

Temporal windows of perceptual organization: Evidence from crowding and uncrowding

Alessia Santoni 

Department of Psychology,
Vita-Salute San Raffaele University, Milan, Italy
Department of Psychology and Cognitive Sciences,
University of Trento, Trento, Italy
Department of Psychology,
University of California, Santa Cruz, CA, USA



Luca Ronconi 

Department of Psychology and Cognitive Sciences,
University of Trento, Trento, Italy



Jason Samaha 

Department of Psychology,
University of California, Santa Cruz, CA, USA



Organizing visual input into coherent percepts requires dynamic grouping and segmentation mechanisms that operate across both spatial and temporal domains. Crowding occurs when nearby elements interfere with target perception, but specific flanker configurations can alleviate this effect through Gestalt-based grouping, a phenomenon known as uncrowding. Here, we examined the temporal dynamics underlying these spatial organization processes using a Vernier discrimination task. In Experiment 1, we varied stimulus duration and found that uncrowding emerged only after 160 ms, suggesting a time-consuming process. In Experiment 2, we manipulated the stimulus onset asynchrony (SOA) between the target and flankers. We found that presenting good-Gestalt flankers briefly before the target (as little as 32 ms) significantly boosted uncrowding, even in the absence of temporal overlap between the two stimuli. This effect was specific to conditions in which flankers preceded the target, ruling out pure temporal integration and masking accounts. These findings suggest that spatial segmentation can be dynamically facilitated when the temporal order of presentation allows grouping mechanisms to engage prior to target processing. Moreover, the observed time course indicates that segmentation is not purely feedforward, particularly for stimuli that are likely to recruit higher level visual areas, pointing instead to the involvement of recurrent or feedback processes.

Introduction

In order to organize incoming sensory input, the visual system must identify and group elements across space, as well as integrate information across time. This dual challenge is central to defining coherent and meaningful perceptual units, yet the mechanisms that support it remain only partially understood.

In the spatial domain, the presence of nearby elements in the visual scene can impair target perception, a phenomenon known as crowding. Crowding is a pervasive feature of visual processing, occurring across different levels of stimulus complexity, from simple Gabor patches to alphanumeric characters and natural scenes (e.g., Bernard & Chung, 2011; Ringer, Coy, Larson, & Loschky, 2021; Ronconi, Bertoni, & Bellacosa Marotti, 2016; Tanriverdi & Cornelissen, 2024). Previous research has identified object spacing as the key factor driving crowding: When the distance between the target and flankers falls below an observer's critical spacing at that retinal location, crowding is likely to occur. This principle, originally described by Bouma in 1970 (Bouma, 1970), has since become known as "Bouma's law" (Pelli & Tillman, 2008). Additionally, the strength of crowding is shaped by flanker–target similarity, with target identification errors often reflecting either a bias toward flanker features or mislocalization of the target relative to the flankers (Bernard & Chung, 2011; Pöder & Wagemans, 2007; Zahabi & Arguin, 2014). This and other evidence has contributed to the view that crowding reflects a

Citation: Santoni, A., Ronconi, L., & Samaha, J. (2025). Temporal windows of perceptual organization: Evidence from crowding and uncrowding. *Journal of Vision*, 25(14):5, 1–9, <https://doi.org/10.1167/jov.25.14.5>.

<https://doi.org/10.1167/jov.25.14.5>

Received July 21, 2025; published December 11, 2025

ISSN 1534-7362 Copyright 2025 The Authors



perceptual bottleneck, arising from feature integration or pooling within spatial regions defined by receptive field sizes in the feedforward visual hierarchy (for a comprehensive characterization of models of crowding, refer to Pelli, 2008).

This account of crowding has been challenged in the past decade by a notable exception in which specific flanker configurations alleviate the otherwise detrimental effects of crowding. This effect, known as *uncrowding*, is often attributed to Gestalt grouping principles, whereby flanking elements are perceptually organized into unitary structures based on their features (Malania, Herzog, & Westheimer, 2007; Manassi, Sayim, & Herzog, 2012; Manassi, Sayim, & Herzog, 2013; Sayim, Westheimer, & Herzog, 2010). The spatial arrangement of flankers determines whether they are perceived as part of a unified object with the target or as separate elements, affecting the degree of crowding accordingly (Livne & Sagi, 2007; Saarela, Westheimer, & Herzog, 2010; Tiurina, Markov, Choung, Herzog, & Pascucci, 2022). Unlike crowding, which has been often considered an early, feedforward limitation, uncrowding may reflect a time-dependent process of perceptual organization emerging through recurrent or integrative mechanisms over time (Herzog, Sayim, Chicherov, & Manassi, 2015; Manassi & Whitney, 2018).

