UNIVERSITY OF PADOVA

Department of General Psychology

Master Degree in Cognitive Neuroscience and Clinical Neuropsychology

Final dissertation

Comparing Crowding and Foveal Feedback mechanisms using a Vernier task

Supervisor:
Professor Marco Bertamini

Candidate: Roberta Cessa
Student ID number: 2034625

Academic Year 2022/2023

We see with our brains, not our eyes.

Paul Bach-y-Rita

Contents

Chapter 5

Discussion 65

Conclusion 73

References 75

Abstract

Feedback connections from higher to lower visual areas have shown to play a role in visual perception by modulating attention, simplifying stimulus perception, and providing more detailed representation of objects in the visual field. An fMRI study by Williams et al. (2008) showed that a representation of peripheral stimuli is available in the foveal retinotopic cortex during a shape discrimination tasks, suggesting a new form of feedback. Following investigations showed that this feedback process can be perturbed: when noise is presented in the foveal visual field within a delay from stimulus onset (~150 ms for some studies, 250 ms for others), performance is impaired. Recent evidence has revealed that the well-known mechanism of crowding also seems to have a critical timing in which target-flanker interactions take place. The strongest crowding effects can be observed when the flankers appear with a temporal delay between 100 and 150 ms after the onset of the target (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015). This U-shaped crowding function suggests strong similarities with what typically happens in metacontrast masking, where a mask can compromise the coding of a peripheral target when presented in the same spatial location but with a delay of about 50-150 ms (Chung, 2016; Chung & Patel, 2022; Di Lollo et al., 2000; Harrison & Bex, 2014; Huckauf & Heller, 2004). What do these predominantly peripheral visual mechanisms have in common? And how can we study them? In this work, we wanted to test the hypothesis that foveal feedback only occurs for high level objects by replicating the previous studies with low level stimuli and investigate the temporal and spatial properties of this phenomenon. A first experiment adapted the paradigm of previous studies to investigate if the same phenomenon would also apply to lower-level stimuli. In a second experiment we combined standard foveal feedback paradigms with crowding to better investigate their nature and the relationship between the two mechanisms. In the final and third experiment we tested the characteristics of the foveal noise to understand whether the disturbances could be caused by a simple, lower-level foveal distractor or if they are dependent of the noise characteristics. We observed a peak in Vernier acuity threshold between 50 and 150 ms after stimulus onset in the first two experiments, while the third one did not produce significant results. We propose that the effect of a foveal disruptor in a peripheral visual acuity task is task-dependent, and it cannot be induced by the simple presence of any type of delayed foveal object. These findings suggest a potential link to the concept of cortical reentrant processing (Di Lollo et al., 2000), with recurrent and mechanisms playing a pivotal role in object detection and recognition within precise temporal windows.

Chapter 1

Visual processing and feedback mechanisms

1.1   The relationship between foveal and peripheral visual processing

In the classical view of visual processing, the foveal and the peripheral areas of the visual field have been considered to work independently from each other. This assumption arose from the anatomical and physiological differences in organization present in the human retina and visual pathways. The retina consists of separate regions (Figure 1.1) that can be distinguished by different distribution and connectivity patterns of rods and cones, the photoreceptors.
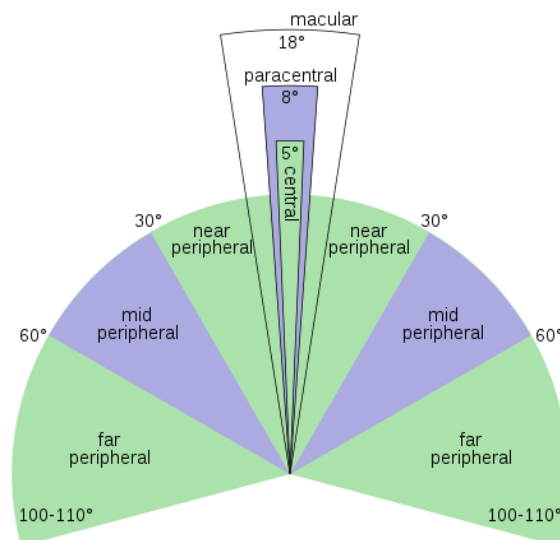
Figure 1.1: The illustration depicts the distinct divisions of the human visual field. The fovea, referred to as the "central" region, spans a range of 5 degrees and is in immediate proximity to the parafovea, the "paracentral" region, which encompasses 8 degrees of the visual field. These sections are both part of the macular area. The remaining portion of the visual field falls within the periphery. Image from Wikimedia Commons, author Zyxwv99 (https://commons.wikimedia.org/wiki/File:Peripheral_vision.svg).

In the center, the fovea[1] extends for about 1 degree of visual angle (0.3 mm) and presents a significant concentration of cones and a total absence of rods. Furthermore, the connections that the photoreceptors in the fovea have with cells in the subsequent layers are one-to-one, leading to smaller receptive fields in the ganglion cells and to an over-

---

[1] Stewart et al. (2020) noted that the term "fovea" is associated with different definitions and sizes. Here, I will refer specifically to the fovea as the central portion of the visual field, while the periphery will be intended as the encompassing area.

representation of the fovea that is maintained throughout all the following stages of visual processing, up to the visual cortex. The area found 1.3 – 1.4 mm away from the foveal center is known as parafovea. Here, rods can be found, cone density starts reducing and all photoreceptors diameter increases (Curcio et al., 1990). The surrounding region is referred to as periphery, where rods predominate and connections with the underlying ganglion cells are mostly more-to-one.

The way in which the architecture of the visual system affects visual perception is an indication that foveal and peripheral vision carry out different functions. Foveal vision is responsible for detailed processing of stimuli and enables high spatial resolution, acuity, and contrast sensitivity. On the other hand, vision processing in the periphery lacks details and information about position and shape of objects, even though it receives most of the information that appears in the visual field. Furthermore, peripheral vision is particularly sensitive to crowding, resulting in challenges in visual discrimination tasks due to inhibitory interactions between the contours of neighboring objects (Levi, 2008).

There is general agreement that foveal and peripheral vision play separate roles in the visual perceptual system. Yet, our perception of the visual environment does not consist of a blurred area surrounding a distinct focal point of attention, but rather of a smooth and cohesive whole (e.g., Martinez-Conde et al., 2004; Treisman & Gelade, 1980). This is the result of the integration processes across different sources that our brain is constantly engaged in. Emerging findings indicate a close interconnection between the fovea and the peripheral regions of the human retina, challenging the previous assumption of their independence (for a review, see Stewart et al., 2020).

1.2   The role of feedback in the predictive processing framework

The classical models of perception consider visual processing as a mainly feed-forward mechanism, implying that visual information coding occurs in a unidirectional manner, with minimal influence from higher-level areas on lower-level areas. According to this perspective, visual inputs are processed via a hierarchy of cortical regions that reflects the functional organization of the ventral visual stream (Riesenhuber & Poggio, 1999).

In the initial stages (primary visual cortex and extrastriate areas) basic features of the visual input such as shape, orientation, motion direction and colors are detected. As we move beyond the extrastriate cortices, associative areas generate increasingly sophisticated representations. Examples are the lateral occipital complex (LOC), that has

a central role in object recognition and can detect shapes independently of low-level visual features (Grill-Spector et al., 2001; Kourtzi & Kanwisher, 2001). Yet, the abundance of projections from higher- to lower-level areas in the visual system makes the hypothesis of an exclusively feed-forward processing too simplistic.

Emerging evidence supports the idea that perception is not exclusively stimulus-driven and that feedback mechanisms play an essential role in human visual perception (Panichello et al., 2013). Most theories about the role of visual feedback mechanisms agree that they primarily serve predictive and facilitatory purposes. This notion becomes particularly relevant when dealing with ambiguous visual information, as feedback mechanisms help with minimizing the possibility of errors and selecting the final conscious representation, based upon prior knowledge and contextual information (Panichello et al., 2013; Teufel & Fletcher, 2020). According to this perspective, feedback plays a crucial role in guiding conscious perception, optimizing lower-level activity, and maximizing the generation of reliable and coherent representations of reality. This theory is established within the framework of predictive coding, which posits that we possess an internal model of the surrounding world that can be employed to anticipate incoming information (Rao & Ballard, 1999).

In the context of visual perception, evidence indicates that neurons in certain visual cortices (e.g., V1, V2, and V4) cease to respond to optimal stimuli extending beyond their classical receptive field (RF) boundaries, as observed in natural images. Rao & Ballard (1999) proposed that this phenomenon is linked to predictive coding. The brain adapts to the regularities present in the visual environment, actively avoiding processing of redundant and predictable information. Neurons possessing extra-classical RF properties respond to the presence of inconsistencies between the expected and perceived visual information, working as predictive estimators (Figure 1.2).

In the context of shape recognition, psychophysiological evidence has highlighted how feedback mechanisms can enhance the resolution and the level of detail of a visual representation in lower-level areas (Jehee et al., 2007). Evidence in monkeys also showed that in figure-ground segmentation tasks feedback connections can amplify the response to optimal stimuli for the receptive fields center, making the salient figure 'pop-out' from the background and suppressing processing of non-relevant contextual elements (Hupé et al., 1998; Klink et al., 2017).

Feedback exhibits an organization that appears to be feature- or stream-specific, which aligns with the goal of adjusting the timing and precision of the feedforward information (Briggs, 2020). Moreover, the literature indicates that feedback and feedforward visual information both exhibit a retinotopic organization (Briggs, 2020; Marques et al., 2018; Wang et al., 2022). This implies that the visual input is processed in the areas of the retina responsible for perceiving that particular input, as well as in the following stages that are interconnected with those specific retinal regions.

Over the past decades, robust research findings have established the modulatory role of feedback in shaping our perception. This has led to significant progress in understanding the intricate network of feedback connections within our brain, which outnumbers even the feedforward connections. Feedback is now recognized as an essential factor in conscious perception, and it does so without undermining the hypothesis of a hierarchical organization of the visual cortex. Instead, feedback optimizes the activation of cortical areas and the resulting processing of visual information, often following a retinotopic and feature-specific logic.

Despite considerable progress, there is still much to be uncovered about feedback mechanisms. Interestingly, fMRI findings from Williams et al. (2008) propose that feedback might serve novel functions and operate in ways that challenge our current assumptions.
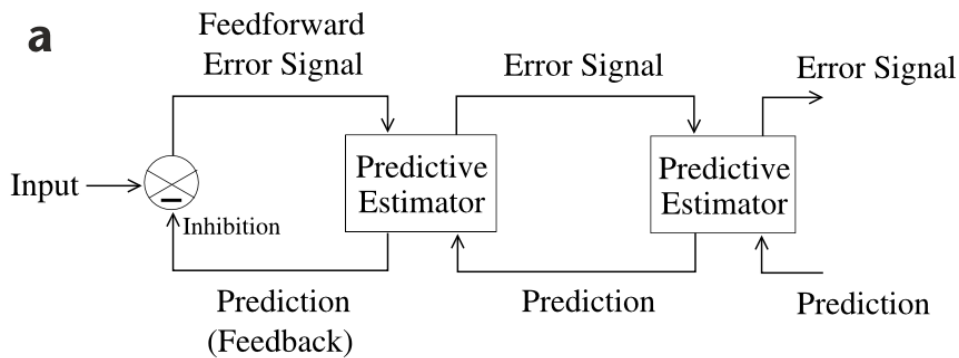
Figure 1.2: General architecture of the hierarchical predictive coding model. Feedback pathways convey predictions of neural activity at each level, while feedforward pathways transmit errors between predictions and actual activity. The predictive estimator (PE) uses these errors to adjust its estimation and generate predictions (Rao & Ballard, 1999).

1.3   Recent findings about foveal feedback

The pioneering study conducted by Williams et al. (2008) revealed the presence of unexpected feedback mechanisms within the foveal retinotopic cortex, where they

detected a definite reconstruction of stimuli originally presented in the peripheral visual field. This discovery suggested that feedback might possess constructive properties that were previously unexplored. Initial evidence emerged from neuroimaging and neurostimulation studies, and over the last decade, behavioral findings have further supported this new theory.

1.3.1 Neuroimaging results: fMRI data

The study by Williams et al. (2008) was initially designed to investigate how position-invariant objects are represented in higher-level cortical areas. Participants were asked to judge couples of novel three-dimensional objects and to determine whether the two items were identical or different with a same-different task (Figure 1.3). The targets were presented in the periphery of the visual field, while the subjects maintained fixation on a central dot throughout the entire duration of the experiment. They then analyzed the pattern of responses from functional magnetic resonance imaging (fMRI) data, which revealed that a representation of peripheral objects presented more than 5 degrees away from the fovea is available in the foveal retinotopic cortex.

For the first time, a representation of visual stimuli was found in a cortical area that is not involved in its feedforward processing. This representation was position-invariant, indicating that the pattern observed in the foveal cortex was unaffected by the original position of the stimulus in the periphery, and behaviorally relevant, being exclusively observed during object discrimination tasks and not, for example, in color discrimination tasks with the same stimuli. Moreover, the strength of this representation showed a correlation with task performance, with a stronger representation being associated with improved task results.
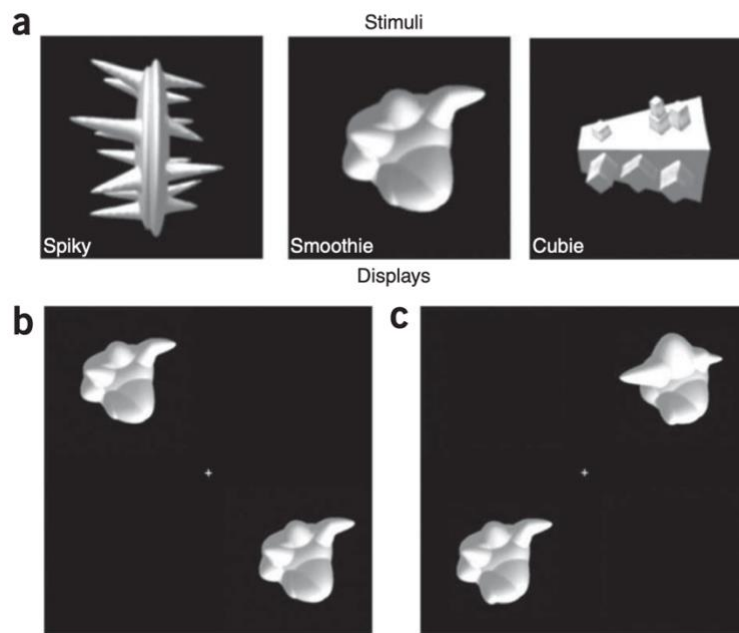
Figure 1.3: Exemplars of stimuli employed in the experiment by Williams et al. (2008). (a) Typical stimuli from each of the three categories available: spiky, smoothie and cubie (retrieved from Op De Beeck et al., 2006). (b) Example of presentation display with two identical 'smoothies' positioned along the left diagonal of the screen. (c) Example of presentation display with two identical 'smoothies' positioned along the right diagonal of the screen.

The authors ruled out the possibility of the peripheral information in the foveal cortex being influenced by eye movements or foveation errors across objects. Furthermore, they excluded the potential misattribution of activation to the foveal region of interest (ROI) when it might have actually originated from peripheral areas or the lateral occipital complex (LOC). This was supported by the fact that the peripheral cortex lacks position invariance and the LOC exhibits task-independence, making them incompatible with the pattern observed in fMRI data. Given these constraints, the authors proposed that the observed effect results from higher-level cortical areas feeding back object information to the lower-level foveal cortex.

1.3.2 Neurostimulation results: TMS data

Three years later, Chambers et al. (2013) extended the research on "foveal feedback" to incorporate transcranial magnetic stimulation (TMS) techniques, to further investigate the role of reconstructed foveal information in extra-foveal areas while also excluding the possibility that it was the mere result of activity occurring elsewhere. TMS is known for its ability to modify the excitability of the targeted cortical areas and induce short-lasting

disruptive effects on the normal functioning of such areas, especially at lower frequencies (Kobayashi & Pascual-Leone, 2003).

The main aim of Chambers et al.'s study (2013) was to determine the critical time window during which foveal feedback occurs. This was done to rule out the influence of known explanations (e.g., attentive processes) and to ultimately provide causal evidence regarding the mechanisms responsible for this constructive effect. This was achieved by observing the impact of transitory deactivation of the foveal cortex on performance in a same-different task. They asked participants to perform the same task as in Williams et al. (2008), but they combined it with a 50ms double TMS pulse (Figure 1.4). The pulse was administered to the posterior termination of the calcarine sulcus, where the foveal cortex is, and to non-calcarine areas 15 mm away from the foveal area, as control sites. TMS was delivered at different timings before and after the onset of the target stimulus (stimulus onset asynchronies, SOAs) ranging from -150 to +500 ms, and with varying degrees of intensity.
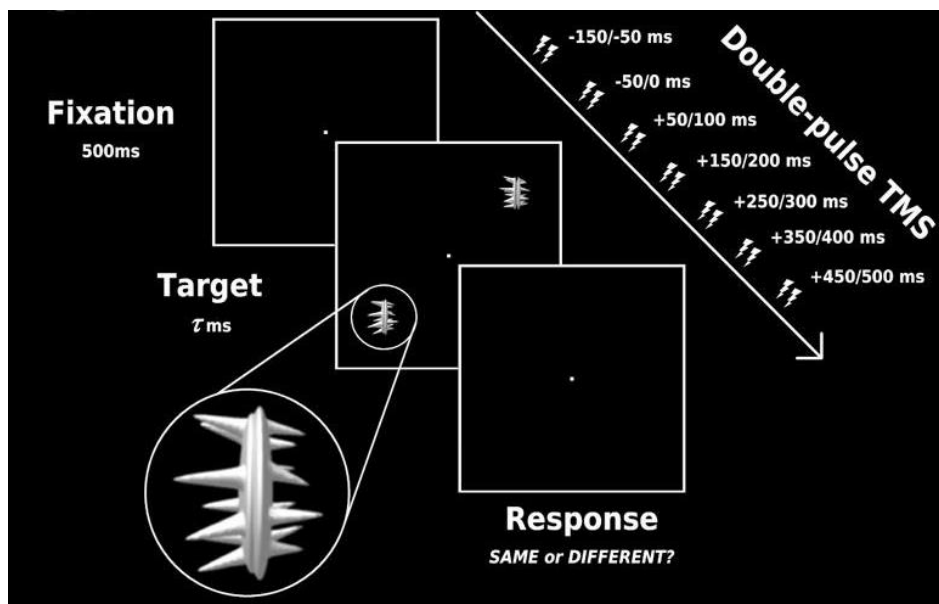
Figure 1.4: Schematic of the experimental paradigm by Chambers et al. (2013). Two objects from the same abstract category of "spikies" from the original experiment (Williams et al., 2008) were presented at opposite diagonal locations. Participants had to determine whether the objects appeared identical or different. To investigate the effects, a double-pulse TMS was administered at 7 different SOAs on the posterior calcarine site.

A drop in performance was observed when the pulse was delivered at +350/400 ms SOAs over the calcarine area. The effect occurred too late to be explained by feedforward processing. It also could not be attributed to attentional priming, as according to this notion the pulse administered before the target's onset should have resulted in the

most significant disruption of performance. In this case, they did not observe any effect at negative SOAs. The use of TMS provided further and causal evidence of the crucial involvement of the foveal cortex in utilizing feedback signals from later visual areas, where information from the peripheral vision is stored, to carry out object recognition tasks.

1.3.3 Behavioral results

After the first two studies by Williams et al. (2008) and Chambers et al. (2013), several studies tried to obtain behavioral evidence of foveal disruption in object recognition tasks. This was unusual, as behavioral evidence typically precedes investigations into the neurobiological aspects, which had already been explored in this case. Most of these studies used a brief display of an incongruent foveal distractor or mask to cause the disruption, in order to interfere with foveal activity and stop the feedback from working. The visual mask should be administered and anticipated to become effective with different SOA to produce a similar disruptive impact to the TMS pulse. Visual stimuli presented in the fovea require between 40 and 120 ms to reach the relevant cortical areas, whereas a TMS pulse exerts an immediate modulatory effect on the activated areas it targets (Foxe & Simpson, 2002; Kammer, 2007; Oletto et al., 2022). The main aim of the behavioral investigations was, indeed, to determine this effective timing.

