantial overlap with saccadic dead time. During this brief period – just before saccade initiation – no changes to the saccade execution are possible anymore. The period of the saccadic dead time (approx. 70-100 ms) is very similar in size to the period during which changes in crowding were observed (approx. 50 ms prior to saccade initiation). It is important to note that our results do not necessarily demonstrate that eye movement preparation does not cause changes in crowding. However, our results do indicate that eye movement preparation does not affect discrimination performance in our paradigm.

We modeled our paradigm to mimic after what we believe the visual system has to do in most natural viewing behavior such as e.g. visual search: select a potential target – out of the usually many available in peripheral vision– for further scrutiny and plan a saccade to its location. One might object that a (forced) orientation discrimination task preceding the eye movement is not natural. Yet, planning an eye movement requires at the least a pre-attentive target selection process (e.g. based on saliency). Therefore, in our view, our paradigm with a forced selection based on orientation followed by an eye movement comes close to the natural situation.

Our present design – with both a target and a reference symmetrically presented left and right of fixation – guaranteed that there was no net advantage on discrimination performance of moving ones eyes early. As in natural viewing behavior, our paradigm requires observers to decide on the location of the target before programming the saccade. Therefore, one could argue that discrimination performance could not have been affected by saccade preparation. However, Allik, Toom, & Luuk (2003) showed that saccade direction and amplitude can be planned separately. In our paradigm observers
could plan saccade direction only after having made the decision about which side contained the target. However, they might have started planning the saccadic amplitude as soon as the stimulus was shown, giving this process a substantial head-start and also leaving sufficient time for any putative enhancement and integration effects to occur.

**FOOTNOTES**

¹ Note that the term ‘target-flanker difference’ refers to the difference in orientation between the base orientation of the target and the flanker orientation. In the experiments, the target and the reference were defined by a tilt offset relative to this base orientation. Hence, for instance when both target base tilt and flanker tilt were 45°, both the target and reference would still be 5° different from the flankers (e.g. 40° and 50°). For the sake of simplicity of the analyses, the target-flanker differences were grouped based on the difference between the base orientation of the target and the flanker orientation (so 0° in the particular case mentioned here).
3.5. References


42