In the temporal domain, crowding effects can still be observed when flankers are not presented simultaneously to the target. For example, Huckauf and Heller (2004) manipulated the stimulus onset asynchrony (SOA) between target and flanker letters and found that strong crowding effects emerged at short temporal delays and persisted even when target and flankers were separated by longer time-windows of ± 150 ms. However, a precise temporal characterization of uncrowding is still lacking. Sayim, Manassi, and Herzog (2014) investigated the interaction between backward masking (i.e., when the target precedes the flankers) and (un)crowding by manipulating both the appearance of the flankers and the SOA between target and flankers. They found that flankers can produce type-B backward masking, with the greatest performance impairments occurring at SOAs of approximately 40 to 60 ms, depending on stimulus characteristics. Notably, the study did not reveal any perceptual advantage when flankers formed a good Gestalt (e.g., cuboids) compared to typical crowding-inducing flankers (e.g., vertical lines), suggesting that Gestalt-based representations may require additional time to emerge and be the results of higher level feedback mechanisms (Sayim et al., 2014). This interpretation is also supported by recent results by Morea, Herzog, Francis, and Manassi (2025), who showed that, although uncrowding effects typically emerged only after 160 ms under standard simultaneous presentation, even a brief 20-ms preview of the good-Gestalt flankers prior to the full flanker-target

display significantly improved Vernier discrimination performance.

The aim of the present study was to investigate temporal dynamics underlying spatiotemporal organization in visual perception. To characterize the temporal dynamics of crowding and uncrowding, we conducted two experiments using a Vernier discrimination task. In the first study, we manipulated stimulus duration while presenting flankers known to elicit either crowding or uncrowding effects. In the second study, we introduced a temporal delay between the flankers and the target Vernier stimulus, such that flankers (inducing either crowding or uncrowding) could appear before or after the target. This design allowed us to systematically map both the directionality and timing of uncrowding effects. Our results first replicated previous findings showing that grouping effects typically emerge relatively slowly. However, this delay was drastically reduced when flankers were presented before the target. Additionally, performance in the two tasks was highly correlated, suggesting that individual accuracy was reliably mapped across different temporal manipulations of the stimuli. We interpret these findings in light of feedback or recurrent mechanisms, suggesting that early presentation of the flankers pre-activates these processes, leading to a marked reduction in the time required for the uncrowding benefit to emerge.

Methods

Participants

Twenty-four participants were recruited from the University of California, Santa Cruz, community. Two participants were excluded due to uncorrected visual impairments, resulting in a final sample of 22 participants (mean age, 22 years; range, 18–35). The final sample included 16 individuals who identified as female, five as male, and one as non-binary. Participants received university research credits as compensation, if interested. The study was approved by the Institutional Review Board at the University of California, Santa Cruz. All participants performed two experiments, the order of which was counterbalanced across participants.

Apparatus and stimuli

Participants were seated in a dimly lit room at a viewing distance of 75 cm, maintained with a chin rest, and completed two tasks, each lasting approximately 25 minutes. Stimuli were presented through Psychtoolbox 3 (Brainard, 1997) for MATLAB (MathWorks, Natick,

MA) on a black background on a VIEWPixx/EEG monitor (VPixx Technologies, Saint-Bruno, QC, Canada) with a 120-Hz refresh rate. Stimulus luminance was set to 80 cd/m². In a Vernier discrimination task, Vernier targets consisted of two vertical lines, with a slight horizontal offset in the lower line. Following an intertrial interval of 500 ms (± 250 ms jitter), Vernier stimuli were randomly presented to either the left or the right of the central fixation dot. The center of the stimulus was positioned 4 degrees of visual angle (DVA) from fixation, and participants were instructed to maintain central fixation throughout the trial. The two vertical lines composing the Vernier stimulus measured 40 arcminutes (') in length each and were separated by a vertical gap of 4'. The horizontal offset between the two lines was set to 1', and offset direction was randomized across trials. Across three experimental blocks, Vernier stimuli were presented either in isolation (baseline condition) or flanked by additional elements (illustrated in Figure 1A). Flankers, when present, were presented symmetrically to the Vernier stimulus along the horizontal axis at a distance of 16'. Flankers could be either vertical lines (44' in length) or rectangular flankers (44' tall \times 116' wide), previously shown to elicit crowding and uncrowding effects, respectively (Manassi et al., 2012). Participants were instructed to indicate the direction of the offset of the lower line relative to the upper line by pressing "M" (right) or "Z" (left) on the keyboard.

Procedure

Experiment 1

Participants were asked to report the offset direction of the Vernier stimulus while stimulus duration was manipulated along the following values: 32, 64, 80, 96, 128, 160, or 200 ms. Vernier and flanker stimuli were always presented simultaneously. The task included a total of 588 trials: 196 baseline trials and 196 trials for each flanker shape (lines and rectangles).