Weldon et al. (2016) used the same stimuli and paradigm as in Williams et al. (2008). The task involved a same-different comparison between two stimuli from the same novel category ("spiky", "cubie" and "smoothie") in the periphery. Additionally, they introduced a stimulus from one of the other two categories as the non-congruent foveal distractor. They examined five different SOAs: two SOAs with the mask presented before the stimulus onset (-267 ms and -117 ms), two with the mask displayed after the stimulus onset (+117 ms and +276 ms), and one with the mask appearing simultaneously with the targets (0 ms). In a second experiment the distractor was presented in non-foveal locations of the visual field. The study revealed that the feedback effect is strongly dependent on time and location: discriminability was only impaired when the distractor appeared at +117 ms after stimulus onset and when it was displayed at fixation.

In a later study, Weldon et al. (2020) further explored whether foveal disruption could affect color perception. Using the same experimental design as in their previous study, participants were now tasked with a color discrimination task. They also examined

performance changes in color discrimination using simpler low-level stimuli, such as homogeneous color patches. Results revealed a significant effect of the central distractor on color discrimination, but only when the peripheral stimuli contained complex shape information. This finding suggests that foveal feedback is comes into play specifically when fine object discrimination is needed.

In the same year, Yu & Shim (2016) proposed a slightly different task. Using the same stimuli as in previous studies, they asked participants to determine whether a stimulus presented in the periphery matched a previously shown target. Further, they manipulated the foveal noise, alternating it with an object from another category, the target itself and a scrambled object. The duration for which the noise was displayed was also reduced compared to previous studies, now set at 33 ms. These distractors were presented in the fovea either simultaneously with the target (0 ms SOA) or at a subconscious delay of +150 ms after stimulus onset. Moreover, their perception was rendered unconscious though backward masking, whereby they were followed by a mask appearing in the exact same spatial location briefly after the stimulus onset. The results revealed that peripheral object discrimination was facilitated when the foil was a higher-level visual object identical to the target. This finding aligned with earlier studies and was specifically observed in the foveal region, but only after some time had passed since the presentation of the target.

Fan et al. (2016) delved deeper into exploring the potential SOAs at which the foveal feedback could be effective. In the leading experiment, participants engaged in the same peripheral object discrimination task from the study by Williams et al. (2008). The stimuli closely resembled those from Op De Beeck et al. (2006) "spiky" category. To assess the effects, a colored dynamic noise was displayed at fixation at with different SOAs: 50, 150, 250, 350 and 450 ms, along with a no-noise condition. The data revealed that the performance was significantly impaired 50 ms and 250 ms after stimulus onset (as shown in Figure 1.5). To explain these dips in performance, the researchers proposed two complementary explanations. The performance drop at 250 ms was seen as consistent with prior findings on the foveal feedback effect and was interpreted as the time window during which the mechanism is effective. On the other hand, the performance drop at 50 ms was attributed to an attentional distraction caused by the partial overlap in time between the noise patch and the target stimuli.
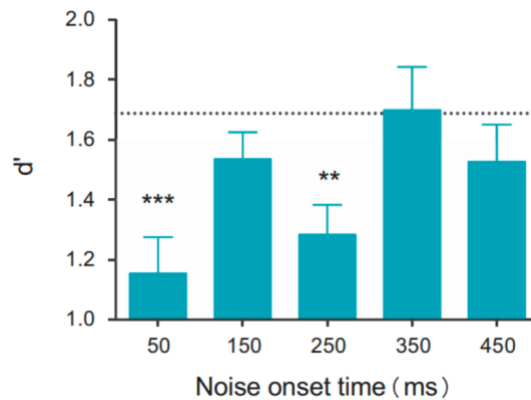
Figure 1.5: Pattern of results from the first experiment by Fan et al. (2016). The d prime (d') values revealed two dips, respectively at 50 ms and 250 ms after stimulus onset.

A recent study by Contemori et al. (2022) extended the first experiment by Fan et al. (2016) with the purpose of better defining the temporal patterns of foveal feedback. To do so, they increased the number of SOAs of the visual mask to 60 and tested the effects of a foveal noise from zero to half a second every 8.33 ms after the targets were displayed. A quartic model revealed a global minimum in d' when the mask appeared 94 ms after stimulus onset (as shown in Figure 1.6), slightly earlier than in the original study by Fan et al. (2016).
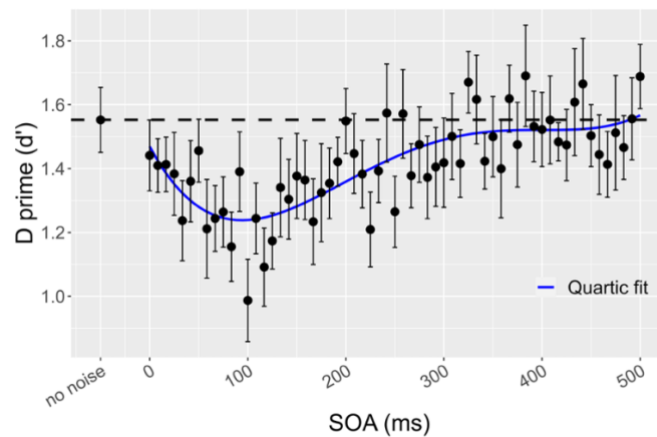
Figure 1.6: d' as a function of SOAs in Contemori et al. (2022). The dashed line indicates the d' in the baseline no-noise condition. Each dot indicated mean d' values for each of the 60 SOAs. The blue line represents the quartic fit.

Finally, Ramezani et al. (2019) explored the impact of foveal disruption on the processing of higher-level visual features. They designed an object categorization task involving natural object images at varying abstraction levels (superordinate, basic, subordinate). The researchers then manipulated the foveal representation of these images using a delayed foveal noise. Their findings revealed that peripheral vision has accurate

superordinate categorization abilities. A decline in performance was observed when a foveal noise appeared with a 300 ms delay, but only when dealing with objects from the basic and subordinate levels. These results provide further evidence of the role that foveal feedback plays in processing more detailed object information at intermediate abstraction levels.

Table 1 offers a comprehensive schematization of the studies conducted about foveal feedback so far. The table focuses on differences in the manipulation of foveal vision, the type of tasks utilized, and the timing of the observed performance dip.

Experiment	Manipulation (duration)	Timing of the Dip	Type of Task
Chambers et al. (2013)	TMS	350 - 400 ms	same/different
Contemori et al. (2022)	foveal mask (83 ms)	94 ms	same/different
Fan et al. (2016)	foveal mask (83 ms)	250 ms	same/different
Fan et al. (2016) – blurred object	foveal mask (83 ms)	no dip	same/different
Fan et al. (2016) – mental rotation	foveal mask (83 ms)	450–550 ms	same/different
Fan et al. (2016) – single object	foveal mask (83 ms)	150 ms	same/different
Ramezani et al. (2019)	foveal mask (100 ms)	300 ms	categorization
Weldon et al. (2016)	foveal mask (117 ms)	117 ms	same/different
Weldon et al. (2020) – color patches	foveal mask (117 ms)	117 ms	same/different
Weldon et al. (2020) – colored objects	foveal mask (117 ms)	117 ms	same/different
Williams et al. (2008)	no	-	same/different
Yu and Shim (2016)	foveal mask (33 ms)	150 ms	target recognition

Table 1. Table adapted from Oletto et al. (2022). The information included reports the type of manipulation of the foveal cortex, the timing of the observed dip in performance accuracy (if present) and the type of task administered to the participants.

In summary, the behavioral studies conducted in the past decade have yielded evidence supporting the crucial role of the foveal cortex in peripheral visual processing. This influence is supported by the fact that peripheral object recognition and categorization can be perturbed by the brief presentation of a foveal delayed noise. This should compromise the feedback of information from higher- to lower-level cortical areas, negatively affecting performance accuracy. These studies have shed light on the time window during which this feedback signal occurs, though some variability still exists. The temporal dynamics seem to range from approximately 100-150 ms (Contemori et al., 2022; Fan et al., 2016; Q. Yu & Shim, 2016) to around 300 ms after the target onset (Ramezani et al., 2019).

1.4   Theoretical interpretations of foveal feedback

The existing literature on feedback signals to the foveal retinotopic cortex from higher-level cortical areas has provided compelling evidence of this intriguing constructive role of feedback. However, the purpose of this mechanism remains a central question: why is foveal vision so critical for highly detailed object recognition in the periphery? The authors have put forth various reasonable explanations to shed light on this phenomenon.

The scratchpad hypothesis Williams et al. (2008) first suggested that foveal feedback could play a crucial role in supporting performance on complex and detailed object discrimination tasks. According to this interpretation, the foveal retinotopic cortex may serve as a high-resolution "scratchpad", storing task-relevant information that is later recruited to optimize performance in the periphery, where visual acuity and resolution are particularly coarse and limited. This aligns with Lee et al.'s (1998) original view of the primary visual cortex (V1), proposed as an alternative to the classical model for image meaning computation. According to the authors, V1 efficiency benefits from intracortical feedback connections at different stages of visual processing. This theory is also consistent with the well-established idea of the pivotal role that recurrent processing plays in shaping the conscious experience of an object (Lamme & Roelfsema, 2000).

A link between the scratchpad hypothesis of foveal cortex and Baddeley's visuospatial sketchpad has also been proposed (Contemori et al., 2022; Oletto et al., 2022). The visuospatial sketchpad is considered responsible for storing visuo-spatial information, which can be later employed to manipulate and revisit task-relevant visual

images (Baddeley, 1988; Baddeley & Hitch, 1974). This is further supported by Contemori et al.'s (2022) finding about a shift in response criterion when the foveal mask was presented at 150 ms SOA, where the performance dip occurred. Participants displayed a more conservative approach in their responses, leaning towards answering "different" more frequently. This observation aligns with the sketchpad hypothesis, suggesting that the foveal mask hinders an accurate reconstruction of the information stored during feedforward processing. However, it is essential to note that while this build-up hypothesis presents a compelling perspective, it should be regarded with caution as it doesn't provide definitive evidence, and few studies alone cannot conclusively confirm this notion.

Predictive remapping theory Some authors have considered predictive remapping as a plausible explanation for foveal feedback mechanisms (Fan et al., 2016; Ramezani et al., 2019). According to this idea, our natural tendency to foveate peripheral task-relevant objects could lead to increased activity in visual cortical areas in anticipation of the upcoming foveation (Melcher, 2007; Rolfs et al., 2011). Ramezani et al.'s (2019) findings seemed to support this notion, as a saccade was required to perceive the details of the natural images used in their paradigm. Fan et al. (2016) explored this hypothesis in a modified version of their original study, by allowing participants to plan a saccade away from the target stimuli. In this case, foveal disruption did not have the expected effect. Based on the predictive remapping view, no significant changes in performance should be detected as an observer should be able to predict the upcoming peripheral stimulus from the very first trial. As shown by Williams et al. (2008) this is not the case since they observed instead a progressive build up in behavioral performance over the course of each experimental block. This cannot be explained by predictive coding, but is rather more consistent with the idea that the foveal retinotopic cortex works as a "scratch pad", storing task-relevant information that is gradually updated and can lead to performance improvement as the experiment progresses (Slotnick et al., 2005; Williams et al., 2008). As a result, this theory remains a subject of debate and might not be the most reliable explanation for foveal feedback mechanisms.

Reverse hierarchy theory Foveal feedback has also been associated with the reverse hierarchy theory (RHT, Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002) which posits that in shape recognition, implicit feedforward connections initially provide basic

stimulus characteristics stored as high-level object and category representation (Figure 1.7). Shortly after, a feedback signal is sent from lower-level areas to add details and fine parameters to the stimulus, enabling a more thorough perception. In this context, foveal feedback could be implicated in achieving perceptual accuracy by providing fine stimulus information to the extrafoveal cortical areas. However, it still differs from Hochstein & Ahissar's (2002) theory in some respects. Foveal feedback is not associated with learning but rather with real-time object discrimination, it exhibits position invariance and is not consistently effective, unlike what the RHT might predict. Instead, it occurs only within a specific short time window, making this explanation only partially valid (Contemori et al., 2022).
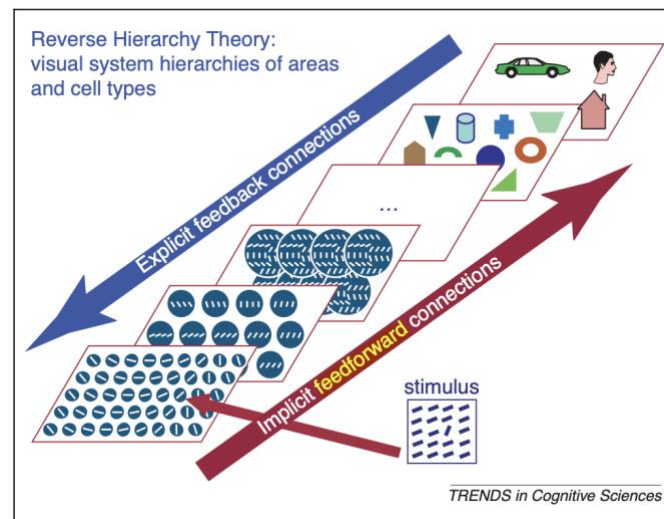
Figure 1.7: Reverse Hierarchy Theory. Higher-level areas first detect basic visual features of the task-relevant stimulus through feedforward processing. Later, explicit feedback connections from lower-level areas join by adding simple feature details to improve perceptual discrimination. Image from Hochstein & Ahissar (2002).

1.5   Conclusions

In summary, over the past years, consistent neuroimaging and behavioral studies have revealed that in tasks involving fine and complex object recognition (Chambers et al., 2013; Contemori et al., 2022; Fan et al., 2016; Weldon et al., 2016; Q. Yu & Shim, 2016) and categorization (Ramezani et al., 2019), peripheral vision alone lacks the precision to accurately process detailed information, and it appears to require support from the foveal retinotopic cortex. As a result, a new role for feedback processes has been proposed, suggesting that they can generate a reconstruction in a cortical area different from the one where, according to the retinotopic framework, the stimulus should be projected. Specifically, this reconstruction can be observed in the foveal retinotopic cortex,

replicating what was initially presented in the periphery of the visual field, starting from around 6.5 degrees away from the fovea.

This feedback signal can be perturbed either by applying a TMS pulse to the calcarine area, where the foveal retinotopic cortex is situated, or by introducing a delayed foveal noise, while presenting the target stimuli in the peripheral visual field. According to the foveal feedback hypothesis, when behavioral foveal disruption occurs, it leads to the simultaneous transmission of two different pieces of information to the foveal retinotopic cortex: the top-down peripheral object representation and the noisy signal from the fovea. This mismatch results in a decline in the accuracy of target processing since the foveal cortex is unable to supply the required detailed information about object shape, causing the representation to remain coarse and approximate.

While the search for the most convincing and unambiguous explanation continues, it seems evident that foveal feedback is the consequence of a built-in mechanism within the human visual system. This mechanism involves peripheral feedback that primes the foveal cortex, enabling more effective processing of visual objects that would otherwise result in inaccurate and superficial recognition tasks (Ramezani et al., 2019). As researchers delve deeper into this intricate interplay between foveal and peripheral vision, we have gained precious insights into the sophisticated mechanisms underlying our visual perception and object recognition abilities. The search for an in-depth understanding of foveal feedback mechanisms is still ongoing, but it promises exciting new perspectives in the fascinating field of visual cognition.

Chapter 2

Crowding and its effects on object recognition

## 2.1 Crowding mechanisms and object recognition

Crowding is another visual phenomenon central to this thesis project. It refers to a highly specific and powerful effect that is commonly observed in everyday life settings, with its most disruptive consequences when occurring in the periphery of the visual field.

In crowding, the perception of a peripheral target is strongly deteriorated by surrounding elements, referred to as flankers. The interactions between target and flankers are characterized by some distinctive features as (i) producing deleterious effects on object recognition tasks, (ii) being locally confined to a precise spatial window (Bouma, 1970) and (iii) requiring feature-specificity between target and flankers. However, as will be address in the next paragraphs, the tenets that have driven most of the research in crowding so far have been shown to be only partially true (Herzog et al., 2015; Herzog & Manassi, 2015). Crowding is, in fact, far more complex and variable than initially thought.

Crowding has been extensively studied over the last century. The paper that provides a first description of the phenomenon is the one from Korte (1923), who wrote about crowing in the following terms: ''It is as if there is a pressure on both sides of the word that tends to compress it. Then the stronger, i.e. the more salient or dominant letters, are preserved and they 'squash' the weaker, i.e. the less salient letters, between them.'' (translation from German by Uta Wolfe appeared in Pelli et al., 2004).

The initial studies on crowding were primarily focused on reading and recognizing letters under cluttered conditions (Bouma, 1970; Toet & Levi, 1992) (Figure 2.1). As the research progressed, it further extended to object recognition in broader contexts, investigating effects of crowing on multiple different types of stimuli like Vernier acuity (Levi et al., 1985; Malania et al., 2007), Gabor patches (Livne & Sagi, 2011; Saarela et al., 2009), and even more complex items, such as faces (Farzin et al., 2009; Martelli et al., 2005). In crowding, what is affected is not the detection of the target itself. This process remains relatively unaffected, allowing us to perceive the presence of the stimulus

clearly. However, crowding impairs the identification of the features of such stimulus (Pelli et al., 2004).
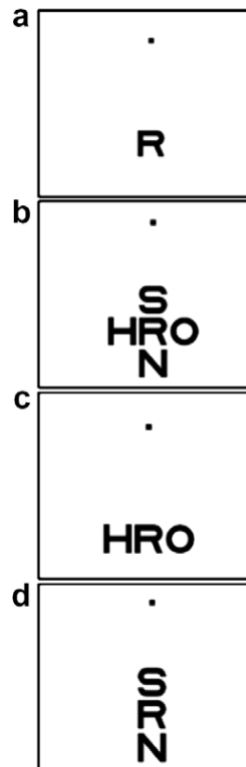
Figure 2.1: Straightforward illustration of crowding involving letters. Crowding effects be perceived when focusing on the dot in the upper part of the squares, while attempting to recognize the target letter 'R.' In scenario (a), the target stands alone and is readily identifiable; however, in (b), the target is surrounded by four flanking letters. In (c), the target is surrounded by two flanking letters placed horizontally, and in (d), it is surrounded by two flanking letters placed vertically. Crowding is present in all the examples except for (a), indicating that it is a robust and easily perceivable phenomenon. Image from Levi (2008).

The classic explanation of crowding is based on low level mechanisms, within the standard feed-forward hierarchal model of vision. In pooling models, neurons in lower areas, coding for basic features such as orientation and contrast, perform a first level of processing with high precision due to their small receptive fields. Information is lost at higher levels of the visual hierarchy, when neurons with larger receptive fields pool features of close-by elements (Parkes et al., 2001; Pelli, 2008; Wilkinson et al., 1997). Recognition of the target is impaired due to averaging or merging of flankers and targets features, suggesting that crowding is a "bottleneck of vision" and part of the limitations of our visual system (Levi, 2008). Other theories explain crowding by lateral interactions, which is neurons responding to flankers that affect the contrast level needed to detect the target by inhibiting neighboring neurons (Wilkinson et al., 1997), or by substitution, i.e., when the distractor is erroneously confused with the target (Ester et al., 2014). Being

proven to be a standard and ubiquitous phenomenon in real life situations, its understanding became fundamental to unravel the underlying mechanisms of human perception and behavior.

2.2 Spatial properties of crowding: comparing the classical framework with new perspectives

Based on local interactions of nearby elements, all the aforementioned models (Levi, 2008; Wilkinson et al., 1997; Ester et al., 2014) agree with the assumption that flankers can only affect the perception of the target if they are placed within half its eccentricity (Figure 2.2). This gap of potential interaction is called Bouma's window (Bouma, 1970). According to this interpretation, elements placed outside this window should not have any impact on how the target is perceived. Hence, the global configuration of the stimulus should not matter and, most importantly, adding more flankers should only worsen target perception, since it increases the energy of the disturbing factors or noise.
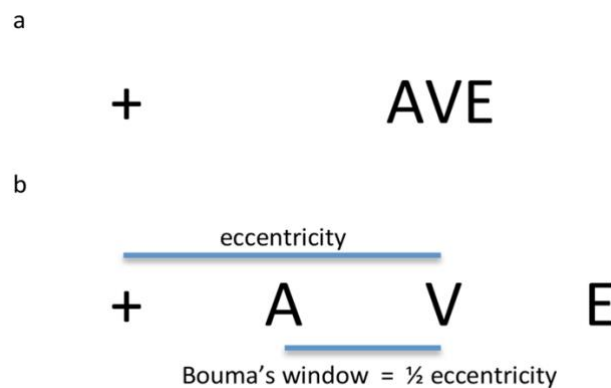
Figure 2.2: When fixating the cross, the flanking letters deteriorate the recognition of the target letter V. This can happen because the flankers are located within the crowing area, as determined by Bouma's law (a). The flanking letters are too far from the target and outside Bouma's window. In this case, target recognition is not affected by the flankers and crowding does not occur (b). Image retrieved and adapted from Doerig et al. (2019).