Experiment 2

Participants performed a Vernier discrimination task where the duration of both the Vernier and flanker stimuli was fixed at 40 ms. However, target and flanker stimuli were presented either simultaneously (i.e., a SOA of 0 ms) or with a variable SOA, randomly selected from ± 32 , ± 64 , ± 80 , ± 96 , ± 128 , or ± 160 ms (Figure 1B). Negative SOA values indicate trials in which the flankers preceded the Vernier stimulus, whereas positive SOA values indicate trials in which the flankers followed the Vernier stimulus. The task included a total of 616 trials: 56 baseline trials and 280 trials for each flanker shape (lines and rectangles).

For both experiments, participants completed three separate blocks of trials: one for the baseline condition (no flankers), one with line flankers, and one with rectangle flankers. Block order was counterbalanced across participants. Before each experiment, participants completed 20 practice trials with feedback to familiarize themselves with the procedure.

Statistical analyses

For Experiment 1, we investigated the effect of stimulus duration on Vernier discrimination. With this aim, stimulus duration (seven levels: 32, 64, 80, 96, 128, 160, and 200 ms) and flanker shape (lines vs. rectangles) were input as within-subject factors in a two-way repeated-measures analysis of variance (ANOVA) on accuracy data. For Experiment 2, a two-way repeated-measures ANOVA was conducted on accuracy data to examine the interaction between flanker shape (lines vs. rectangles) and stimulus SOA (12 levels: ± 32 , ± 64 , ± 80 , ± 96 , ± 128 , and ± 160 ms). Sphericity was assessed using Mauchly's test, and Greenhouse–Geisser correction was applied where necessary. For each ANOVA, post hoc *t*-tests were performed to further investigate the relationship between flanker shape and either stimulus duration or SOA. In the second experiment, additional post hoc *t*-tests were conducted to compare accuracy at same-magnitude positive and negative SOAs for each flanker shape separately. As a control analysis to account for possible effects of attentional capture based on flanker configuration in the second experiment, we conducted an additional repeated-measures ANOVA on median reaction times (RTs) with SOA and flanker configuration as within-subject factors. Any difference in median RTs between flanker configuration (either a main effect or an interaction with SOA) might reflect a differential engagement of exogenous attention based on good-Gestalt stimulus properties. False discovery rate (FDR) (Benjamini & Hochberg, 1995) was used to control for the family-wise type-I error rate across multiple comparisons. For each ANOVA, effect sizes were quantified using generalized eta squared (GES), and effect sizes for post hoc *t*-tests were reported as the absolute value of Cohen's *d*. Accuracy data from the baseline condition were not included in the ANOVAs; instead, they were examined to ensure that participants performed above chance level.

Finally, two-tailed Pearson's correlation analyses were conducted to assess whether individual accuracy scores based on flanker shape were reliably measured across the two temporal manipulations (stimulus duration vs. SOA). Correlations were performed on mean accuracy scores obtained for each subject and for each flanker configuration by averaging across all stimulus durations in Experiment 1 (32–200 ms) and SOA levels for Experiment 2 (± 32 to ± 160 ms).

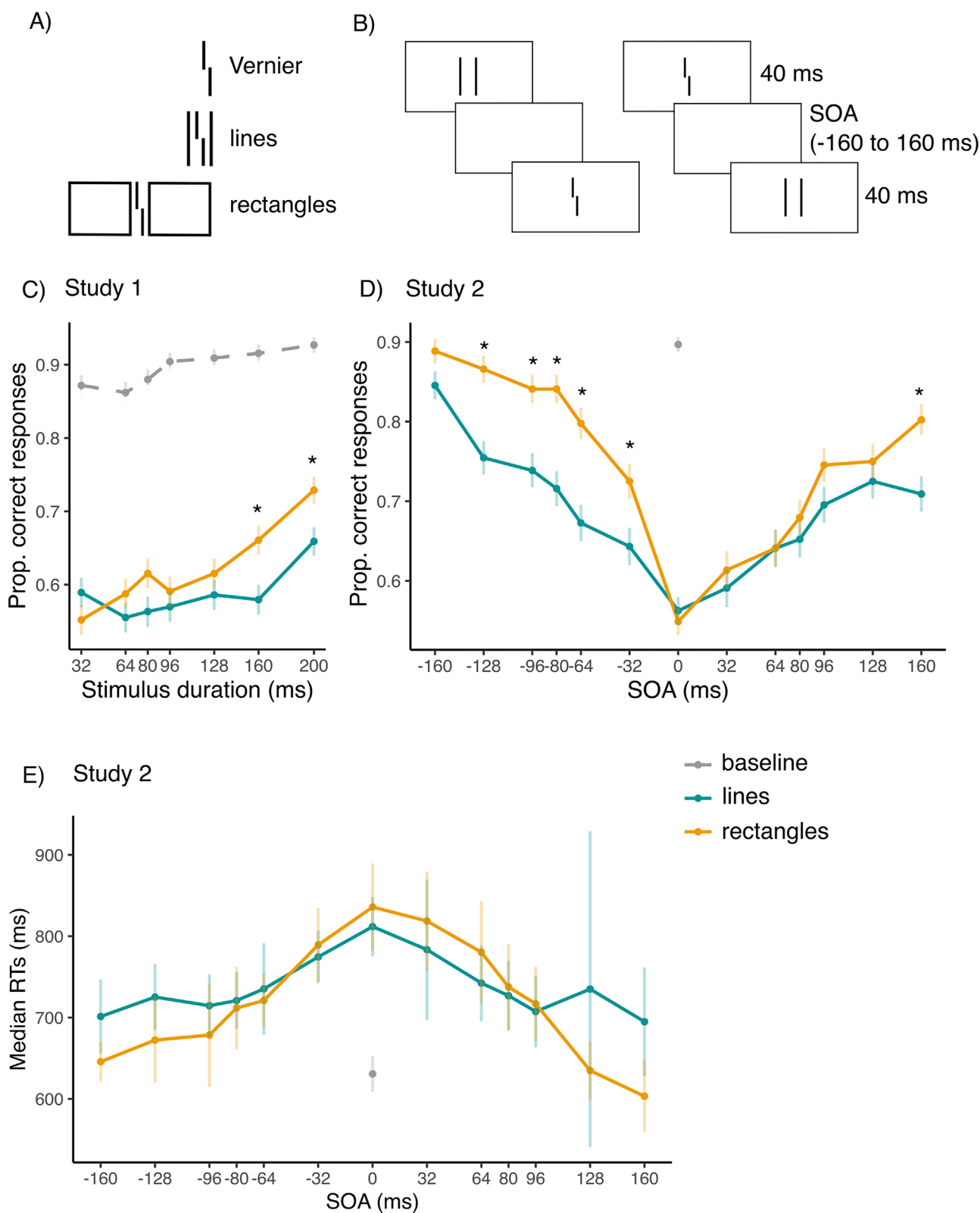


Figure 1. (A) Stimuli used in the Vernier discrimination task. Vernier stimuli were either presented in isolation (baseline condition) or flanked by lines or rectangles across separate blocks of trials. (B) Schematic representation of Experiment 2. A temporal delay was introduced between the flankers and the Vernier target, such that flankers either preceded (left panel) or followed (right panel) the Vernier. Stimuli are enlarged for illustrative purposes. On each trial, stimuli were randomly presented to the left or right of a central fixation point. (C, D) Results from Experiments 1 and 2 show that the uncrowding effect emerged around 160 ms when stimuli were presented simultaneously, but at a SOA of 32 ms when flankers preceded Verniers. Asterisks represent p values from t -tests

←
comparing accuracy for flanker shape (lines vs. rectangles). (E) In Experiment 2, median RTs showed no modulation as a function of flanker type (lines vs. rectangles). Error bars represent standard error of the mean (*SEM*).

Results

Experiment 1

Results from Experiment 1 are depicted in [Figure 1C](#). A two-way repeated-measures ANOVA was conducted to examine the effects of stimulus duration (seven levels: 32, 64, 80, 96, 128, 160, and 200 ms) and flanker shape (lines vs. rectangles) on Vernier discrimination accuracy. The analysis revealed a significant main effect of stimulus duration, $F(3.47, 72.89) = 8.12$, $p < 0.001$, $GES = 0.11$, and a significant main effect of flanker shape, $F(1, 21) = 6.63$, $p = 0.018$, $GES = 0.02$. Importantly, the analysis showed a significant interaction between stimulus duration and flanker shape, $F(4.60, 96.70) = 2.51$, $p = 0.039$, $GES = 0.02$, suggesting that the effect of duration on accuracy differed between the two flanker shapes. As our main goal was to assess *when* the differences emerged, we conducted post hoc pairwise comparisons between the two flanker shapes at each stimulus duration. This revealed significantly higher accuracy in the presence of rectangular flankers compared with flanker lines at 160 ms, $t(21) = -2.77$, $p_{FDR} = 0.040$, $d = 0.59$ and 200 ms, $t(21) = -2.83$, $p_{FDR} = 0.040$, $d = 0.60$, respectively. No significant differences were found at shorter durations (32–128 ms; all $p > 0.08$).