2.2.1 When more is better: beyond Bouma's law

Recent evidence has indicated that object recognition and crowding can be influenced by elements situated beyond Bouma's window (for reviews, see Herzog et al., 2015; Herzog & Manassi, 2015). Researchers have observed opposite effects of crowding when additional flankers stick out the classical crowding window. For instance, Vickery et al. (2009) found increased crowding effects when flankers were presented simultaneously with a masked target (Figure 2.3). This phenomenon, referred to as "supercrowding," exhibited more than just a combination of masking and crowding effects. Instead, it

displayed an over-additive outcome, where flankers located well beyond the critical crowding window significantly impacted target discrimination, with maximum effects observed at 0.7 eccentricity.
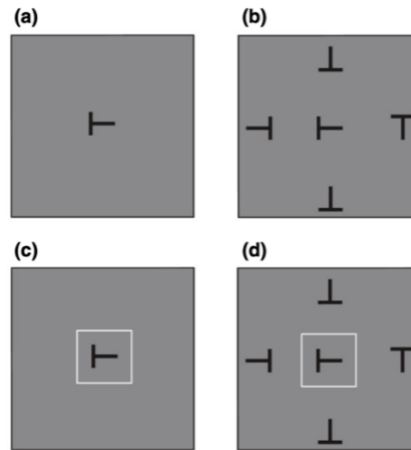
Figure 2.3: In Vickery et al. (2009), participants were asked to indicate the orientation of a T (a). As expected, when the flankers were presented outside Bouma's window, crowding did not occur (b). Performance was unaffected when the target was enclosed in a white square (c). The combination of condition b and c resulted in the most pronounced impact of crowding on performance (d).

On the other hand, evidence has also been found for a release of crowding as elements are added outside Bouma's window. One case is that from Manassi et al. (2013), who observed *un*crowding when more flankers were added next to the ones closest to the target (Figure 2.4). This has been explained by the authors in terms of grouping by different Gestalt principles. According to their interpretation, the neighboring squares group together with the central one by shape similarity (Wertheimer, 1938), thus releasing the grouping by common region (Palmer, 1992) between the target and the central square.
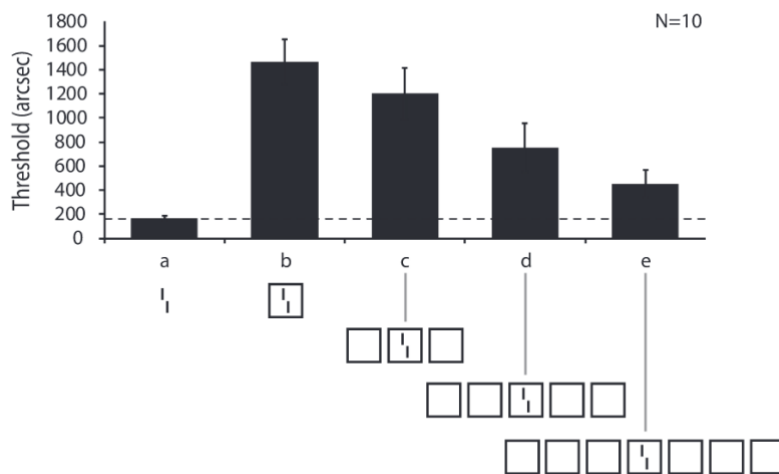
Figure 2.4: In Manassi et al. (2013), participants were asked to discriminate the Vernier offset direction, which determined the baseline performance. In scenario (b), the threshold reaches its highest

26

value, indicating a decline in performance attributed to classical crowding mechanisms. However, as progressively more flankers were added on both sides of the already flanked target (in scenarios c, d, and e), the thresholds decreased, and performance exhibited a linear recovery, approaching a level similar to the baseline.

There is agreement that crowding heavily relies on a critical spacing between flankers and the target, which is known as Bouma's window. However, it is essential to note that the impact of crowding is not limited solely to this spatial window, as elements outside of it can still significantly influence object recognition.

Taken together, these new findings have been summarized in the idea that "more can be better" (Herzog et al., 2015; Levi & Carney, 2009; Manassi et al., 2012; Põder, 2007; Rosen & Pelli, 2015; Vickery et al., 2009). This presents a novel perspective on the spatial attributes of crowding phenomena, deviating from the original assumptions. Increased distracting elements can sometimes enhance performance, challenging the limits set by Bouma's critical window. This underscores the broader impact of the entire visual scene's configuration on target perception, going beyond the influence of single elements alone.

2.2.2 Feature specificity and global configuration

Initially, the investigation of crowding focused on simple visual stimuli including letters, Vernier, and Gabor patches. This research primarily suggested that crowding was specific to low-level visual features, maintaining identical characteristics between the flankers and the target, such as color (Kooi et al., 1994; Põder, 2007; Põder & Wagemans, 2007) and orientation (Andriessen & Bouma, 1976). However, low level feature interaction between the target and flankers has also been demonstrated to be an unreliable predictor of crowding, not holding true in all cases. It seems instead that the global configuration of the visual scene can impact object recognition, leading to a more nuanced understanding of crowding phenomena.

Manassi et al. (2012) presented an intriguing demonstration of this, comparing the impact on Vernier offset discrimination by manipulating both flankers' colors and spacing. As can be seen in Figure 2.5, significant performance changes from the baseline (where crowding occurs) were only observed in 2.5a and 2.5c. In configuration a, the target was surrounded by ten red flankers arranged in a butterfly-like shape. Condition c retained the same configuration but with flankers alternating between red and green colors. This is particularly relevant since it cannot be attributed to the mere presence of

the red lines. When the red lines were presented alone with the same spacing as in condition c but without the green lines (2.5e), they did not produce any discernible effect.

This finding suggests that the task's outcome is significantly influenced by the global configuration of the stimuli, and the specificity of color, while possibly necessary, is not sufficient to induce crowding.
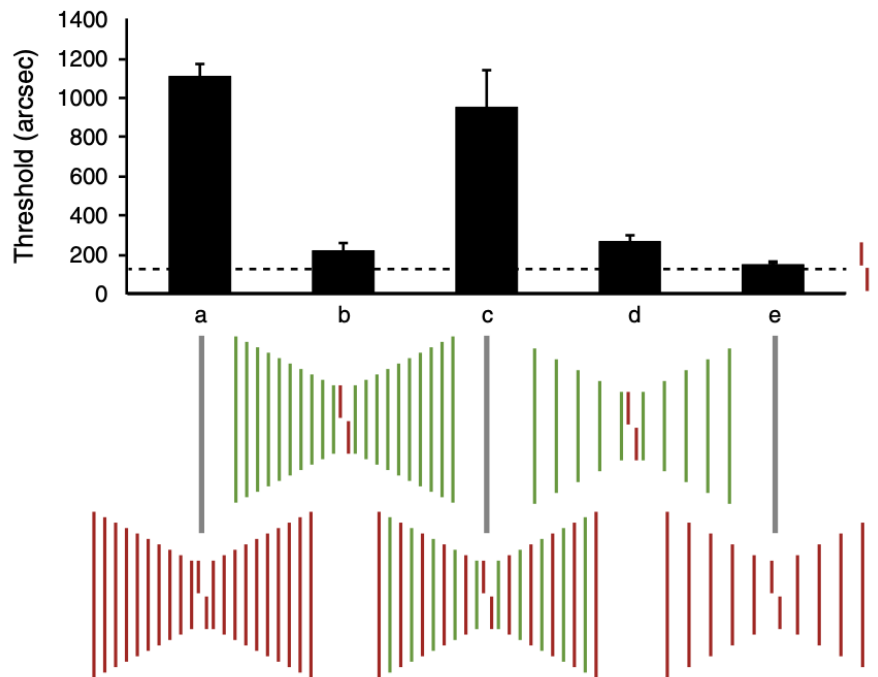
Figure 2.5: The stimuli used by Manassi et al. (2012) and their impact on Vernier offset discrimination, compared to the baseline condition with Vernier only. The red Vernier could be flanked by ten red (a), green (b) or alternated (c) flankers per side with increasing length. Spacing was also manipulated, creating two conditions where the number of flankers per side was reduced to five (d and e). Notably, performance deterioration was observed only in conditions a and c, where at least part of the lines was red, and ten flankers per side were presented.

Another example comes from Livne & Sagi (2011), who employed different configurations of Gabor patches in an orientation discrimination task. Crowding exhibited considerable variation depending on whether the Gabors formed a smooth contour or a star-like pattern (Figure 2.6). The overall configuration of the flankers played a crucial role, as crowding was notably weaker when the Gabors comprised a smooth contour compared to the star-like arrangement.
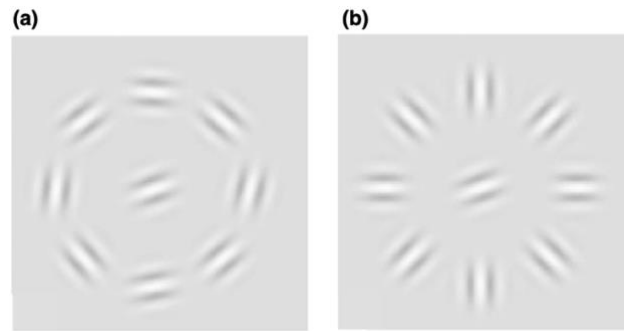
Figure 2.6: Stimuli employed by Livne & Sagi (2011). Flankers surrounding the target could form a smooth contour (a) or display a star-like arrangement (b). In condition (a), flankers group together and crowding is much less effective than in condition (b).

This evidence supports the notion that local models alone cannot fully account for the complexities of crowding (Clarke et al., 2014). While spatial extent does play a role, it is the specific stimulus configuration that ultimately determines crowding effects. Configurational influences are not minor modulations of crowding; they demonstrate substantial effect sizes and have the potential to reshape the pattern of results.

To address the challenge of including global aspects into models, extensive research has explored various approaches. Notably, a number of models has been tested, and the evidence indicates that incorporating grouping-like components significantly enhances model performance (Doerig et al., 2019). These findings underscore the importance of considering both local and global factors to attain a more comprehensive understanding of crowding phenomena.

2.3 Neural mechanisms behind crowding

The investigation into the spatial characteristics of crowding naturally transitions to an examination of the neural mechanisms involved in this phenomenon. Initially, a series of models based on local interactions was proposed as potential explanation for crowding. Some focused on visual features pooling (Pelli et al., 2004; Wilkinson et al., 1997; Zahabi & Arguin, 2014), object substitution (Ester et al., 2015; Strasburger et al., 1991; Zhang et al., 2012), and limited attentional resolution (He et al., 1996; Intriligator & Cavanagh, 2001).

As research progressed, it became clear that these early models were insufficient in fully capturing the complexity of crowding and its diverse manifestations. In response to this limitation, researchers began to shift their focus towards alternative mechanisms, with particular focus on grouping and global configuration (Bornet et al., 2021; Clarke et

al., 2014; Duangudom et al., 2007; Francis et al., 2017; Herzog et al., 2015; Herzog & Sayim, 2022; Manassi et al., 2013). These models showed promising potential in offering a more comprehensive account of the various behaviors exhibited by crowding. By incorporating grouping and global configuration, researchers have been able to gain deeper insights into how crowding operates and the intricate neural processes at play during this perceptual phenomenon.

2.3.1 Early approaches: pooling, object substitution and attentional resolution

Pooling models Traditional pooling models are rooted in a hierarchical perspective of visual processing (Riesenhuber & Poggio, 1999) (Figure 2.7A). According to this framework, stimuli are initially processed by neurons in V1 with small receptive fields. As the visual information progresses through the system, it undergoes integration at higher-level stages, where lower-level visual features are pooled together by neurons with larger receptive fields, forming a coherent perceptual whole (Figure 2.7B and C). In this hierarchical view, crowding is explained as a consequence of visual features from elements that are closely located in the visual space (such as the target and its flankers) becoming mixed-up during the hierarchical processing. This fusion leads to a deterioration of object recognition, as the distinctive features of individual elements become confounded and the ability to perceive and discriminate the target accurately is impaired (Pelli et al., 2004; Wilkinson et al., 1997; Zahabi & Arguin, 2014). In other words, crowding appears to involve an excessive integration of features over a large area due to the absence of smaller receptive fields, resulting in an ambiguous signal that becomes jumbled with the flanking elements.

Object substitution Object substitution models propose an explanation for crowding based on the notion that features of the target can be confused with those of the flankers. According to Ester et al. (2014, 2015), this is in accordance with a probabilistic substitution model of visual data processing, indicating that simple feature averaging, the basis of pooling, might not sufficiently explain the altered final perception. This model doesn't entirely exclude the role of pooling, as the authors stress that their findings stem from studies involving low target-flanker similarity, unlike most studies on crowding. However, their results seem to be more consistent with the idea that, during some trials

(thus the term 'probabilistic'), subjects struggle to distinguish the different features linked to the target and those linked to the flankers.
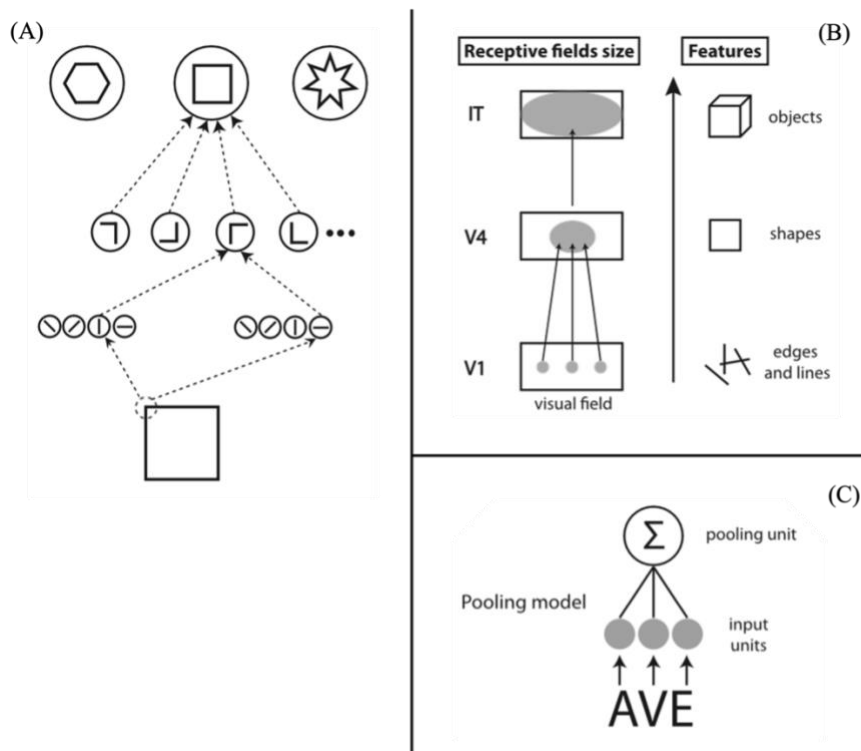
Figure 2.7: An outline of the hierarchical object recognition model proposed by Riesenhuber and Poggio (1999). In the early stages, cells engage in processing the visual object's low-level features, such as lines and their orientations. Subsequently, the signal is transmitted to higher-level cells, where these features are pooled together, ultimately enabling the recognition of the entire object (A). Neurophysiological depiction of the hierarchical local model of object recognition, emphasizing the progressive increase in the size of receptive fields as one moves up the hierarchy toward areas responsible for processing more sophisticated features (B). A simplified representation of pooling models, wherein each visual input is managed by a singular input unit in the brain. Through a feedforward process, all information is combined, finally generating a visual object with all the merged input characteristics. In crowding, the wider receptive fields can lead to the mix-up of each visual feature, producing an altered final perception (C).

Attentional resolution Other models have also investigated the role of attention in crowding. These models are built upon the concept that early visual stages manage to process low-level visual features with sufficient accuracy. Issues arise when attention comes into play to select the specific features of the target (Intriligator & Cavanagh, 2001). Within this framework, when the target and its flankers are too close, the selection process might erroneously capture task irrelevant information and subsequently mix up their features. This selection operates with what the authors describe as a "coarse grain", a level of detail significantly broader than visual resolution. Their assertion is that this coarser level of detail represents an unchangeable aspect of how attention accesses spatial

locations, independently from the capacity limit of attention (He et al., 1996; Intriligator & Cavanagh, 2001).

2.3.2 New perspectives: grouping and global configuration

The studies that contributed to shaping the models mentioned earlier were grounded in hierarchical and local processing of visual features. These models converged in the notion that crowding served as a low-level bottleneck (Levi, 2008), allowing investigations into feature separation only at the early stage of vision. Yet, some researchers began to distance themselves from these models configuration, exposing the limitations of these initial theories (Bornet et al., 2021; Clarke et al., 2014; Duangudom et al., 2007; Francis et al., 2017; Herzog et al., 2015; Herzog & Sayim, 2022; Manassi et al., 2013).

Local pooling models struggle to account for the influence of complex configurations on crowding strength, which led the most recent models to relocate crowding cortical mechanisms progressively higher within the visual hierarchy. Moreover, contrary to prior assumptions, crowding's intensity is tied to the holistic stimulus configuration, which provides further support to the involvement of higher-level processing (Herzog & Manassi, 2015).

Revealing the weaknesses of crowding initial explanations led to a shift towards a central role of perceptual grouping. According to this hypothesis, crowding happens when both flankers and the target are grouped, and its effects can vary based on different grouping cues. While crowding clearly impairs object recognition at multiple levels, it preserves the integrity of visual information, thereby impacting subsequent higher-level visual processing (Manassi & Whitney, 2018).

Perceptual grouping exploits visual cues such as texture, shades and depth cues to segregate visual targets from their backgrounds, inducing perception of objects edges and boundaries (Foley et al., 2012; Grossberg et al., 1997). Grouping is a good candidate to explain crowding phenomena as it is highly context-dependent, hence accounting for the most recent findings about crowding, with a particular focus on the fact that more flankers can reduce crowding and shape-shape interactions take place (Manassi et al., 2016). Given that this is true, it is not possible to only explain such mechanisms in terms of a feedforward model, but we must take into account both the global configuration of the visual scene and the possibility of recurrent interactions taking place in the visual processing stream that can manipulate perception as much as feedforward ones.

One potential explanation of this is that crowding takes place during recurrent processes where high- and low-level information dynamically influences each other, sometimes producing interference and disrupting perception (Foley et al., 2012; Grossberg et al., 1997). This view allows for an interpretation of crowding in the sense of a process that does not compromise low-level perception (we can accurately process all the basic features of the target and flankers). Alterations happen at the interactions level and are determined by the appearance of the stimuli, as we have evidence that global configuration and shape-shape relationships (objective measures) do matter and affect perception (Herzog et al., 2015). Appearance also depends on internal status of the viewer and there is evidence of a correlation between crowding mechanisms and subjective reports of how much the target stands out (Malania et al., 2007). In line with the principles of Gestalt theory, it seems that when a textured background assumes its own identity as a configuration, its disruptive impact diminishes. However, rules should be defined that determine how feature elements group together during a Vernier task, and by extension, how they function as distinct units in perception.

While grouping has gained substantial ground as an explanatory mechanism, its role in the genesis of crowding remains enigmatic (Herzog & Manassi, 2015; Herzog et al., 2015). Furthermore, the potential role of recurrent models in this context awaits further elucidation, though it is beyond the scope of this work, and it is not fully understood at present.

2.4   Temporal properties of crowding

While the spatial aspects of grouping were well defined and extensively investigated, little is known about its temporal properties. A first investigation of neural correlates of grouping and its timing was carried out by Chicherov et al. (2014). In their high-density electroencephalography (EEG) study, participants discriminated the offset of a Vernier alone and flanked by arrays of vertical lines that could be longer, shorter, or same length as the Vernier. Participants had the best performance when the Vernier was flanked by longer lines (thus the Vernier ungrouped from the lines), and the worst performance when target and flanker had the same length (the Vernier grouped with the lines). On the electrophysiological level, the reduction in performance was reflected by suppression of the N1 component (~180 ms after stimulus onset).