Experiment 2

Results from Experiment 2 are depicted in [Figure 1D](#). The repeated-measures ANOVA on flanker shape (lines vs. rectangles) and SOA (12 levels: ± 32 , ± 64 , ± 80 , ± 96 , ± 128 , and ± 160 ms) revealed a significant main effect of SOA, $F(6.87, 144.18) = 32.96$, $p < 0.001$, $GES = 0.31$, and a significant main effect of flanker shape, $F(1, 21) = 22.88$, $p < 0.001$, $GES = 0.06$. A significant interaction between SOA and flanker shape was found, $F(7.69, 161.51) = 3.00$, $p = 0.004$, $GES = 0.03$. Post hoc tests comparing accuracy for flanker shapes at each SOA revealed significantly higher accuracy for rectangular flankers across negative SOAs between -128 and -32 ms, as well as at the positive SOA of 160 ms (see [Table 1](#) for detailed statistics). No significant differences were found at 0 ms or positive SOAs up to 128 ms ($p_{FDR} > 0.10$). Post hoc *t*-tests revealed that Vernier accuracy significantly differed between negative and positive SOAs of the same magnitude across all SOA levels when rectangular flankers were presented

Comparison (SOA)	<i>t</i> (21)	<i>P</i> (FDR-corrected)	Cohen's <i>d</i>
-128 ms	-3.99	0.003	0.85
-96 ms	-3.70	0.004	0.79
-80 ms	-4.65	<0.001	0.99
-64 ms	-3.38	0.006	0.72
-32 ms	-3.37	0.006	0.72
0 ms	0.66	0.56	0.14
32 ms	-0.72	0.56	0.15
64 ms	0	1	<0.001
80 ms	-0.81	0.56	0.18
96 ms	-2.01	0.11	0.43
128 ms	-0.68	0.56	0.15
160 ms	-4.64	<0.001	0.98

Table 1. Results from post hoc *t*-tests comparing accuracy for flanker shape (lines vs. rectangles) at each SOA level in Experiment 2. Bold terms indicate significant *P* values after FDR correction ($P < 0.05$).

Comparison (SOA)	<i>t</i> (21)	<i>P</i> (FDR-corrected)	Cohen's <i>d</i>
-160 vs. 160 ms	3.82	0.002	0.81
-128 vs. 128 ms	3.57	0.002	0.76
-96 vs. 96 ms	3.57	0.002	0.76
-80 vs. 80 ms	6.81	<0.001	1.45
-64 vs. 64 ms	4.50	<0.001	0.96
-32 vs. 32 ms	3.21	0.004	0.69

Table 2. Results from post hoc *t*-tests comparing accuracy between negative and positive SOA levels when presenting a rectangular flanker shape in Experiment 2. Bold terms indicate significant *P* values after FDR correction ($P < 0.05$).

(see [Table 2](#) for detailed statistics). The same post hoc *t*-tests performed when line flankers were presented revealed a significant difference only at 160 ms, $t(21) = 5.68$, $p_{FDR} < 0.001$, $d = 1.21$ (all other $p > 0.15$).

As a control analysis for potential confound effects of attentional capture across conditions, we ran a repeated-measures ANOVA on median RTs with SOA and flanker shape as within-subject factors. The ANOVA revealed a main effect of SOA, $F(3.77, 79.16) = 8.77$, $p < 0.001$, $GES = 0.02$, but no main effect of flanker shape, $F(1, 21) = 0.17$, $p = 0.68$, $GES = 0.068$, nor a flanker shape \times SOA interaction, $F(4.13, 86.69) = 1.87$, $p = 0.12$, $GES = 0.12$, indicating no difference in attentional capture across conditions ([Figure 1E](#)).