The results are consistent with the grouping hypothesis: the Vernier groups better with the flankers when they are of the same length. Hence, the authors suggested that grouping emerges slowly in the brain, likely as a result of feedback connections from higher areas.
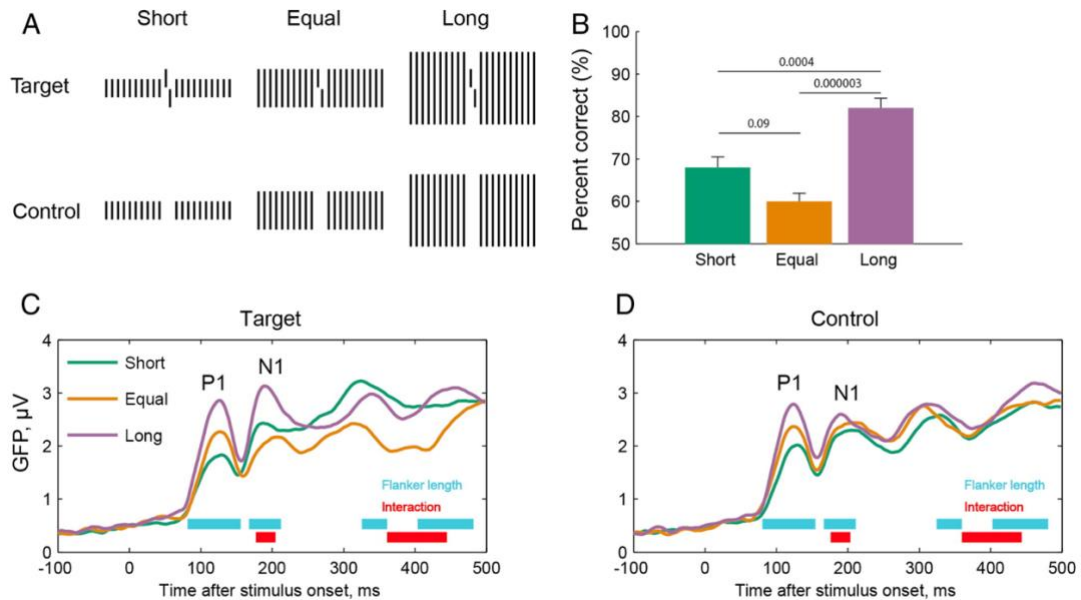
Figure 2.8: Stimuli employed by Chicherov et al. (2014). In the target condition, flanker lengths could be shorter, equal, or longer than the target Vernier. In the control condition, only the flankers were present, with no target displayed (A). The behavioral outcomes in the target setup were analyzed based on the percentage of correct responses for each flanker length. Optimal performance was observed when the flankers were longer than the target, moderate results were seen with shorter flankers, and the worst performance occurred when lengths were equal. These findings suggested that more grouping led to poorer performance, rather than an increased flanking energy (Herzog, 2022) (B). Measurement of global field power (GBF), an index of the EEG of all the electrodes taken together, in the target condition. The P1 component demonstrated its strongest GFP between 100-135 ms after stimulus onset for longer lines, intermediate for equal lengths, and lowest for shorter flankers. Regarding the N1 component, the most robust GFP was recorded 176-202 ms after stimulus onset for shorter lines (C). Measurement of global field power (GBF) in the control condition. P1 amplitudes are the same in (C), while N1 amplitudes show no differences across the three flanker conditions, suggesting that crowding effects are strongly task dependent.

A handful of studies has tried to investigate temporal properties of crowding with letters from a behavioral perspective, modulating the presentation time of the flankers with respect to the target (Tkacz-Domb & Yeshurun, 2017, 2021; Yeshurun et al., 2015). Before these paradigms were introduced, the almost totality of the behavioral studies on crowding has relied on the presentation of a flanker simultaneously to the presentation of the target, which has never allowed to identify a potential time window over which crowding actually occurs.

The first studies (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015) focused on 'pure' temporal crowding, where the presentation of the flankers was delayed with respect to that of the target. Figure 2.9 shows a schematic of the paradigm used in Tkacz-Domb & Yeshurun (2017). However, spacing was not manipulated, and the visual stimuli were presented in the same spatial location. Both studies revealed that temporal crowding does, in fact, occur: identification of targets was compromised only at smaller stimulus onset asynchronies (SOAs), i.e., temporal delays between that and the flankers. Further, this temporal effect was longer than in ordinary masking, that is typically effective when the mask is presented around 100-150 ms after the stimulus (Breitmeyer & Ogmen, 2000; Enns, 2004; Enns et al., 2000).
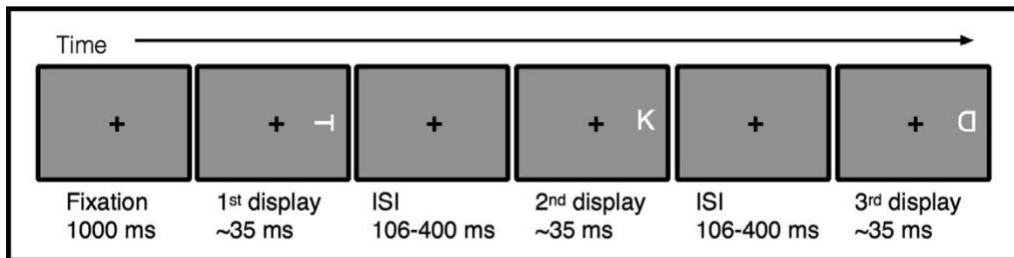
Figure 2.9: In the first study by Tkacz-Domb & Yeshurun (2017), they used a series of letters that could appear on either side of a central fixation cross. Each letter display was separated by a gap of time (inter-stimulus interval or ISI), which was randomly selected from the following options: 106, 129, 153, 177, 200, 235, 259, 306, 353, and 400 ms. The target letter consistently appeared in the same time spot within each trial but changed from one trial to another. Participants' task was to indicate the target's orientation and they received auditory feedback in response.

In a following study (Tkacz-Domb & Yeshurun, 2021) the authors also investigated the effect of temporal crowding on the continuous perceived orientation of a target line. The dependent variable was the response error based on the difference between real and perceived orientation of the target (Figure 2.10).
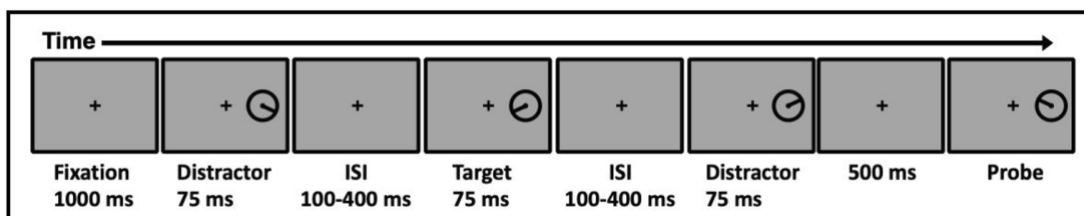
Figure 2.10: In the second study by Tkacz-Domb & Yeshurun (2021), they used a crowded condition where a series of three visual stimuli appeared either on the right or left side of a circle. The sequence included a distractor, followed by the target, and then the probe. In contrast, the uncrowded condition only displayed the target. These visual stimuli were separated by a specific time interval (inter-stimulus interval, ISI), which ranged from 100 to 400 milliseconds after stimulus presentation. The ISI remained consistent within each trial but varied between different trials. Participants were required to rotate the probe to match the perceived orientation of the target.

The results of first the two studies (Tkacz-Domb & Yeshurun, 2017; Yeshurun et al., 2015), however, do not provide evidence of an unequivocal difference between temporal crowding and classical masking. In the third study, Tkacz-Domb & Yeshurun (2021) claim that temporal crowding, spatial crowding and masking are indeed three different phenomena. This is because they observed that temporal crowding (i) impaired the precision of target's encoding, even when there was half a second between the presentation of stimuli and target, (ii) increased substitution errors, but (iii) did not affect the signal-to-noise ratio (SNR) of the target. Recent evidence also revealed that temporal crowding affected performance in the fovea as well (Sahar & Yeshurun, 2022).

Manassi & Herzog (2012) did an attempt to combine backward masking and crowding to investigate how the two differ in stimulus processing and what the temporal properties of grouping in crowding are. Their findings revealed that when a mask was presented after the target, thresholds remained on a constant level. Surprisingly, when the mask was presented before the target, thresholds changed accordingly to the type of mask employed.

Taken all this evidence together, the most accredited explanation is that vision acts in long-lasting periods of recurrent processes. This processing has a high spatiotemporal resolution, hence it can be affected by manipulation happening in this temporal and spatial window, and it is completely unconscious, only the output is conscious (Herzog, 2022). To better understand the role of temporal dynamics in crowding, it is necessary to further delve into the role of masking and its commonalities and differences with crowding mechanisms. This will be covered in Chapter 3 of this work.

2.5   Crowding and the Vernier task

The Vernier scale was originally introduced by the French mathematician Pierre Vernier in 1631. To interpret it, an observer engages in a Vernier acuity task, which requires detecting any misalignment between two adjacent vertical segments (Figure 2.11). The smallest noticeable gap between these segments, referred to as the offset, has been measured to be between 2 and 5 seconds of arc (Westheimer, 1987; Westheimer & McKee, 1977). This feature places Vernier acuity in the category of hyperacuity (Westheimer, 1975), that includes tasks with a threshold smaller than the size of foveal cone receptive field, which is approximately 30 arcseconds (Hu et al., 2021).
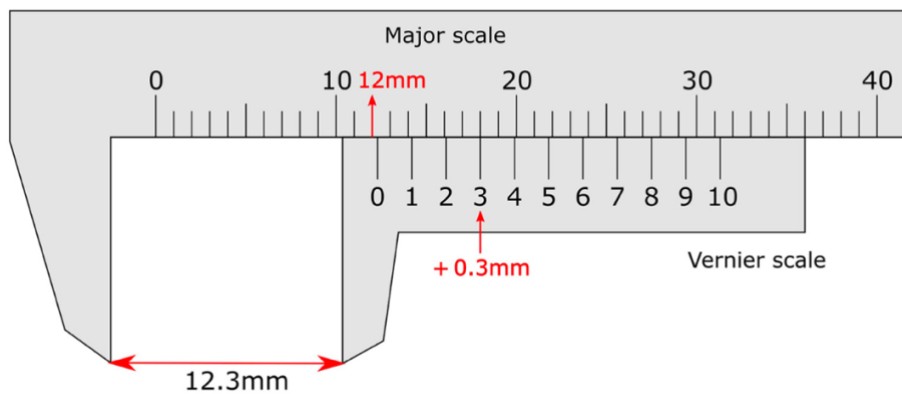
Figure 2.11: A caliper with the main scale on top of the Vernier scale used to measure the external size of an object. The offset is represented by the misalignment between an upper segment in the main scale and an adjacent segment in the lower Vernier scale.

Vernier lines have been widely employed in research to examine crowding phenomena (for instance, Fahle & Edelman, 1993; Herzog & Fahle, 1997, 1999; Levi et al., 1985; Levi & Klein, 1985; Levi & Waugh, 1994; Malania et al., 2007, 2007; Westheimer & Hauske, 1975). This acuity type is particularly prone to the disruption caused by surrounding distracting elements. This suggests the involvement of a cortical mechanism that processes visual inputs from a broader retinal area than where the stimuli are situated (Hu et al., 2021).

Using Verniers in visual perceptual tasks offers a distinctive advantage. They are simple to replicate and make tasks easily comprehensible for observers. Additionally, they provide a precise tool to investigate various aspects of human perception (Hu et al., 2021).

2.6   Conclusions

Crowding is an intricate and multidimensional phenomenon that can strongly affect the way we see things in everyday life. Recent research has expanded on its potential as a tool for understanding the gaps in existing models of human vision. This has been accomplished by investigating the effects of more elaborate stimuli and thorough manipulations, which have revealed an unexpected new role of grouping and global configuration in determining crowding mechanisms (Bornet et al., 2021; Clarke et al., 2014; Duangudom et al., 2007; Francis et al., 2017; Herzog et al., 2015; Herzog & Sayim, 2022; Manassi et al., 2013).

These findings have opened the discussion about the neural level at which these mechanisms occur, challenging the traditional pooling models and shifting the focus on

higher-level processes. In this view, the low-level representation of visual inputs remains intact. What is compromised is the subsequent combination of information from target and flankers, which get mixed up along the stream of visual processing. Hence, crowding cannot be defined as a bottleneck of lower lever vision (Herzog et al., 2015; Levi, 2008).

Yet, a lot of questions remain unanswered when it comes to understanding how visual crowding works. These questions include the fact that grouping alone is a necessary but not sufficient mechanism to explain why crowding causes a deterioration in performance, but most likely only represent an intermediate step that eventually leads to such disruptions (Herzog et al., 2015). Moreover, if crowding strongly correlates with how the global visual scene appears, it should be analyzed and understood what exactly is meant by *appearance* and how the subjective perception of a stimulus can change crowding manifestation (Malania et al., 2007; Herzog et al., 2015).

Another big question mark is left when it comes to analyzing the temporal dynamics of crowding. So far, the available work on temporal crowding has focused on either its behavioral characteristics without considering the spatial variables as well (Tkacz-Domb & Yeshurun, 2017, 2021; Yeshurun et al., 2015). Other works have only focused on the temporal aspects of crowding neural correlates (Chicherov et al. 2014). Overall, the existing evidence suggests some shared characteristics between temporal aspects of crowding and masking (Duangudom et al., 2007; Tkacz-Domb & Yeshurun, 2021), thus proposing an interesting perspective to consider for further investigations. In Chapter 3, I will delve into the behavioral evidence regarding the spatiotemporal extent of crowding, as previous studies have strongly linked this phenomenon with the dynamics of masking.

In conclusion, crowding is characterized by simple yet effective designs, making it an ideal tool for studying a wide range of visual processes. While both temporal and spatial aspects contribute to the complexity of the phenomenon, research has primarily focused on the spatial aspects, leaving much to be explored regarding its temporal dynamics. One of the aims of this work is to further delve into the temporal window in which crowding occurs.

Chapter 3

Relations between crowding and masking

3.1   Differences and similarities between crowding and masking

Visual masking and crowding are two perceptual phenomena that can induce changes in how we perceive visual stimuli when they are presented alongside specific patterns. Chapter 2 extensively explained how crowding works and the factors that can influence its manifestations. Crowding is sometimes compared to a particular type of masking, where the mask[2] appears in a different location on the retina than the target (Huckauf & Heller, 2004). This is typically referred to as "lateral masking", distinct from "ordinary masking" where the visual mask appears in the same retinal position as the target, although typically with a slight time delay (Pelli et al., 2004; Polat & Sagi, 1993).

As illustrated in Chapter 2, crowding appears to be sensitive to various factors, like the overall configuration of the flankers in the visual scene and their proximity to the target. Interestingly, all forms of masking share this sensitivity with crowding, together with the specificity for spatial frequencies and the reduction of the effect with low-contrast flankers (Chung et al., 2001). This might initially suggest a strong similarity between these two mechanisms, even implying their equivalence.

The literature from the last two decades has provided contrasting evidence. In particular, studies that focused on the spatial differences between crowding and masking revealed a substantial difference between the two (Chung et al., 2001; Pelli et al., 2004). On the other hand, studies that have emphasized the temporal dynamics of these two mechanisms pointed out that similarities between crowding and masking might be more than what emerged from the investigation of their spatial properties (Chung, 2016; Harrison & Bex, 2014; Huckauf & Heller, 2004; Y. Yu et al., 2020, 2021).

3.1.1 Is crowding different from lateral masking?

The main distinction between crowding and masking lies in the level at which they impact visual information. In crowding, the information about the target is mixed with that of the mask, leading to challenges at the level of target recognition. On the other hand, masking

---

[2] Typically, in the context of masking, what interferes with perception of the target is known as the "mask". In crowding, this is referred to as "flanker".

involves the complete replacement of the target by the mask, resulting in an irreversible loss of the target signal. This differentiation aligns with the concept of a multilevel processing of visual information, wherein detection occurs initially, followed by the assembly of components for recognition (Herzog et al., 2015; Herzog & Manassi, 2015; Manassi & Whitney, 2018). Crowding appears to affect perception at higher processing levels, leaving the detection of the target's presence relatively unaffected (e.g., Chakravarthi & Cavanagh, 2009; Levi, 2008; Manassi & Whitney, 2018; Parkes et al., 2001; Pelli et al., 2004; Pelli & Tillman, 2008; Strasburger & Malania, 2013). In contrast, masking compromises perception from the very first stages. In other words, masking impairs feature detection while crowding can start interfering with perception at the level of feature identification and integration (Parkes et al., 2001; Pelli et al., 2004).

Parkes et al. (2001) also suggested that crowding and lateral masking are essentially different, after assessing observers' ability to estimate the average orientation of a Gabor patch surrounded by other similar patches. Even though we are not able to explicitly report the features of what we have seen, we can still have a sense of the general texture, suggesting that in crowding the signal is simply confused and jumbled rather than lost, as in masking.

There is strong evidence that when it comes to spatial properties of the visual scene, such as texture perception (Levi et al., 2002; Parkes et al., 2001) and critical spacing (Bouma, 1970; Strasburger et al., 1991; Toet & Levi, 1992). However, when the focus is shifted on predominantly temporal aspects of two phenomena, we can come to slightly different conclusions.

3.1.2 Temporal dynamics of masking and the time window for mask-target interactions

While temporal aspects of crowding have received little exploration, a detailed exploratory study conducted by Huckauf & Heller (2004) has examined the temporal dynamics in masking. Visual temporal masking relates to disruptions in object recognition tasks due to mask presentation at the target's location, but with a temporal delay that can occur either before (forward or paracontrast masking) or after (backward or metacontrast masking) the target display.

They employed a classic crowding paradigm, in which participants were asked to recognize flanked letters appearing in the periphery of the visual field. The innovative

aspect of their study consisted in the manipulation of both the timing and the spatial location at which the flankers were presented. They examined changes in performance across different flanker positions, including three spacings from the target (0.4, 1, and 2 degrees of visual angles, as depicted in Figure 3.1) and three eccentricities, quantifying the distance between the center of the screen and the stimulus center, with measurements of 1, 4, and 7 degrees of visual angle (illustrated in Figure 3.2). In addition, flankers could appear with a variable temporal delay with respect to the stimulus (stimulus onset asynchrony, SOA), from -150 ms to +150 ms, every 50 ms. Since this focus of this thesis is to analyze the effects of perceptual disruption following the onset of the stimulus, I will only discuss the interpretations of results in the right halves of Figure 3.1A and 3.1B, that is those regarding the case in which the flankers appeared at different delays only *after* the stimulus was displayed.

At small spacings (0.4°) and large eccentricities (7°) we can see what the authors define as Type A masking, a monotonically increasing masking function in which performance is progressively better as the delay between target and mask gets bigger. A very different masking function is observed when the mask is instead presented at intermediate spacings and eccentricities after the onset of the stimulus. When the flankers appeared at 1° or 2° from the target or at 1° or 4° from the center of the screen, a dip in the performance was observed at 50 ms SOA. This has been defined by the authors as Type B masking, in which a U-shaped masking function appears, and target recognition is impaired at intermediate SOAs (between 50 and 150 ms after stimulus onset).

The first effect has been explained in terms of integration and interruption of information processing (Scheerer, 1973): in case of shorter SOAs, the perceptual information about masks and target is integrated, finally causing a disturbed processing of the visual input. When SOAs are longer, interruption comes into play, as the delayed presentation of the mask reduces the time available for processing the initial target information. On the other hand, Type B masking must invoke a completely different explanation. The most plausible one suggested by the authors refers to a pioneering study by Di Lollo et al. (2000) which suggests that information processing is based on interactive activation and it includes both feedforward and feedback signals that take place with different timings. The idea is that higher-level cortical areas send feedback signals to lower-level ones, causing interference at some point with the incoming signal.

This would link Type B masking to crowding for two main reasons: (i) both crowding and Type B masking increase with increase in eccentricity and (ii) if Di Lollo et al.'s (2000) assumption is true, both also involve higher level mechanisms and recurrent signals (e.g., Bornet et al., 2021; Clarke et al., 2014; Duangudom et al., 2007; Francis et al., 2017; Herzog et al., 2015; Herzog & Sayim, 2022; Manassi et al., 2013). This distinction agrees with Huckauf and Heller's (2004) claim concerning variations in spacing and eccentricity: higher-level information becomes accessible when there is a combination of larger spacing and smaller eccentricity, whereas very small spacings and larger eccentricities do not permit the feedback of detailed information or even their processing.

In a third study, Huckauf & Heller (2004) also observed that Type B masking is more pronounced when the flankers are letters as the stimulus than when they are non-letters, providing further evidence of the involvement of higher-level recurrent interactions that are more pronounced when there is congruence between the type of stimuli used as flankers and target. These findings can also be interpreted as an additional support that crowding is not a mere bottleneck of vision (Levi, 2008) and is not limited to grouping of adjacent visual information. It is instead the result of interference at higher levels of the visual processing stream, that act through specific (but probably still unknown) mechanisms that happen at specific timing and cannot be reproduced if the required conditions are not fulfilled.
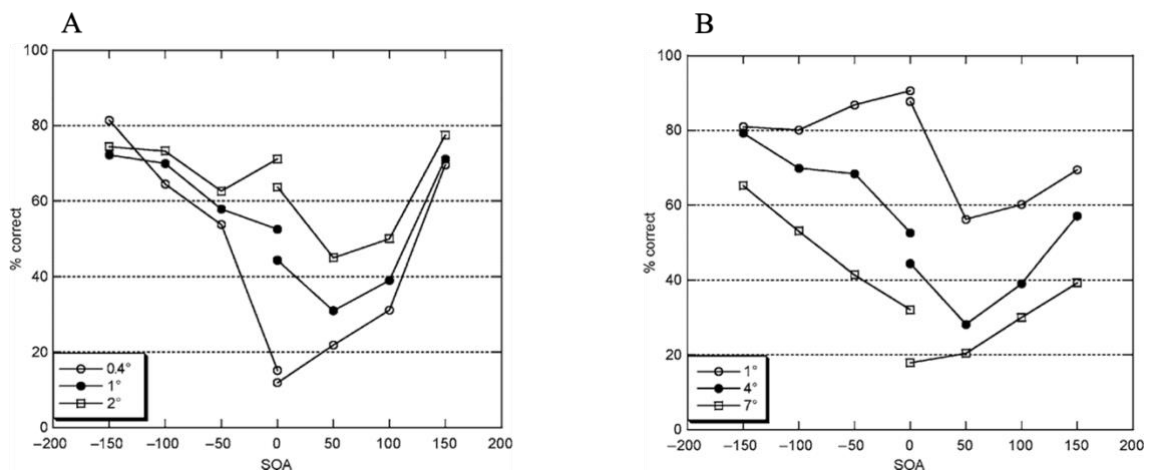
Figure 3.1: Results from study 1A and 1B in Huckauf & Heller (2004). Focusing on the masking function obtained with positive SOAs, the smallest spacing value and the largest eccentricity value (0.4° and 7° respectively) produce a monotonically increasing function, with % of correct response being the lowest when the target and the mask appear simultaneously and progressively increasing as the temporal delay between the two stimuli increases, too. For intermediate spacings (1° and 2°) and

eccentricities (1° and 4°) the functions appear U-shaped, with worst performance in correspondence of the 50 ms delay between target and mask.

Many of the results about object recognition tasks employing a delayed foveal noise in the context of foveal feedback (Chambers et al., 2013; Contemori et al., 2022; Fan et al., 2016; Weldon et al., 2016, 2020; Q. Yu & Shim, 2016) show a performance impairment trajectory that mirrors that of Type B masking. However, some aspects of these studies should be considered as potential confounds. For instance, at the largest eccentricity value of 7°, type B masking was no longer observed, and maximal crowding appeared again when the target-mask presentation was simultaneous. Chung (2016) claimed that this was caused by the size of the stimuli, which was kept identical throughout each condition of the study, inducing a potentially confounding effect given by letter size and spacing and not a genuine consequence of variations in eccentricity.

Further research has confirmed the importance of timing between a target and its mask, in addition to the central role of space dependence (Chung, 2016; Chung & Patel, 2022; Harrison & Bex, 2014; Y. Yu et al., 2020, 2021). It is relevant to note that most of these studies have used letters from the English alphabet to investigate the presence of a critical timing in which target-mask interactions occur, limiting the possibility to extend results to different, lower-level types of stimuli. Harrison & Bex (2014) identified a critical window of 45 milliseconds in which they observed the most significant drop in performance. This finding suggests that this specific timing corresponds to when mask-target interactions have the most pronounced detrimental impact. Similarly, Chung (2016) revealed that maximal crowding occurs when the target and the flankers do not overlap in time. Among the other results, she observed a dip in performance (percentage of correct responses) when a target letter presented at 10° eccentricity was flanked by adjacent distractors. The dip would shift towards 50 ms SOA as the distance between target and flankers increased (as can be observed in Figure 3.2), consistently with the results from Huckauf & Heller (2004), in which U-shaped masking functions were more visible as target-flanker eccentricity increased. These results also show that, unlike in Huckauf & Heller (2004), the U-shaped function is maintained for larger eccentricities that exclusively involve the periphery of the visual field.

Similar results were observed in another study from the same author (Chung & Patel, 2022), that concluded that the spatiotemporal configuration of target and flankers is what eventually determines the strength and manifestation of crowding. While neither

the timing of the flankers nor the position of target alone can determine the strongest crowding effects, they are both necessary. In particular, the authors found again a U-shaped crowding function in the performance of all the participants tested when there was a target-flanker delay between 20 and 100 ms approximately (Figure 3.3), and the effect was modulated by the spacing between the two stimuli, consistently with previous findings described so far.
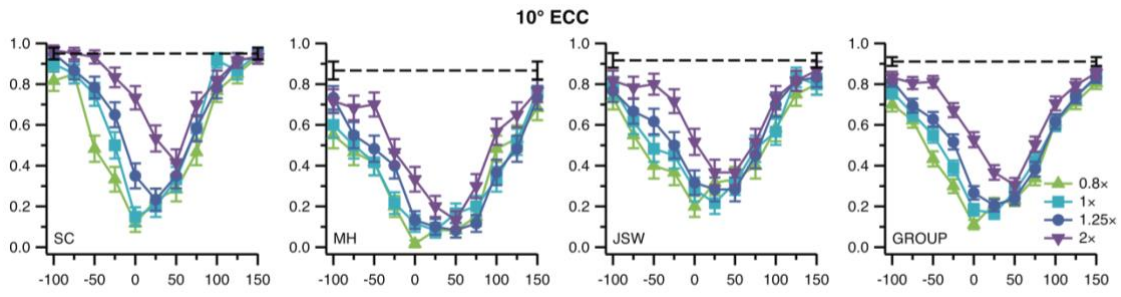
Figure 3.2: Results the first experiment by Chung (2016), with target duration fixed at 50 ms at eccentricity of 10º. The first three graphs represent the performance (proportion of correct responses) for each of the three participants, and the last one is the average of the individual results. The dip in all the participants' performance changed with the spacing between letters. At smaller spacing values, the worst performance was observed around 0 ms SOA, that is when the letters appeared simultaneously. As spacing increased, the dip would shift to the right, ultimately leading to strongest crowding effects at 50 ms SOA, when the spacing was twice the height of the stimuli.
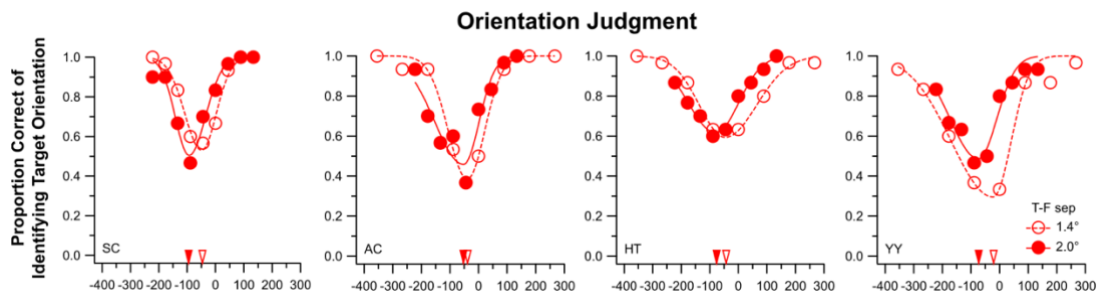
Figure 3.3: Results from the orientation judgement task by Chung & Patel (2022) showing the relationship between the proportion of correct responses and different target-mask SOAs in an orientation discrimination task. The dip in the performance was observed when the target was presented approximately 20 to 100 ms before the target.

Other more recent studies arrived at similar conclusions, but found a later dip with respect to the aforementioned studies. Y. Yu et al. (2020, 2021) reported that manipulating the temporal presentation of the flankers, the maximal crowding effect was observed when flankers appeared 100-150 ms after target onset. Here, they also manipulated presentation time, which is another crucial factor that is known to influence crowding dynamics (Chung & Mansfield, 2009; Tripathy et al., 2014).

In conclusion, these findings collectively suggest a trade-off between the spatial and temporal properties of crowding (Chung, 2016). This relationship is what allows to observe analogous variations in performance through the manipulation of timing and spacing in contrasting directions. Furthermore, emerging evidence suggests that the interactions between masks and targets occur within a narrow time window, typically ranging from 50 to 150 ms after target onset.

3.2   Masking the flankers (Chakravarthi & Cavanagh, 2009)

One more study that has a particular relevance for explaining the purpose of this work is that from Chakravarthi & Cavanagh (2009), who investigated the relationship between masking and crowding through an innovative approach. Their goal was to deepen the understanding of the locus of crowding, and they did this by investigating whether a crowded object could be recovered through the technique of masking the flankers.

As shown in Figure 3.4, the researchers used three types of masks that interacted with the target at different processing levels, from both a spatial and temporal point of view: (i) noise masking (at a lower level), (ii) metacontrast masking (at a mid-level), and (iii) object substitution masking (at a higher level). Their investigation revealed a recovery in accuracy when masking by noise and metacontrast masking were applied, while no differences were observed between the crowded target condition and the masked flankers in the case of object substitution trials. These results provide further details about the temporal extent of crowding dynamics, placing it beyond the earliest stages of visual processing but before the efficacy of object substitution, as evidenced by the persistence of noise and metacontrast masking effects on the flankers, but not of those of object substitution.
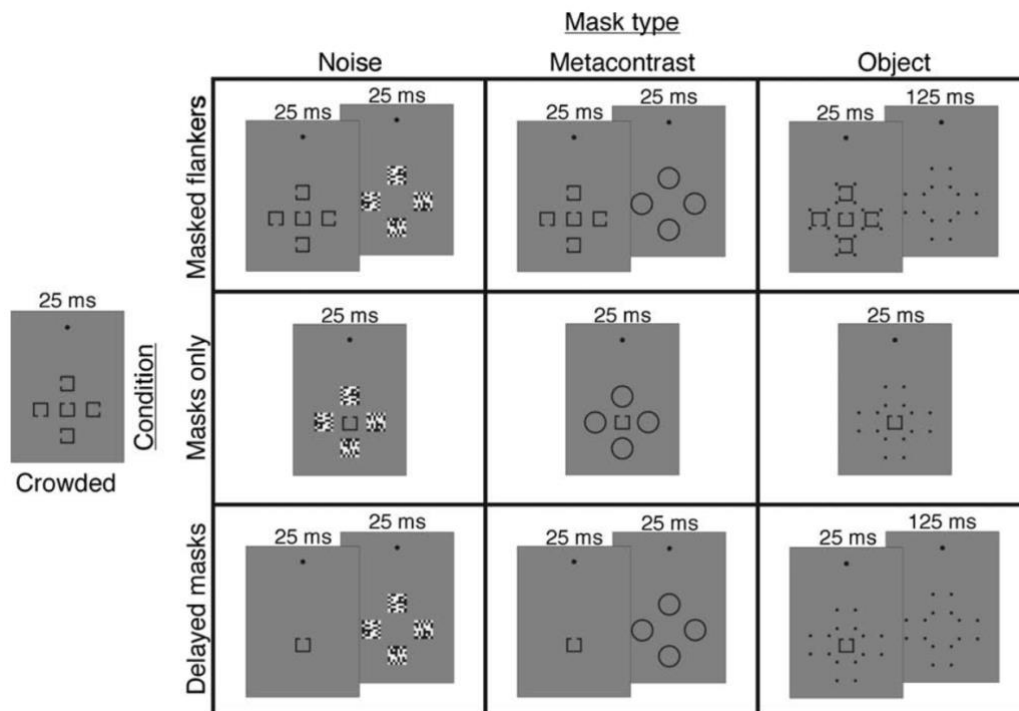
Figure 3.4: Description of the type of masks employed by Chakravarthi & Cavanagh (2009). The experiment included 9 test conditions along with a baseline where only the crowded target was presented. They varied the impact of the mask on object recognition by using noise masking, metacontrast masking, and object substitution masking. Each of them affects visual processing at a different level, here from the lowest to the highest. They also included conditions where the target was masked but not crowded, as well as a condition where the mask appeared after the target.

3.3 Conclusions

According to Strasburger (2020), recent years have seen a significant shift in attention from low-level visual acuity to higher-level processes within crowding. He labels this transformation as a "paradigm shift." The growing interest in the temporal dynamics of crowding can be seen as part of these higher-level considerations.

The debate regarding the differences and similarities between crowding and masking remains ongoing. As shown in several studies (Parkes et al., 2001; Pelli et al., 2004; Strasburger et al., 1991; Toet & Levi, 1992), these two phenomena share certain traits but remain fundamentally distinct. They affect different stages along the visual processing pathway: masking impacts object recognition and codification from early levels, while crowding influences higher stages, affecting the integration of lower-level stimulus features.

Huckauf & Heller (2004) highlighted the division of masking into at least two separate processes, which become discernible when temporal aspects are considered. This

led to further studies that, accounting for temporal dynamics, suggested more commonalities between these phenomena than initially believed (Chakravarthi & Cavanagh, 2009; Chung, 2016; Chung & Patel, 2022; Harrison & Bex, 2014; Y. Yu et al., 2021).

Efforts have been made to explain the spatiotemporal effects of crowding and masking, with recent findings indicating the influence of the increasing size of human perceptual fields as also eccentricity increases (Lev & Polat, 2015). These insights provide some clarity on the intricate interplay of space and time in crowding, although many aspects remain unexplored, and the exact extent of the crowding temporal window is still to be unambiguously defined.

Chapter 4

Comparing Crowding and Foveal Feedback mechanisms using a Vernier task

4.1 Introduction

In the opening chapters of this study, I provided a summary of the key discoveries in the context of foveal feedback and crowding, the two mechanisms central to the experiment we conducted and that will be described in the next paragraphs. Investigating their characteristics, also including the contribution of temporal dynamics of metacontrast masking to our understanding of crowding, represents an essential step to clarify the motivations of this research.

Crowding and foveal feedback are visual phenomena that mostly affect object recognition in the periphery of the visual field. In both cases, the perception of a target is impaired by the presence of confounding elements that, unlike in ordinary masking, appear in a different spatial location from that of the target. In crowding, the flankers are positioned close to the target (Bouma, 1970), inducing a mix-up of information from adjacent visual elements. In the special case of foveal feedback, however, the noise is always displayed in the fovea and, when presented with a critical delay from the target, can disrupt information processing by the generation of a cortical representation in an area different from that corresponding to the retinal position of the stimulus.

As to masking, there is general agreement that it causes perceptual disruptions when a distractor appears after a certain time delay relative to the target and occupies the same location as the target. On the other hand, crowding leads to disruptions when a distractor appears simultaneously with the target but occupies a nearby location. By relying on foveal feedback mechanisms, we would expect disturbances to occur when noise is presented both after a time delay and in a different location compared to the target. This means that the feedback mechanisms can influence perception under conditions involving both temporal and spatial differences between the target and the distractor.

We considered one common element that these phenomena might share, the involvement of recurrent signals between higher-level and lower-level areas. This was

first theorized by Di Lollo, Enns, and Rensink (2000), who focused on the role of "iterative reentry" to account for metacontrast masking and masking by object substitution. Given recent findings concerning foveal feedback mechanisms, which suggest their pivotal role in recognizing objects with a delayed mask, including the creation of entirely new high-level cortical representations for stimuli that don't align with retinal input, we hypothesized that these mechanisms could also exert an influence on crowding. Multiple studies have suggested the presence of a temporal window during which target-flanker interactions take place, producing maximal crowding between approximately 50 and 150 ms after the target has been presented. Assuming this as true, we could interpret such finding according to Di Lollo's (2000) view of re-entrant processing. If target-flanker interactions occur in this time window, we can agree that this is consistent with the critical timing that has been observed for recurrent processing in foveal feedback. Given these similarities especially in the temporal domain, we thought that foveal feedback might affect target-flanker interactions in crowding. Therefore, we wondered whether it would be possible to obtain similar results in performance disruption by using a classical crowding paradigm. We wanted to test this in terms of Vernier acuity, which has never been investigated either in foveal feedback or temporal crowding paradigms, as we want to focus on effects on lower-level stimuli. The paradigm we designed also allowed us to better investigate noise characteristics. We explored whether the foveal disruptor needed to be a complex, attention-grabbing object, or if even simpler, low-level foveal stimuli could induce disruption. Furthermore, we investigated whether the complexity of the foveal stimulus impacted the degree of impairment. We aimed at determining if the observed changes in performance (such as a dip in accuracy and a peak in Vernier thresholds) were solely a byproduct of the task itself, or if the foveal location of the disruptor influenced performance on peripheral targets.

Experiment A adapted the paradigm from previous studies (Contemori et al., 2022; Fan et al., 2016) to investigate whether the disruptive effect of a dynamic noise patch would also apply to a low-level target as a Vernier. Given that a dynamic noise is a relatively extensive and disruptive stimulus, we anticipated its effects to remain observable even when tasks are less challenging and involve lower-level stimuli, such as a Vernier offset.

In Experiment B, we combined traditional foveal feedback paradigms with crowding to gain a deeper understanding of these phenomena and their relationship. We aimed to explore the nature of the noise and determine whether a similar effect could be obtained with a less invasive noise compared to a large, dynamic, and colorful patch. Our assumption was that the noise would exhibit specificity with respect to the target, suggesting that even a simpler distractor could impact performance significantly, as long as it possesses the required characteristics, and this association could be related to crowding effects.

Lastly, in Experiment C, we examined the impact of rotating the foveal flankers. This allowed us to further investigate whether the effects observed in the previous paradigm were tied to the specificity and visual characteristics of the noise, particularly its interaction with the target in a task-dependent manner. Alternatively, we aimed to determine whether the effect could be induced by any type of foveal distractor, regardless of its specificity or visual features. We did not expect to observe any variations in performance when the noise employed is of a low-level nature and does not align with the characteristics of the target.

4.2 General methods

4.2.1 Participants

Participants were paid students of the École Polytechnique Fédérale de Lausanne (EPFL), who provided written informed consent before the experiment and were naïve as to the purpose of the experiments. They were all above 18 years of age and had normal or corrected-to-normal vision, with a visual acuity of 1.0 (corresponding to 20/20) or better in at least one eye, measured with the Freiburg Visual Acuity Test (Bach, 1996). All participants were told that they could quit the study at any time, but not one asked to. The study was approved by the local ethics committee under the Declaration of Helsinki (apart from preregistration). The selection of the number of participants for each experiment was based on prior research on crowding effects conducted within the laboratory and on initial observations gathered through pilot studies involving volunteers from the lab.

In Experiment A, data from nine out of the eleven participants were used for analysis. In Experiment B, data from twelve participants out of the fifteen who took part were considered. In Experiment C, data from twelve out of the fourteen participants who participated were included in the analysis. All the excluded participants exhibited a

ceiling effect, and there was no noticeable difference between their baseline performance and their performance in trials with noise.

The presence of individuals who consistently exhibited ceiling effects across all experimental conditions could be attributed to several factors. One possibility is that the task itself may not have posed a significant challenge, particularly for individuals with unusually high Vernier acuity, resulting in a lack of detectable effects within the paradigm we employed. Additionally, it is important to note that some participants may not have strictly followed the instructions, possibly foveating the target despite being explicitly instructed to maintain their gaze on the fixation cross. In anticipation of our results, it is worth acknowledging that the observed effect was relatively modest, suggesting that individual differences may have played a substantial role in determining its detectability.