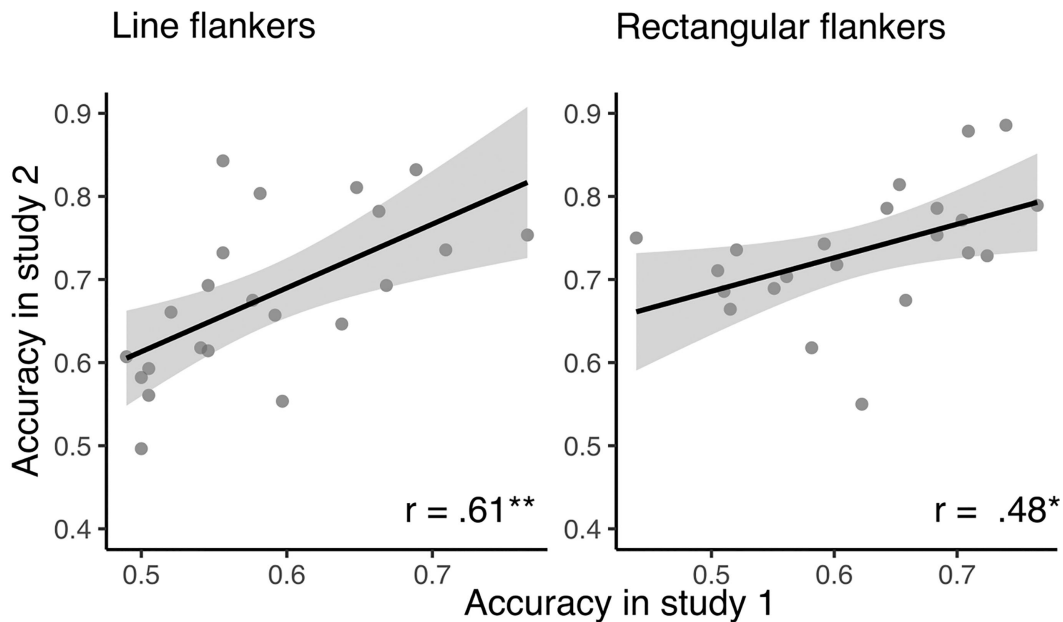


Figure 2. Pearson's correlations between mean accuracy in Experiments 1 and 2, conducted separately for flanker configuration (left panel, line flanker blocks; right panel, rectangular flanker blocks). Dots represent individual subjects; gray shaded areas represent 95% confidence intervals. * $p_{FDR} < 0.05$, ** $p_{FDR} < 0.01$.

Correlations

Accuracy scores averaged for stimulus duration/SOA level across the two experiments were positively correlated for both line flanker blocks, $r(20) = 0.61$, $p = 0.003$, and rectangular flanker blocks, $r(20) = 0.48$, $p = 0.025$, indicating that accuracy was reliably measured across the two tasks and flanker configurations (Figure 2).

Discussion

Perception depends on dynamic grouping and segmentation processes that unfold across both space and time. Here, we extended previous work by examining the time course of spatial segmentation using a Vernier discrimination paradigm known to produce strong crowding and uncrowding effects depending on the flanker configuration. Rectangular flankers, in particular, have been shown to alleviate crowding compared to simple vertical lines, as they are perceptually grouped into a single unit under Gestalt principles (Herzog et al., 2015; Manassi et al., 2012). Whereas earlier studies have largely focused on how flanker properties influence spatial grouping (e.g., Manassi et al., 2012; Manassi et al., 2013), our focus was on when these effects emerge during stimulus processing. In our first experiment, we mapped the time

frame of the uncrowding benefit by varying stimulus duration across a broad range (20–200 ms) and found that uncrowding emerged only when the stimulus was presented for at least 160 ms. This replicates previous findings by Morea et al. (2025) and Doerig et al. (2020), who reported uncrowding effects emerging at 160 ms. Together, these results suggest that spatial segmentation and grouping build up over time, likely via either recurrent or feedback processing (Herzog et al., 2015). However, a difference from previous studies is that in our paradigm stimulus location was randomized across hemifields, minimizing stimulus expectancy and preparatory spatial attention. Despite this, we observed that the same temporal window was necessary for uncrowding to emerge, indicating that the time course of spatial segmentation is likely not contingent on stimulus location predictability.

Our first experiment confirmed that spatial organization unfolds over relatively long time scales, but our second experiment demonstrated that, when flankers precede the target, the uncrowding effect emerges more rapidly. In this experiment, we systematically varied the SOA between the Vernier and the flankers, both presented for 40 ms, across a broad temporal range. First, we showed that, even when flankers and Verniers are not presented simultaneously, crowding and uncrowding effects are still observed, demonstrating the temporal extent of spatial segmentation mechanisms. Furthermore, we found uncrowding benefits emerging at SOAs as early as 32 ms and persisting up to 128 ms when flankers

preceded the target. Beyond this point, the temporal separation between the two stimuli was large enough that the temporal crowding effect was minimal to begin with. At an SOA of -32 ms, the total visual stimulation amounted to 72 ms, about half the presentation duration expected to elicit uncrowding under the conditions of our first experiment. This suggests that Gestalt-based uncrowding can emerge much faster than previously thought when there is a brief temporal offset between the flanker and target.

Moreover, an SOA of -32 ms involved an 8-ms temporal overlap between flankers and the Vernier, but similar uncrowding effects were also observed at longer negative SOAs without any overlap (i.e., SOAs < -64 ms). For example, even at an SOA of -128 ms, corresponding to an 88-ms gap between flanker offset and Vernier onset, significant uncrowding still occurred. This highlights the temporal stability of Gestalt grouping, which can alleviate crowding of subsequent targets even after a substantial delay. Interestingly, we also observed an unexpected significant effect of uncrowding at an SOA of $+160$ ms; however, because nearby SOAs did not show a similar effect it is unclear how robust this difference is.

It is important to note that, although we refer to the results in Experiment 2 in the context of (un)crowding mechanisms, the temporal delay introduced between flankers and the Vernier significantly deviates with classic (un)crowding paradigms. Consequently, other factors may contribute to the observed effects, and it is possible that the results in Experiment 2 were driven by additional influences of temporal integration and masking. By introducing this temporal offset, the task allowed us to investigate temporal integration windows (TIWs) across different perceptual conditions; for example, we showed that uncrowding stimuli define a narrower TIW as compared to crowding stimuli (see the width of the curve in Figure 1D). However, it is important to note that the observed uncrowding benefits cannot be solely attributed to temporal integration mechanisms. Indeed, this effect was specific to negative SOAs (i.e., when flankers preceded the target), implicating rapid preview mechanisms that extend beyond simple temporal integration of the two stimuli. If temporal integration were the only mechanism at play, the uncrowding benefit should have been comparable for positive and negative SOAs of equal magnitude.

On the other hand, our results are also only partially compatible with masking mechanisms. Asymmetries between forward and backward masking are well documented, with backward masking typically exerting a stronger disruptive effect on perception (Enns & Di Lollo, 2000). Using forward- and backward-masked Vernier targets embedded in a stream of crowding-inducing lines, Hermens et al. (2009) showed that backward masking (i.e., when the mask followed the

target) led to a greater increase of Vernier thresholds (i.e., worse performance). However, our findings indicate that crowding strength was comparable for positive and negative SOAs of equal magnitude when crowding stimuli were presented (with the exception of the SOA equal to 160 ms). Conversely, we found asymmetry between positive and negative SOAs of equal magnitude using uncrowding stimuli. According to the masking account, backward masking effects are often interpreted as arising from temporally delayed feedback from higher to lower level visual processes, which can disrupt the perception of the target stimulus (Di Lollo, Enns, & Rensink, 2000). In line with this idea, Huckauf and Heller (2004) showed that crowding was reduced when targets were followed by letter-like nonletters as opposed to letter flankers. They attributed this effect to reduced top-down activation: Because letter-like nonletters engage the letter level more weakly, they generate less feedback to early visual areas, thereby weakening the crowding effect. Additional results investigating backward masking and stimulus configurations come from Sayim et al. (2014), who reported stronger masking at SOAs around 40 to 60 ms (i.e., type-B masking) regardless of flanker configuration (i.e., flankers could be either lines or cuboids). This lack of uncrowding effect for backward masking is consistent with our results at positive SOAs, further corroborating the idea that Gestalt representations might operate on feedback processes that require additional processing time (Wagemans et al., 2012). However, in our study, performance was maximally impaired at short and zero SOAs rather than intermediate SOAs as in Sayim et al. (2014).

We suggest that when flankers form good Gestalts spatial grouping processes can rapidly segment flankers from targets and do so even in the absence of temporal overlap between the stimuli. It has been previously hypothesized that crowding occurs when stimuli fall within the same processing stage (Manassi & Whitney, 2018). In this view, preactivating relevant Gestalt structures may engage top-down/lateral spatial segmentation mechanisms that facilitate subsequent target processing. In our study, when the target preceded the flankers, no such advantage was observed, further arguing against a purely masking-based account.

Importantly, our conclusions about the involvement of feedback or recurrent mechanisms are likely to apply specifically to the relatively complex flanker configurations used in our paradigm, which presumably engage higher level visual areas responsible for spatial grouping across objects and shapes. It remains an open question whether similar temporal dynamics would be observed with simpler stimuli that rely predominantly on early visual processing stages, where crowding might still arise from more purely feedforward mechanisms.

Finally, it is possible that exogenous attention contributed to the behavioral improvement observed at negative SOAs compared with their positive counterparts. Previous studies have shown that both endogenous and exogenous attention can modulate the strength of crowding (Gong, Liu, Chen, & Sun, 2024; Kewan-Khalayly & Yashar, 2022; Yeshurun & Carrasco, 1999) and that objects organized according to Gestalt principles can attract exogenous attention faster (Kimchi, Yeshurun, & Cohen-Savransky, 2007). However, our control analysis performed on RTs showed no significant difference between conditions, suggesting that any attentional effects in our paradigm were equally present in the two flanker configurations and therefore cannot fully account for the uncrowding effects observed here.

Conclusions

Our results demonstrate that spatial segmentation, although relatively slow under standard uncrowding conditions, can be significantly accelerated when uncrowding flankers are presented prior to the target stimulus. We interpret this finding as evidence that a rapid presentation of the flankers, even for just tens of milliseconds, facilitates the emergence of Gestalt-based benefits by initiating grouping mechanisms that have a sustained impact on subsequent target perception. In turn, these results offer further support for the idea that spatial segmentation and grouping can be mediated by recurrent or feedback processing in the case of relatively complex configurations that likely engage higher level visual areas.