4.2.2 Apparatus and procedures

Before undergoing the real experiment, each participant performed the Freiburg visual acuity test (Bach, 1996) to measure binocular visual acuity. Landolt-C optotypes with a gap in one of eight possible orientations ("up," "down,", "left", "right", "up left", "up right", "down left" and "down right") were presented. The size of the optotype changed adaptively, and the orientation of the gap was chosen randomly. Participants were asked to indicate the direction of the gap by using a keyboard with eight different buttons, each of them positioned in one of the eight possible locations of the gap (Figure 4.1). The task was carried at a viewing distance of 2 meters from the screen and was repeated three times, once for each individual eye and once with both eyes open. The results returned the decimal visual acuity value; only participants with a binocular visual acuity ≥ 1 were included in the study.
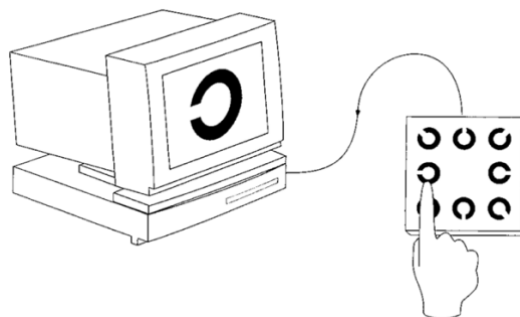
Figure 4.1: A representation of the Freiburg visual acuity task (Bach, 1996). A Landolt-C optotype appears on the screen and the observer is asked to press a button that corresponds to the position of the opening in the viewed stimulus. The size of the optotype reduces with correct responses and increases with incorrect responses.

Stimuli from the main experiment were presented on an Asus VG248QE LCD monitor. Screen resolution was 1920x1080 pixels for a screen size of 53x30 cm and it was operated at 120 Hz refresh rate. Background luminance of the screen was below 1 cd/m^2. Luminance of the stimuli was 80 cd/m^2. The stimuli and the experimental paradigm were generated with custom-made scripts written in MATLAB and the Psychophysics Toolbox extension (Brainard, 1997), running on a Windows-based machine. Experiments were performed in a quiet and dimly illuminated (0.5 lx) room. Participants had a viewing distance from the screen of 75 cm and kept their head on a chin rest throughout the whole experiment, to avoid excessive head movement and facilitate fixation.

Before the first block of the experiment, participants received instructions about the task through a presentation. Following this, they engaged in a training phase where auditory feedback was given to help them understand and familiarize with the task. The feedback was then removed during the test trials (for the role of feedback in trials, see Herzog & Fahle, 1999). The participants were instructed to keep their gaze fixed on a red cross that remained visible throughout the entire experiment, positioned 9º to the left of the screen's center. The target was a Vernier, a small visual element that appeared on the right of the fixation cross (the center of the stimulus was at 12º from the cross and was corrected to 9º if the participant could not perform the task). The Vernier offset was adjusted adaptively during the experiment following a PEST function (Taylor & Creelman, 2005). The lines got closer as the participant provided correct responses, while they got further with wrong responses.

Each session was split into seven short blocks, each consisting of 80 trials, repeated two times (for a total of 1120 trials). Every block featured a distinct SOA, indicating the timing at which a foveal noise would appear on the screen after the target onset. Additionally, a control condition presented only the Vernier. Participants in both experiments performed two repetitions of these blocks, leading to a total of four experimental sessions that enabled a test-retest comparison. Some participants completed all four sessions in the same day, while others chose to divide them across different days.

The task involved determining the direction of Vernier offset, that is participants had to indicate whether the lower line of the Vernier appeared shifted more to the right or left compared to the upper line. They were instructed to constantly fixate on the red

dot. To provide their responses, participants used two distinct buttons they held in their hands while performing the task. If they perceived a leftward offset, they pressed a red button held by their left hand. Conversely, for a rightward offset, they pressed a green button held by their right hand.

4.2.3 Vernier acuity threshold and threshold elevation

Performance is expressed in threshold values, with higher values indicating lower sensitivity and, hence, worse performance. Thresholds were defined as the horizontal offset (in seconds of arc) at which participants could correctly discriminate 75% of the directions. We used a PEST staircase procedure to adapt thresholds during the experiment (Taylor & Creelman, 2005). The offset range went from 10 to 2000 arcsec. The starting offset was 200 arcsec. Thresholds were determined after fitting a cumulative Gaussian to the data (Wichmann & Hill, 2001b, 2001a). We compared the different thresholds at each of the SOAs. First, we calculated them separately for each individual attempt and then computed a grand average.

In parallel to the threshold values, we also computed threshold elevation (TE) for both the experiments. We converted the thresholds (T) for each SOA into a multiple of the threshold values for the unmasked condition (TU). Each threshold of the masked conditions was divided by the threshold that the subject showed in the unmasked Vernier.

$$TE_{SOA} = T_{SOA}/TU$$

4.3 Experiment A – dynamic noise patch and Vernier

4.3.1 Stimuli

The Vernier target consisted of two white vertical bars on a black background. Each segment of the Vernier was 2400 arcsec long and 108 arcsec wide. The two segments were separated by a vertical gap of 240 arcsec, for a total length of the target of 5040 arcsec. The upper line would appear in the same position at each trial, with an eccentricity of 12º with respect to the fixation cross. Only one participant performed the task with the Vernier at 9º of eccentricity from the fixation, as they reported to be struggling in perceiving the Vernier in the far periphery.

The lower line position would shift along the horizontal axis according to the adaptive procedure. The position of the bottom line with respect to the upper one was randomized through the trials and could be either to the left or to the right (Figure 4.2A).

The number of left and right offsets was balanced and was presented as a pseudorandom sequence.

The foveal noise was the same dynamic patch used by Contemori et al. (2022) and Fan et al. (2016). The patch was a square of 7° × 7° of size (Figure 4.2b). The perception of movement was induced by the multiple smaller squares that the noise was made up of, whose color changed repeatedly during the presentation of the patch.

Figure 4.2: The stimuli used in Experiment A. Three examples of the Vernier with different offsets and degrees of separation between the upper and the lower line. The offset could be either to the left or to the right. The distance between the two segments would shift based on the participant's responses according to the adaptive procedure (A). A frame of the dynamic foveal noise (B).

4.3.2 Experimental paradigm

In the experiment, trials were arranged into 7 short blocks that lasted around 4 minutes each. A red fixation cross appeared on the screen for 1000 ms, followed by the target stimulus that was displayed for 20 ms. The target consisted of a white Vernier made up of two vertical lines with a variable vertical offset. All the stimuli were presented on a black background. After the presentation of the target, a noise would appear for 83 ms, centered at fixation and with 6 different SOAs: 0, 50, 100, 150, 200 and 250 ms. The SOA changed across each block but remained consistent within trials of the same block. There was also a control condition in which only the unmasked Vernier appeared. The order of the block for each SOA was randomized for all participants. After one target-noise sequence was displayed, observers were asked to perform a two-alternative forced choice (2AFC): left or right, depending on the offset they perceived between the two lines constituting the Vernier. Based on the participant's response, the Vernier offset would change according to the adaptive procedure described in 4.2.3. Figure 4.3 provides a schematic of the experimental paradigm described above.
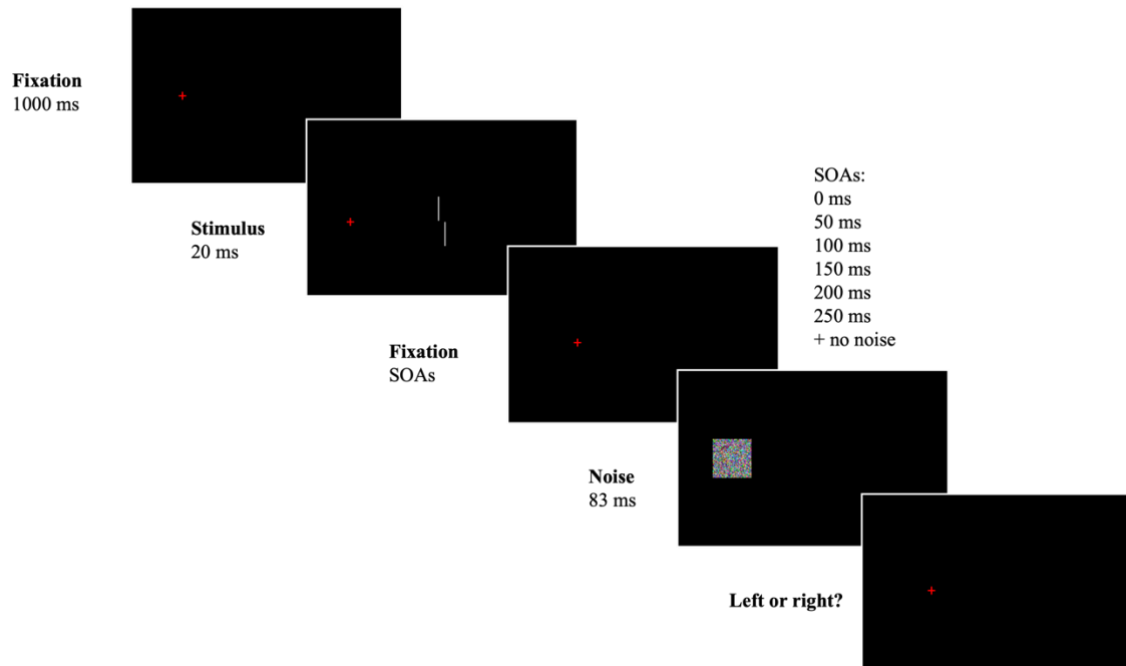
Figure 4.3: Trial structure of Experiment A (*n* = 9). A red fixation cross appeared on the screen and remained visible throughout the whole experiment. After 1000 ms, the peripheral stimulus was displayed on the right of the fixation cross for 20 ms, 9º or 12º away. Then, the colored dynamic noise was presented in the fovea. Each block of trials had a different SOA for the noise presentation, which ranged from 0 to 250 ms. There was also one block in which only the unmasked target appeared. Finally, the observer was asked to express the perceived side of the target Vernier offset.

4.3.3 Results

For each participant, we determined the threshold values for every SOA as well as the Vernier-only condition. To do this, we computed the average of their results from both the first and second attempts. Subsequently, we conducted pairwise t-tests to compare the mean threshold values at each SOA with the no-noise condition. This analysis aimed to identify any significant discrepancies between the control data and the experimental results for each participant. To account for the potential risk of inflated false positives due to multiple comparisons, we applied the Bonferroni-Holm correction method (Holm, 1979). This is a sequentially rejective version of the Bonferroni correction for multiple comparisons and strongly controls the family-wise error rate at level alpha. The analysis revealed a significant difference against control for the 50 ms SOA ($t(8) = -5.42$; $p = 0.004$), as depicted in Figure 4.4.

While in previous studies (Contemori et al., 2022; Fan et al., 2016) a disruption for 50 ms SOA was attributed to attention due to an overlap between target and noise, in our study this explanation does not hold since this overlap is not present as the Vernier is displayed for 20 ms only. Thus, this points towards a longer sensitive window in which

the foveal feedback can be disrupted compared to what previously hypothesized. After observing in simple, lower-level stimuli the same phenomenon involved in complex stimulus processing, we designed a second study to tentatively explain the spatial and temporal characteristics of the foveal feedback through a crowding-like paradigm.
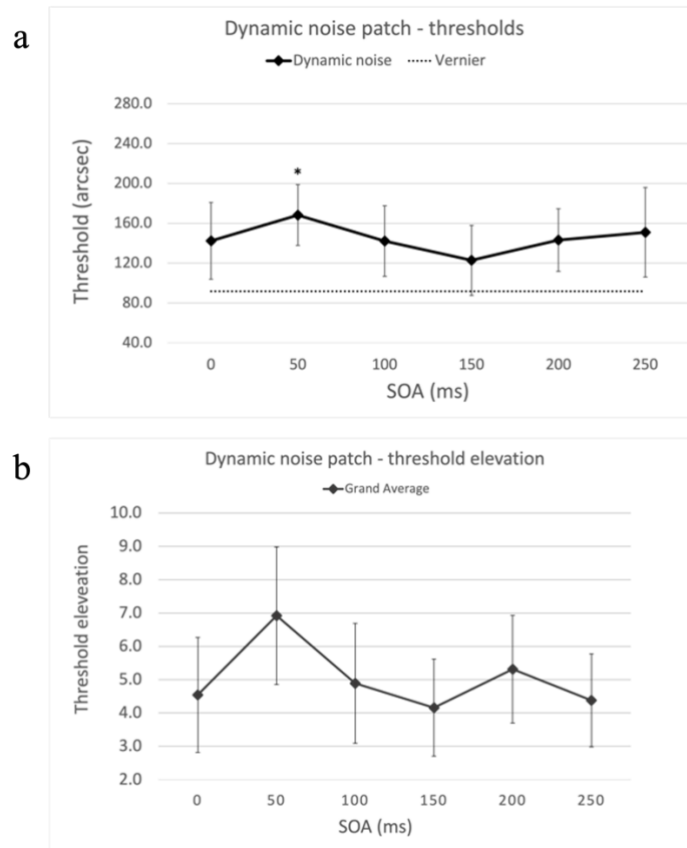

Figure 4.4: Results of Experiment A ($n = 9$). On the top, the mean values of Vernier offset discrimination thresholds with a delayed foveal noise patch. The dashed horizontal line represents the mean threshold for the no-noise baseline condition (a). On the bottom, the mean values of threshold elevation ratio for the same experiment (b). The x-axis shows the different SOA conditions, while on the y-axis the average thresholds are reported in arcsec. Error bars are ±1 Standard Error (SE). There is an increase in threshold values when the noise appears 50 ms after the target onset (T = 168.2).

4.4 Experiment B – vertical flankers and Vernier

4.4.1 Stimuli and procedures

The target stimulus was the same Vernier as in Experiment A (Figure 4.2A). The foveal noise was replaced by a pair of flankers (Figure 4.4). They appeared on both sides of the fixation cross. The flankers had the same length (5040 arcsec) and the same width (108 arcsec) of the whole target Vernier and had a ~400 arcsec (0.11°) gap between each other.
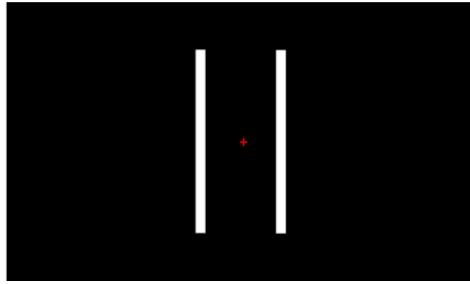
Figure 4.5: The vertical flankers presented in the fovea in Experiment B.

4.4.2 Experimental paradigm

The paradigm was the same employed in Experiment A. The foveal dynamic noise was replaced by two flankers of the same length as the Vernier target. We also employed slightly different SOAs: 0, 50, 150, 250, 350 and 450 ms. There was also a control condition in which only the unmasked Vernier appeared. We decided to raise the maximum SOA value from 250 to 450 in the current experiment. The rationale behind this decision resulted from observations in Experiment A, where there seemed to be a prolonged influence of noise, although not statistically significant, even at the 250 ms mark after target onset. Our aim was to establish that this effect did not persist at even longer SOAs, as its presence could have posed a challenge to the reentrant interaction theory. This theory relies on the absence of noise effects for extended durations (Di Lollo et al., 2000). However, we made the deliberate choice to maintain the total number of SOAs at six, as an increase would have made the task excessively time-consuming and demanding for our participants. As in Experiment A, observers were asked to perform a 2AFC task on the perceived offset between the two lines of the Vernier. Figure 4.6 provides a schematic of the experimental paradigm described above.
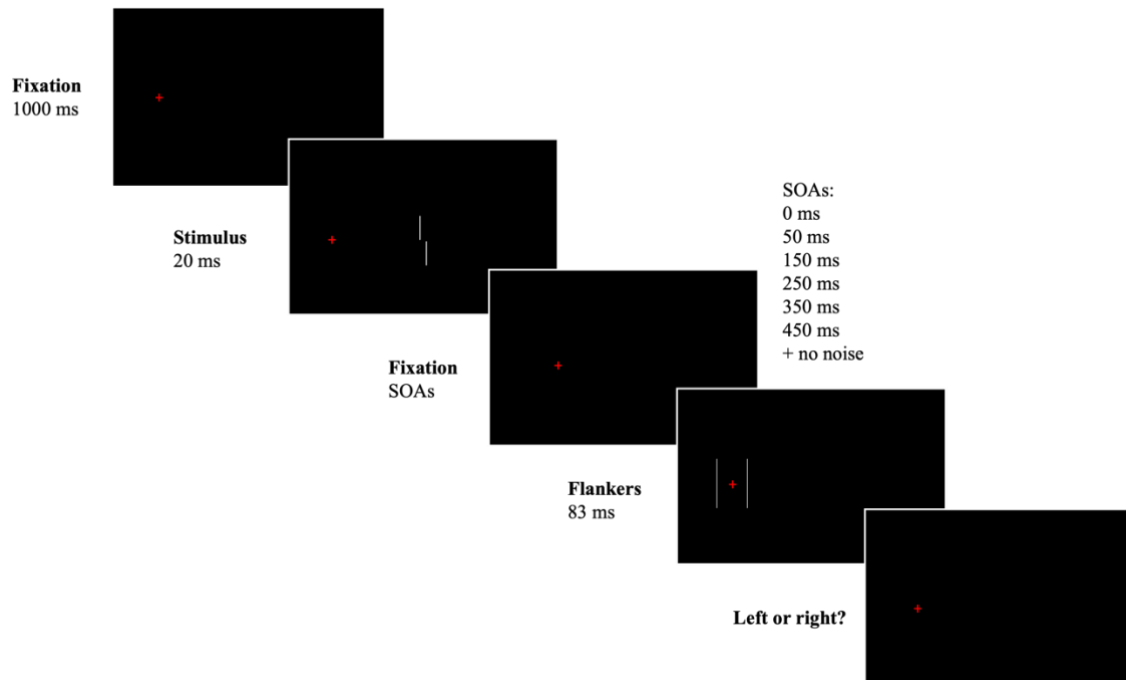
Figure 4.6: Trial structure of experiment B. A red fixation cross appeared on the screen and remained visible throughout the whole experiment. After 1000 ms, the peripheral stimulus was displayed on the right of the fixation cross for 20 ms, 9° or 12° away. Then, the flankers were presented in the fovea, on both sides of fixation. Each block of trials had a different SOA for the flanker presentation, which ranged from 0 to 450 ms. There was also one block in which only the target appeared. Finally, the observer was asked to express the perceived side of the target Vernier offset.

4.4.3 Results

We performed the same analyses employed in Experiment A, employing pairwise Student's t-tests with Bonferroni-Holm correction for multiple comparisons. This time, the analyses revealed a significant difference against control when the flankers appeared 150 ms after the target onset ($t(11) = -3.33$; $p = 0.04$).

Experiment B confirmed the hypothesis of Experiment A and revealed a longer duration of the sensitive window (Figure 4.7). A reduction in performance can be found for the SOA of 150 ms, also for simple foveal noise such as two vertical lines. This suggests that the foveal feedback process is highly sensitive to any kind of stimulus presented in the specific processing window of 50-150 ms after stimulus onset. Indeed, foveal noise presented later than 150 ms does not seem to significantly impact performance, suggesting that the feedback mechanism may be shorter than that. Interestingly, vertical lines appear to induce an even greater impairment in Vernier discrimination than the one found in Experiment A with the dynamic noise patch, while one would expect them to be less effective being a simpler, smaller stimulus. This could be explained by crowding

effects in time, since the spatial characteristics of the vertical lines are closer to the Vernier than the dynamic noise patch and are known to induce strong crowding.
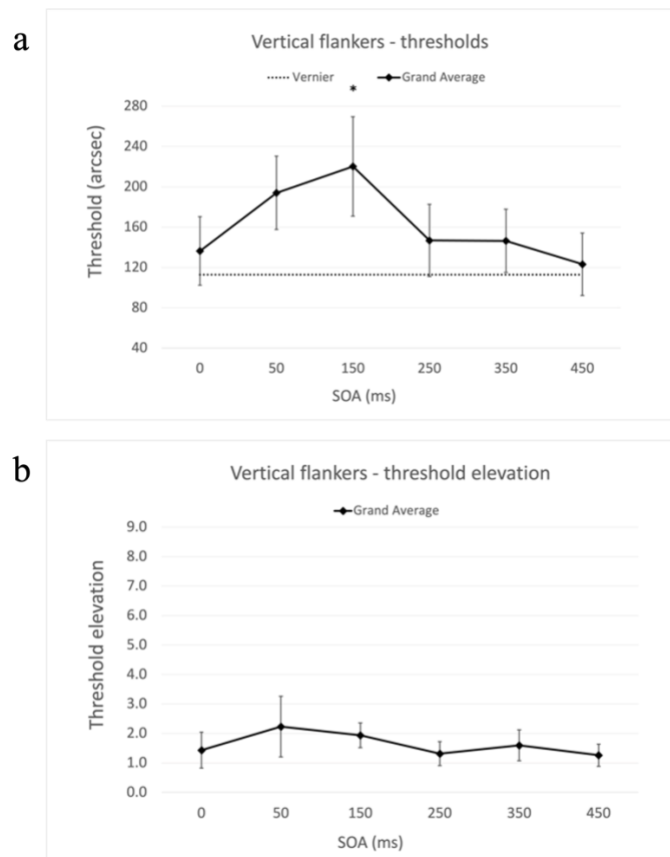

Figure 4.7: Results of Experiment B (*n* = 12). On the left, the mean values of Vernier offset discrimination thresholds with delayed foveal vertical flankers. The dashed horizontal line represents the mean threshold for the no-noise baseline condition (a). On the right, the mean values of threshold elevation ratio for the same experiment (b). The x-axis shows the different SOA conditions, while on the y-axis the average thresholds are reported in arcsec. Error bars are ±1 Standard Error (SE). There is an increase in threshold values when the noise appears 150 ms after the target onset (T = 220).

4.5 Experiment C – horizontal flankers and Vernier

4.5.1 Stimuli and procedures

In Experiment C, participants performed two tasks. The first task involved a reproduction of Experiment B, but with a total of only four SOAs and the no-noise control condition. In the second phase of the experiment, we introduced horizontal flankers, which were generated by rotating the vertical ones by 180 degrees, as illustrated in Figure 4.8.
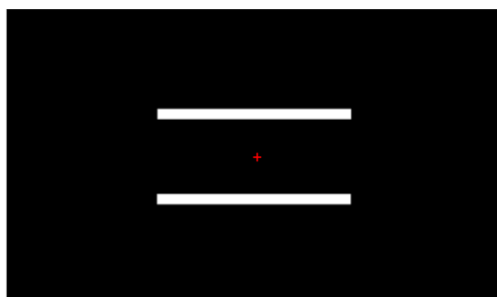
Figure 4.8: The horizontal flankers presented in the fovea in Experiment B. They were the exact same flankers as in Experiment B but rotated 180 degrees relative to the vertical reference.

4.5.2 Experimental paradigm

The experimental setup followed the paradigm used in Experiment B. However, in addition to the vertical foveal flankers, we examined the impact of these same flankers when rotated by 180 degrees. To ensure that the task did not become too long and demanding for the participants, we limited our investigation to four SOAs: 0, 50, 150, and 450 ms. There was also a condition where only the unflanked Vernier was presented. Observers engaged in the same 2AFC task, assessing the perceived offset between the two lines of the Vernier, as was done in Experiment A and B. A schematic representation of this experimental setup is provided in Figure 4.9.
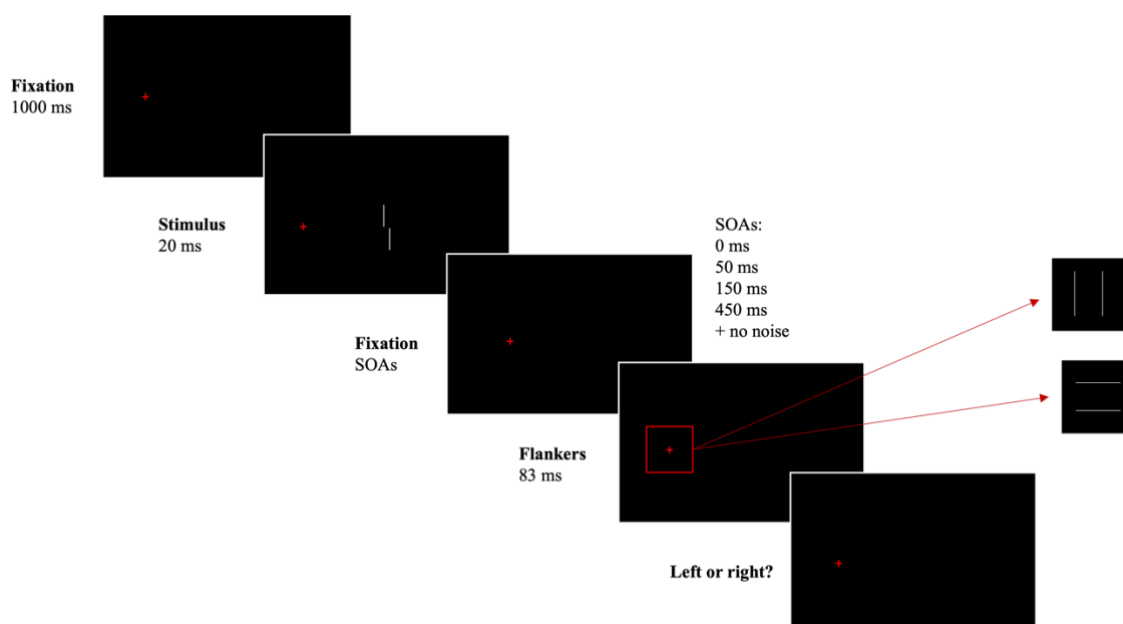
Figure 4.9: Trial structure of experiment C. A red fixation cross appeared on the screen and remained visible for the entire duration of the experiment. After a 1000 ms delay, a peripheral stimulus was presented for 20 ms to the right of the fixation cross, positioned either 9 or 12 degrees away. Subsequently, the flankers were displayed in the central visual field, positioned on both sides of the fixation point. In the first phase of the experiment, the flankers were oriented vertically, while in the second part, they were rotated by 180 degrees. Each block of trials featured a different SOA for the flanker presentation, ranging from 0 to 450 ms. One block only presented the target without flankers. Finally, participants were asked to indicate the perceived side of the offset in the target Vernier.

4.5.3 Results

We conducted the same statistical analyses as in Experiments A and B, employing pairwise Student's t-tests with Bonferroni-Holm correction for multiple comparisons. The results indicated a significant difference from the control condition only when the vertical flankers were presented 50 ms after the target onset ($t(11) = -3.51$; $p = 0.039$).

These findings corroborated our initial hypothesis regarding the specificity of the noise. In fact, there were no significant performance differences detected when the flankers were displayed horizontally. This suggests that when the characteristics of a simple noise, such as two flanking lines, do not align with those of the peripheral target, no significant alterations in performance can be detected.

It is important to highlight that our replication of results from Experiment B was only partial. Notably, the peak in performance with vertical lines was observed earlier, at 50 ms, as opposed to the previous study where it occurred at 150 ms. We attribute this discrepancy to the absence of intermediate SOAs, which would have allowed for a more precise estimation of the threshold peak location. We hypothesize that the peak likely falls somewhere between these two timing points. To conduct a more comprehensive and precise investigation, additional SOAs would be beneficial. However, our desire to maintain a manageable experiment for participants led us to reduce the number of SOAs available.
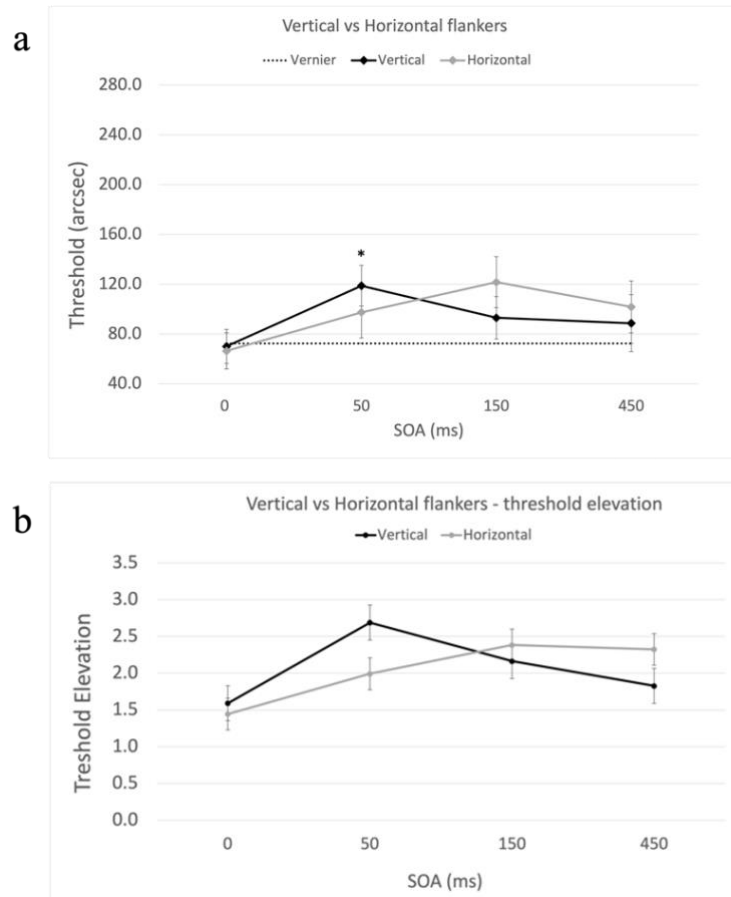
Figure 4.10: Results of Experiment C (*n* = 12). On the top, the mean values of Vernier offset discrimination thresholds with delayed foveal vertical flankers. The black line indicates the results observed with the vertical flankers. The gray line shows performance changes observed with the horizontal flankers. The dashed horizontal line represents the mean threshold for the no-noise baseline condition. On the bottom, the mean values of threshold elevation ratio for the same experiment (b). The x-axis shows the different SOA conditions, while on the y-axis the average thresholds are reported in arcsec. Error bars are ±1 Standard Error (SE). There is an increase in threshold values when the vertical flankers appear 150 ms after the target onset (T = 119.2).

Chapter 5

Discussion

5.1 Summary of findings and discussion of results

5.1.1 Main findings

In the opening chapters of this study, I provided a summary of the key discoveries in the context of foveal feedback and crowding, the two mechanisms central to the experiment we conducted. The investigation of their characteristics, also including the contribution of temporal dynamics of masking to our understanding of crowding, represents an essential step to understand the motivations of this research.

For a long time, crowding implied referred to a phenomenon of information integration over space, in which adjacent stimuli are pooled together, making it difficult to correctly recognize the central target due to the presence of the flankers (e.g., Bouma, 1970; Korte, 1923; Toet & Levi, 1992). Recent evidence has revealed that the spatial properties of crowding are not as fixed as initially thought, but that elements outside Bouma's critical window can affect the perception of a target, as well as the global configuration of the visual scene (Doerig et al., 2019; Herzog et al., 2015; Livne & Sagi, 2007; Manassi et al., 2012, 2013; Saarela et al., 2009).

These aspects of crowding, although innovative, were predominantly associated with the spatial characteristics of these mechanisms. In the same years, fresh perspectives began to emerge also concerning the temporal dimensions of crowding. Initially overlooked, these temporal aspects have proven to be equally significant and have suggested potential parallels between crowding and metacontrast masking (Chakravarthi & Cavanagh, 2009; Chung, 2016; Chung & Patel, 2022; Harrison & Bex, 2014; Y. Yu et al., 2021). Even though the number of studies on the topic is rapidly growing, still little is known about the temporal extent of crowding. Moreover, most of these studies have used letters of the English alphabet as stimuli, as in the classical crowding paradigms. Letters are inherently multifeatured complex stimuli (Greenwood et al., 2009; Nandy & Tjan, 2007; Wolford, 1975) and, to our knowledge, no one has so far investigated the temporal dynamics of crowding with lower-level stimuli, such as the Vernier that we employed in our paradigm.

This paradigm was born out of the idea of combining crowding paradigm with another more recent phenomenon, that of foveal feedback (Williams et al., 2008). While we did not anticipate encountering the same pronounced and detrimental effects of crowding on object detection by placing the flankers in the fovea, we did expect to observe a modest impact of a simpler foveal distractor when introduced at a specific timing. Thus, we tested whether this peak (considering we measured a threshold) is a mere byproduct of the task or if its location within the fovea would indeed influence the performance on a peripheral target.

The findings of the studies described in Chapter 4 hold potential for revealing common traits of crowding, metacontrast masking, and foveal feedback processes. Regarding the outcomes of Experiments A and B, they align with our initial hypothesis, indicating that a large foveal patch can influence the perception of a peripheral target, also when the target exhibits simpler and lower-level visual features. Furthermore, this effect can also be induced by task-relevant, yet simpler distractors, such as vertical flankers placed in the fovea.

In contrast, the results from Experiment C present a less definitive picture, only partially confirming our initial hypotheses. Specifically, we did not observe any significant impact of horizontal flankers on the peripheral target, suggesting that a simple but non-task-relevant distractor does not produce the same disruptive effects as a noise with similar simplicity but task relevance. However, it is worth noting a subtle increase in threshold values around the 150 ms SOA (as shown in Figure 4.10a), even though this observation did not reach statistical significance. Additionally, we replicated the effect observed with vertical lines but with an earlier SOA compared to Experiment B. These outcomes underscore the need for further investigation, which may involve a larger sample size, as the effect size of this phenomenon could be sufficiently small to elude detection with only 12 participants. Exploring a broader range of SOA levels could enable a more precise definition of the critical timing, too.

5.1.2 Interpretation of results

We propose that the phenomena described above may be linked to the concept of cortical re-entrant processing, as outlined in previous studies (Di Lollo et al., 2000; Enns et al., 2000). Initially introduced to explain metacontrast masking, specifically the object substitution type, we believe this theory might also provide a suitable explanation for at

least part of our observations. According to this theory, "the circuit actively searches for a match between a descending code, representing a perceptual hypothesis, and an ongoing pattern of low-level activity" (Enns et al., 2000). Within this framework, disruptions in performance arise when there is a misalignment between the timing of the target's onset and that of the mask. This misalignment gives rise to a conflict, where the information being transmitted to higher-level visual processing contradicts what is concurrently being processed at lower levels. This conflict is described by the authors as an "overwriting operation," wherein the dynamically evolving information at the input level clashes with the information being relayed to higher processing stages (refer to Figure 5.1).

Our findings align with this theory and confirm the presence of a sensitive time window during which incoming information from the mask (or flankers) may overlap with that of the target being fed back to lower levels. Recent evidence related to foveal feedback provides further support to the notion that this mechanism can manifest even when there is a considerable spatial separation between the two sources of visual information, with the target appearing in the periphery of the visual field and the mask displayed in the foveal area. Our results offer additional evidence that this phenomenon indeed occurs, even with lower-level stimuli, encompassing both the mask and the target.

We have reasons to believe that also in this context the foveal retinotopic cortex may function as a high-resolution "scratchpad" where task-relevant information is stored. When foveal feedback is involved, this mechanism may aim to later recruit the stored information to optimize performance in the periphery, where visual acuity and resolution are particularly coarse and limited (Contemori et al., 2022).

We can summarize all these results by suggesting that the visual system might try to project information about peripheral object to an area of the brain that allows for detailed processing of visual objects—namely, the foveal region of the visual cortex. The foveal cortex is engaged not only for complex tasks that require fine detail processing, but also for simple tasks that would ideally need shorter processing times and less heavy computations. A series of studies have emphasized the significant disparities between peripheral and foveal vision, particularly concerning the potential feedback from higher- to lower-level visual areas (Zhaoping, 2017, 2019; Zhaoping & Ackermann, 2018). These studies propose that the lower spatial resolution observed in the peripheral visual field needs compensation through feedback processes originating from areas beyond V1. In

this context, the initial percept can be strengthened (and, therefore, disrupted) by incoming higher-level information (Zhaoping, 2017). Furthermore, they argue that this feedback is weaker or even absent in the periphery (Zhaoping, 2017; Zhaoping & Ackermann, 2018). If a peripheral stimulus lacks detailed representation in higher-level areas, these areas will encounter limitations in providing feedback related to that stimulus. As a result, feedback from peripheral cortical locations to V1 weakens, and it may gradually reduce or vanish as we move downstream along the visual pathway. This notion aligns with the concept of foveal vision compensating for peripheral vision, assuming the role of reinforcing weak peripheral signals within the foveal cortex to facilitate a more robust and stable perception (Williams et al., 2008). These assumptions taken together account for a model of cortical connectivity where reentrant processing assumes a pivotal role in object detection and recognition. These processes operate within precise time windows, and when they are not synchronized, performance disruptions can occur.

Even for a straightforward task such as detecting a Vernier offset with distant flankers, it is plausible that a form of crowding effect over time may be evident, even though with limited intensity and within earlier time windows. This is distinct from the extended time frames, reaching 250 ms as observed in some foveal feedback paradigms (Fan et al., 2016). Instead, our findings are more consistent with temporal crowding paradigms (Chakravarthi & Cavanagh, 2009; Chung, 2016; Chung & Patel, 2022; Harrison & Bex, 2014; Y. Yu et al., 2021), even though most of them have employed letters and strongly manipulated eccentricity as part of their experimental design, which we did not do.
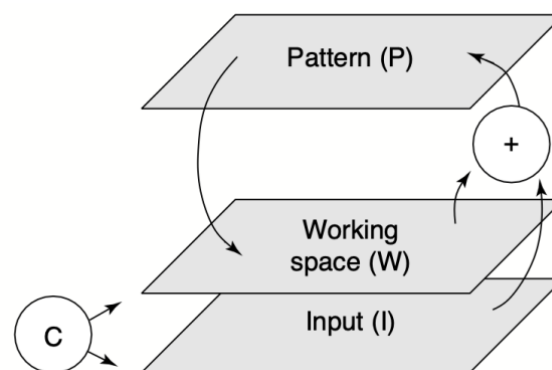
Figure 5.1: Schematic of the computational model for object substitution masking (Di Lollo et al., 2000; Enns et al., 2000). A visual content initially causes activation of all three layers. A weighted sum of the information is produced in the working space (W) once it has received both feedforward

signal from the input layer (I) and feedback information from the pattern layer (P). This allows for a direct comparison (C) between the bidirectional flow of signal.

5.2 Limitations

It is important to acknowledge that this research project was not free from limitations that may have influenced our results. These limitations encompass methodological constraints, sample size considerations, and data collection challenges.

One potential limitation of our study is the fact that we could not confirm whether participants consistently maintained their gaze on the fixation cross rather than foveating the target. However, we tried to limit the effects of this with very brief presentation times to discourage saccades, as the brevity of target presentation makes eye movements impractical, as highlighted by prior research (Weldon et al., 2016).

Furthermore, we believe that some participants might have found our task too easy. We intentionally opted for a relatively simple task to explore the concept of a foveal feedback mechanism within classical crowding paradigms that typically involve lower-level stimuli. Traditional foveal feedback paradigms often incorporate complex and detailed stimuli in object recognition tasks. However, we attempted to mitigate this by employing large eccentricities and brief presentation times.

In Experiment C, a potential limitation is the overlap in representation between the Vernier and the horizontally rotated flankers. The horizontally rotated flankers occupy a position that, if they align precisely with the target, can "cut" into the target representation. This overlapping condition complicates the interpretation of results, as it falls between the complete overlap seen with the dynamic noise patch in Experiment A and the non-overlap observed with the vertical flankers in Experiment A.

Additionally, participants were aware of the target's expected location, in contrast to foveal feedback paradigms where the target's side often changes pseudorandomly. We chose to align with the crowding paradigm, and we already introduced the unpredictable variable of the Vernier offset's side, even though it consistently appeared on the right of the fixation cross.

Further investigation into the role of spacing should also be recommended. Our research has focused on temporal aspects of crowding, perhaps even more than most studies have. While spacing alone is not sufficient to fully explain the complexity of crowding mechanisms, it remains a necessary factor and it is possible that it may have affected the effects we observed. It could be beneficial to explore the manipulation of

eccentricity, as has been undertaken in prior research (Chung, 2016; Chung & Patel, 2022; Harrison & Bex, 2014; Y. Yu et al., 2021), to gain a more comprehensive understanding of these phenomena.

Lastly, the use of different SOA values to suit the specific requirements of each paradigm may have limited our ability to precisely identify the peak timing. This was evident in our observations, where we initially recorded a peak at 150 ms in Experiment B and then at 50 ms in Experiment C. We suggest that the absence of intermediate SOA values might have contributed to this variation and that the peak likely falls between these two values.

While these limitations might influence certain aspects of the interpretation, it is essential to understand their context and potential impacts. We are planning upcoming research activities to address and mitigate these limitations and to enhance the robustness and comprehensiveness of our investigations.

5.3   Future directions

Based on our previous findings, we believe that an additional experiment could clarify some of the aspects that remain unclear in these studies. We would like to extend the paradigm from Experiment B by adding two more conditions that would affect the spatial location of the flankers. We want to test whether any effect on performance can be detected when the flankers are presented in fovea, in the periphery and in the parafovea. This means that, by keeping the presentation of the target in the right periphery, we would also shift the presentation of the flankers in three different spatial locations. The first, in the fovea, has been tested in Experiment B and C of this present study, and revealed an effect of temporal manipulation between 50 and 150 ms. The last, in the periphery enclosing the target, has been assessed by several different crowding studies (e.g., Chicherov et al., 2014; Malania et al., 2007; Manassi et al., 2012; Westheimer, 1975) and is known to have a strong impact of the Vernier. But we do not know yet how it behaves with a temporal delay. The final condition, with the flankers exactly halfway from the fixation and the target, has never been tested before.

This hypothesis should find confirmation in the absence of a discernible peak in thresholds when we position the flankers at an intermediate distance. It would serve as a control to check that the position of the flankers indeed plays a role in altering performance and that the effects from the experiments described here are not an artifact

of basic perceptual activity. It would also be interesting to point out a non-linear relationship between the distance of the flankers and the corresponding change in performance.

Conclusion

In this thesis, I tried to explore the intricate interplay between three mechanisms of visual perception: foveal feedback, crowding, and temporal dynamics. The investigation described has shed light on fundamental mechanisms that underlie the perception of objects in cluttered visual scenes.

Our novel paradigm, crafted to investigate the implications of foveal feedback within crowding paradigms, has produced intriguing findings. While we did not anticipate replicating the pronounced and detrimental effects of crowding on object detection by placing the flankers in the fovea, we expected a subtle impact of a simpler foveal distractor under specific timing conditions. This phenomenon has led us to consider whether this observed threshold peak, indicative of enhanced sensitivity, is a mere task artifact or if its location within the fovea indeed influences peripheral target performance. Our findings from Experiments A and B have reinforced our initial hypothesis, confirming that a foveal distractor can influence the perception of a peripheral target, even when the target exhibits simpler and lower-level visual features. This influence can also be induced by task-relevant, yet simpler distractors, such as vertical flankers placed in the fovea. The results from Experiment C presented a more nuanced picture, only partially confirming our initial hypotheses. As expected, horizontal flankers, which are simple but non-task-relevant distractors, did not produce the same disruptive effects as task-relevant noise with similar simplicity. Nevertheless, there was a subtle increase in threshold values around the 150 ms SOA, albeit not statistically significant, which should be more carefully assessed.

In summary, our research has provided valuable insights into the intricate dynamics of foveal feedback, crowding, and temporal aspects of visual perception. These findings suggest a potential link to the concept of cortical re-entrant processing, with reentrant mechanisms playing a pivotal role in object detection and recognition within precise temporal windows. Even for more straightforward tasks, such as Vernier offset detection with distant flankers, our studies have suggested the presence of temporal crowding effects, albeit with limited intensity and within earlier time windows. These insights expand our understanding of visual perception, offering new avenues for future research and exploration beyond traditional paradigms.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Andriessen, J. J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, *16*(1), 71–78. https://doi.org/10.1016/0042-6989(76)90078-X

Bach, M. (1996). The Freiburg Visual Acuity Test—Automatic Measurement of Visual Acuity. *Optometry and Vision Science*, *73*(1), 49.

Baddeley, A. D. (1988). Imagery and Working Memory. In M. Denis, J. Engelkamp, & J. T. E. Richardson (Eds.), *Cognitive and Neuropsychological Approaches to Mental Imagery* (pp. 169–180). Springer Netherlands. https://doi.org/10.1007/978-94-009-1391-2_15

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Bornet, A., Choung, O.-H., Doerig, A., Whitney, D., Herzog, M. H., & Manassi, M. (2021). Global and high-level effects in crowding cannot be predicted by either high-dimensional pooling or target cueing. *Journal of Vision*, *21*(12), 10. https://doi.org/10.1167/jov.21.12.10

Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature*, *226*(5241), Article 5241. https://doi.org/10.1038/226177a0

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, *62*(8), 1572–1595. https://doi.org/10.3758/BF03212157

Briggs, F. (2020). Role of Feedback Connections in Central Visual Processing. *Annual Review of Vision Science*, *6*(1), 313–334. https://doi.org/10.1146/annurev-vision-121219-081716

Chakravarthi, R., & Cavanagh, P. (2009). Recovery of a crowded object by masking the flankers: Determining the locus of feature integration. *Journal of Vision*, *9*(10), 4–4. https://doi.org/10.1167/9.10.4

Chambers, C. D., Allen, C. P. G., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex*, *49*(1), 327–335. https://doi.org/10.1016/j.cortex.2012.03.007

Chicherov, V., Plomp, G., & Herzog, M. H. (2014). Neural correlates of visual crowding. *NeuroImage*, *93*, 23–31. https://doi.org/10.1016/j.neuroimage.2014.02.021

Chung, S. T. L. (2016). Spatio-temporal properties of letter crowding. *Journal of Vision*, *16*(6), 8. https://doi.org/10.1167/16.6.8

Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast

properties of crowding. *Vision Research*, *41*(14), 1833–1850. https://doi.org/10.1016/S0042-6989(01)00071-2

Chung, S. T. L., & Mansfield, J. S. (2009). Contrast polarity differences reduce crowding but do not benefit reading performance in peripheral vision. *Vision Research*, *49*(23), 2782–2789. https://doi.org/10.1016/j.visres.2009.08.013

Chung, S. T. L., & Patel, S. S. (2022). Spatial and temporal proximity of objects for maximal crowding. *Vision Research*, *194*, 108012. https://doi.org/10.1016/j.visres.2022.108012

Clarke, A. M., Herzog, M. H., & Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. Global and feedforward vs. Feedback distinctions in modeling visual perception. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01193

Contemori, G., Oletto, C. M., Cessa, R., Marini, E., Ronconi, L., Battaglini, L., & Bertamini, M. (2022a). Investigating the role of the foveal cortex in peripheral object discrimination. *Scientific Reports*, *12*(1), 19952. https://doi.org/10.1038/s41598-022-23720-w

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481–507. https://doi.org/10.1037/0096-3445.129.4.481

Doerig, A., Bornet, A., Rosenholtz, R., Francis, G., Clarke, A. M., & Herzog, M. H. (2019). Beyond Bouma's window: How to explain global aspects of crowding? *PLOS Computational Biology*, *15*(5), e1006580. https://doi.org/10.1371/journal.pcbi.1006580

Duangudom, V., Francis, G., & Herzog, M. H. (2007). What is the strength of a mask in visual metacontrast masking? *Journal of Vision*, *7*(1), 7. https://doi.org/10.1167/7.1.7

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, *44*(12), 1321–1331. https://doi.org/10.1016/j.visres.2003.10.024

Enns, J. T., Lollo, V. D., Enns, J. T., & Lollo, V. D. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352. https://doi.org/10.1016/S1364-6613(00)01520-5

Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1022–1033. https://doi.org/10.1037/a0035377

Ester, E. F., Zilber, E., & Serences, J. T. (2015). Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *Journal of Vision*, *15*(1), 4. https://doi.org/10.1167/15.1.4

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of

stimulus orientation, range and of feedback. *Vision Research*, *33*(3), 397–412. https://doi.org/10.1016/0042-6989(93)90094-D

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, *113*(41), 11627–11632. https://doi.org/10.1073/pnas.1606137113

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision*, *9*(6), 18. https://doi.org/10.1167/9.6.18

Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology*, *65*(1), 77–117. https://doi.org/10.1016/j.cogpsych.2012.02.001

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

Francis, G., Manassi, M., & Herzog, M. H. (2017). Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review*, *124*(4), 483–504. https://doi.org/10.1037/rev0000070

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences*, *106*(31), 13130–13135. https://doi.org/10.1073/pnas.0901352106

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Grossberg, S., Mingolla, E., & Ross, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences*, *20*(3), 106–111. https://doi.org/10.1016/S0166-2236(96)01002-8

Harrison, W. J., & Bex, P. J. (2014). Integrating Retinotopic Features in Spatiotopic Coordinates. *Journal of Neuroscience*, *34*(21), 7351–7360. https://doi.org/10.1523/JNEUROSCI.5252-13.2014

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), Article 6598. https://doi.org/10.1038/383334a0

Herzog, M. H. (2022). The Irreducibility of Vision: Gestalt, Crowding and the Fundamentals of Vision. *Vision*, *6*(2), 35. https://doi.org/10.3390/vision6020035

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141. https://doi.org/10.1016/S0042-6989(97)00043-6

Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, *39*(25), 4232–4243. https://doi.org/10.1016/S0042-6989(99)00138-8

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh

look at object recognition. *Current Opinion in Behavioral Sciences*, *1*, 86–93. https://doi.org/10.1016/j.cobeha.2014.10.006

Herzog, M. H., & Sayim, B. (2022). Crowding: Recent advances and perspectives. *Journal of Vision*, *22*(12), 15. https://doi.org/10.1167/jov.22.12.15

Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, *15*(6), 5. https://doi.org/10.1167/15.6.5

Hochstein, S., & Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hu, M. L., Ayton, L. N., & Jolly, J. K. (2021). The Clinical Use of Vernier Acuity: Resolution of the Visual Cortex Is More Than Meets the Eye. *Frontiers in Neuroscience*, *15*, 714843. https://doi.org/10.3389/fnins.2021.714843

Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics*, *66*(4), 584–595. https://doi.org/10.3758/BF03194903

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787. https://doi.org/10.1038/29537

Intriligator, J., & Cavanagh, P. (2001). The Spatial Resolution of Visual Attention. *Cognitive Psychology*, *43*(3), 171–216. https://doi.org/10.1006/cogp.2001.0755

Jehee, J. F. M., Roelfsema, P. R., Deco, G., Murre, J. M. J., & Lamme, V. A. F. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons—Explaining crowding phenomena. *Brain Research*, *1157*, 167–176. https://doi.org/10.1016/j.brainres.2007.03.090

Kammer, T. (2007). Masking visual stimuli by transcranial magnetic stimulation. *Psychological Research*, *71*(6), 659–666. https://doi.org/10.1007/s00426-006-0063-5

Klink, P. C., Dagnino, B., Gariel-Mathis, M.-A., & Roelfsema, P. R. (2017). Distinct Feedforward and Feedback Effects of Microstimulation in Visual Cortex Reveal Neural Mechanisms of Texture Segregation. *Neuron*, *95*(1), 209-220.e3. https://doi.org/10.1016/j.neuron.2017.05.033

Kobayashi, M., & Pascual-Leone, A. (2003). Transcranial magnetic stimulation in neurology. *The Lancet Neurology*, *2*(3), 145–156. https://doi.org/10.1016/S1474-4422(03)00321-1

Kooi, F., Toet, A., Tripathy, S., & Levi, D. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*, 255–279. https://doi.org/10.1163/156856894X00350

Korte, W. (1923). *Über die Gestaltauffassung im indirek- ten Sehen. Zeitschrift für Psychologie*, 93, 17-82.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of Perceived Object Shape by the Human Lateral Occipital Complex. *Science*, *293*(5534), 1506–1509. https://doi.org/10.1126/science.1061133

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, *38*(15–16), 2429–2454. https://doi.org/10.1016/s0042-6989(97)00464-1

Lev, M., & Polat, U. (2015). Space and time in masking and crowding. *Journal of Vision*, *15*(13), 10. https://doi.org/10.1167/15.13.10

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Levi, D. M., & Carney, T. (2009). Crowding in Peripheral Vision: Why Bigger Is Better. *Current Biology*, *19*(23), 1988–1993. https://doi.org/10.1016/j.cub.2009.09.056

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*(2), 167–177. https://doi.org/10.1167/2.2.3

Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research*, *25*(7), 979–991. https://doi.org/10.1016/0042-6989(85)90208-1

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, *25*(7), 963–977. https://doi.org/10.1016/0042-6989(85)90207-X

Levi, D. M., & Waugh, S. J. (1994). Spatial scale shifts in peripheral vernier acuity. *Vision Research*, *34*(17), 2215–2238. https://doi.org/10.1016/0042-6989(94)90104-X

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, *7*(2), 4. https://doi.org/10.1167/7.2.4

Livne, T., & Sagi, D. (2011). Multiple levels of orientation anisotropy in crowding with Gabor flankers. *Journal of Vision*, *11*(13), 18. https://doi.org/10.1167/11.13.18

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, *7*(2), 1. https://doi.org/10.1167/7.2.1

Manassi, M., & Herzog, M. H. (2012). Temporal properties of grouping in crowding. *Journal of Vision*. (Unpublished manuscript)

Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, *16*(3), 35. https://doi.org/10.1167/16.3.35

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10), 13–13. https://doi.org/10.1167/12.10.13

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision*, *13*(13), 10. https://doi.org/10.1167/13.13.10

Manassi, M., & Whitney, D. (2018). Multi-level Crowding and the Paradox of Object Recognition in Clutter. *Current Biology*, *28*(3), R127–R133. https://doi.org/10.1016/j.cub.2017.12.051

Marques, T., Nguyen, J., Fioreze, G., & Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. *Nature Neuroscience*, *21*(5), Article 5. https://doi.org/10.1038/s41593-018-0135-z

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

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), Article 3. https://doi.org/10.1038/nrn1348

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, *10*(7), Article 7. https://doi.org/10.1038/nn1917

Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision*, *7*(2), 5. https://doi.org/10.1167/7.2.5

Oletto, C. M., Contemori, G., Bertamini, M., & Battaglini, L. (2022). The Role of Foveal Cortex in Discriminating Peripheral Stimuli: The Sketchpad Hypothesis. *NeuroSci*, *4*(1), 9–17. https://doi.org/10.3390/neurosci4010002

Op De Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *The Journal of Neuroscience*, *26*(50), 13025–13036. https://doi.org/10.1523/JNEUROSCI.2481-06.2006

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, *24*(3), 436–447. https://doi.org/10.1016/0010-0285(92)90014-S

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive Feedback and Conscious Visual Experience. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00620

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), Article 7. https://doi.org/10.1038/89532

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, *18*(4), 445–451. https://doi.org/10.1016/j.conb.2008.09.008

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 12. https://doi.org/10.1167/4.12.12

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition.

Nature Neuroscience, *11*(10), 1129–1135. https://doi.org/10.1038/nn.2187

Põder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research*, *71*(6), 641–645. https://doi.org/10.1007/s00426-006-0053-7

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2), 23. https://doi.org/10.1167/7.2.23

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, *33*(7), 993–999. https://doi.org/10.1016/0042-6989(93)90081-7

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

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. https://doi.org/10.1038/14819

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*(2), Article 2. https://doi.org/10.1038/nn.2711

Rosen, S., & Pelli, D. G. (2015). Crowding by a repeating pattern. *Journal of Vision*, *15*(6), 10. https://doi.org/10.1167/15.6.10

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, *9*(2), 5–5. https://doi.org/10.1167/9.2.5

Sahar, T., & Yeshurun, Y. (2022). *Temporal Crowding with Central Vision Reveals the Fragility of Visual Representations* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/vmguc

Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung*, *36*(1), 71–93. https://doi.org/10.1007/BF00424655

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual Mental Imagery Induces Retinotopically Organized Activation of Early Visual Areas. *Cerebral Cortex*, *15*(10), 1570–1583. https://doi.org/10.1093/cercor/bhi035

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2. https://doi.org/10.1167/jov.20.12.2

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*(6), 495–508. https://doi.org/10.3758/BF03212183

Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *Journal of Vision*, *13*(1), 24. https://doi.org/10.1167/13.1.24

Taylor, M. M., & Creelman, C. D. (2005). PEST: Efficient Estimates on Probability Functions. *The Journal of the Acoustical Society of America*, *41*(4A), 782–787. https://doi.org/10.1121/1.1910407

Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews Neuroscience*, *21*(4), 231–242. https://doi.org/10.1038/s41583-020-0275-5

Tkacz-Domb, S., & Yeshurun, Y. (2017). Spatial attention alleviates temporal crowding, but neither temporal nor spatial uncertainty are necessary for the emergence of temporal crowding. *Journal of Vision*, *17*(3), 9. https://doi.org/10.1167/17.3.9

Tkacz-Domb, S., & Yeshurun, Y. (2021). Temporal crowding is a unique phenomenon reflecting impaired target encoding over large temporal intervals. *Psychonomic Bulletin & Review*, *28*(6), 1885–1893. https://doi.org/10.3758/s13423-021-01943-8

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*(7), 1349–1357. https://doi.org/10.1016/0042-6989(92)90227-A

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tripathy, S. P., Cavanagh, P., & Bedell, H. E. (2014). Large crowding zones in peripheral vision for briefly presented stimuli. *Journal of Vision*, *14*(6), 11. https://doi.org/10.1167/14.6.11

Vickery, T. J., Shim, W. M., Chakravarthi, R., Jiang, Y. V., & Luedeman, R. (2009). Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision*, *9*(2), 12. https://doi.org/10.1167/9.2.12

Wang, M., Hou, Y., Magrou, L., Autio, J., Misery, P., Coalson, T., Reid, E., Xu, Y., Lamy, C., Falchier, A., Zhang, Q., Poo, M.-M., Dehay, C., Glasser, M., Hayashi, T., Knoblauch, K., Van Essen, D., Shen, Z., & Kennedy, H. (2022). *Retinotopic organization of feedback projections in primate early visual cortex: Implications for active visio*. https://doi.org/10.1101/2022.04.27.489651

Weldon, K. B., Rich, A. N., Woolgar, A., & Williams, M. A. (2016). Disruption of Foveal Space Impairs Discrimination of Peripheral Objects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00699

Weldon, K. B., Woolgar, A., Rich, A. N., & Williams, M. A. (2020). Late disruption of central visual field disrupts peripheral perception of form and color. *PLOS ONE*, *15*(1), e0219725. https://doi.org/10.1371/journal.pone.0219725

Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A source book of Gestalt psychology* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Westheimer, G. (1975). Editorial: Visual acuity and hyperacuity. *Investigative Ophthalmology & Visual Science*, *14*(8), 570–572.

Westheimer, G. (1987). Visual acuity and hyperacuity: Resolution, localization, form. *American Journal of Optometry and Physiological Optics*, *64*(8), 567–574.

https://doi.org/10.1097/00006324-198708000-00002

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research*, *15*(10), 1137–1141. https://doi.org/10.1016/0042-6989(75)90012-7

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research*, *17*(8), 941–947. https://doi.org/10.1016/0042-6989(77)90069-4

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329. https://doi.org/10.3758/BF03194545

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America, A, Optics, Image Science & Vision*, *14*(9), 2057–2068. https://doi.org/10.1364/JOSAA.14.002057

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439–1445. https://doi.org/10.1038/nn.2218

Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, *82*(3), 184–199. https://doi.org/10.1037/0033-295X.82.3.184

Yeshurun, Y., Rashal, E., & Tkacz-Domb, S. (2015). Temporal crowding and its interplay with spatial crowding. *Journal of Vision*, *15*(3), 11. https://doi.org/10.1167/15.3.11

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

Yu, Y., Wu, Q., Dai, Q., Yang, J., Takahashi, S., Ejima, Y., & Wu, J. (2021). The effect of flanker category on crowding is modulated by processing time. *International Journal of Mechatronics and Automation*, *8*(2), 100–109. https://doi.org/10.1504/IJMA.2021.115241

Yu, Y., Wu, Q., Yang, J., Takahashi, S., Ejima, Y., & Wu, J. (2020). Temporal Factors for the Flanker Category Effect under Visual Crowding. *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, 1419–1424. https://doi.org/10.1109/ICMA49215.2020.9233761

Zahabi, S., & Arguin, M. (2014). A crowdful of letters: Disentangling the role of similarity, eccentricity and spatial frequencies in letter crowding. *Vision Research*, *97*, 45–51. https://doi.org/10.1016/j.visres.2014.02.001

Zhang, J.-Y., Zhang, G.-L., Liu, L., & Yu, C. (2012). Whole report uncovers correctly identified but incorrectly placed target information under visual crowding. *Journal of Vision*, *12*(7), 5. https://doi.org/10.1167/12.7.5

Zhaoping, L. (2017). Feedback from higher to lower visual areas for visual recognition

may be weaker in the periphery: Glimpses from the perception of brief dichoptic stimuli. *Vision Research, 136,* 32–49. https://doi.org/10.1016/j.visres.2017.05.002

Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology, 58,* 1–10. https://doi.org/10.1016/j.conb.2019.06.001

Zhaoping, L., & Ackermann, J. (2018). Reversed Depth in Anticorrelated Random-Dot Stereograms and the Central-Peripheral Difference in Visual Inference. *Perception, 47*(5), 531–539. https://doi.org/10.1177/0301006618758571