Keywords: crowding, perceptual organization, Gestalt grouping, vernier discrimination

Acknowledgments

The authors thank Michael Herzog and Martina Morea for the helpful discussions. We also thank Emily Lincoln, Vrishab Nukala, and Gaia Minari for their help in data collection. The present work was performed by A.S. in fulfillment of the requirements for obtaining a PhD degree at Vita-Salute San Raffaele University, Milan, Italy.

Commercial relationships: none.

Corresponding author: Jason Samaha.

Email: jsamaha@ucsc.edu.

Address: Department of Psychology, University of California, Santa Cruz, CA 95064, USA.

References

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, 11(8):1, 1–16, <https://doi.org/10.1167/11.8.1>.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(241), 177–178, <https://doi.org/10.1038/226177a0>.
- 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., Schmittwilken, L., Sayim, B., Manassi, M., & Herzog, M. H. (2020). Capsule networks as recurrent models of grouping and segmentation. *PLoS Computational Biology*, 16(7), e1008017, <https://doi.org/10.1371/journal.pcbi.1008017>.
- Enns, J. T., & Di Lollo, V. (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](https://doi.org/10.1016/S1364-6613(00)01520-5).
- Gong, M., Liu, T., Chen, Y., & Sun, Y. (2024). Dissociable effects of endogenous and exogenous attention on crowding: Evidence from event-related potentials. *Brain Sciences*, 14(10), 956, <https://doi.org/10.3390/brainsci14100956>.
- Hermens, F., Herzog, M. H., & Francis, G. (2009). Combining simultaneous with temporal masking. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 977–988, <https://doi.org/10.1037/a0014252>.
- 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, 1–18, <https://doi.org/10.1167/15.6.5>.
- 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>.
- Kewan-Khalayly, B., & Yashar, A. (2022). The role of spatial attention in crowding and feature binding. *Journal of Vision*, 22(13):6, 1–16, <https://doi.org/10.1167/jov.22.13.6>.

- Kimchi, R., Yeshurun, Y., & Cohen-Savransky, A. (2007). Automatic, stimulus-driven attentional capture by objecthood. *Psychonomic Bulletin & Review*, *14*, 166–172, <https://doi.org/10.3758/BF03194045>.
- Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, *7*(2):4, 1–12, <https://doi.org/10.1167/7.2.4>.
- Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, *7*(2):1, 1–7, <https://doi.org/10.1167/7.2.1>.
- Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10):13, 1–14, <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, 1–10, <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>.
- Morea, M., Herzog, M. H., Francis, G., & Manassi, M. (2025). Dynamics of vision: Grouping takes longer than crowding. *Journal of Vision*, *25*(12):16, 1–17, <https://doi.org/10.1167/jov.25.12.16>.
- 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., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature neuroscience*, *11*(10), 1129–1135, <https://doi.org/10.1038/nn.2187>.
- Pöder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2):23, 1–10, <https://doi.org/10.1167/7.2.23>.
- Ringer, R. V., Coy, A. M., Larson, A. M., & Loschky, L. C. (2021). Investigating visual crowding of objects in complex real-world scenes. *i-Perception*, *12*(2), 2041669521994150, <https://doi.org/10.1177/2041669521994150>.
- Ronconi, L., Bertoni, S., & Bellacosa Marotti, R. (2016). The neural origins of visual crowding as revealed by event-related potentials and oscillatory dynamics. *Cortex*, *79*, 87–98, <https://doi.org/10.1016/j.cortex.2016.03.005>.
- Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision*, *10*(10):17, 1–7, <https://doi.org/10.1167/10.10.17>.
- Sayim, B., Manassi, M., & Herzog, M. (2014). How color, regularity, and good Gestalt determine backward masking. *Journal of Vision*, *14*(7):8, 1–11, <https://doi.org/10.1167/14.7.8>.
- Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, *21*(5), 641–644, <https://doi.org/10.1177/0956797610368811>.
- Tanriverdi, D., & Cornelissen, F. W. (2024). Rapid assessment of peripheral visual crowding. *Frontiers in Neuroscience*, *18*, 1332701, <https://doi.org/10.3389/fnins.2024.1332701>.
- Tiurina, N. A., Markov, Y. A., Choung, O.-H., Herzog, M. H., & Pascucci, D. (2022). Unlocking crowding by ensemble statistics. *Current Biology*, *32*(22), 4975–4981.e3, <https://doi.org/10.1016/j.cub.2022.10.003>.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., . . . Von Der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217, <https://doi.org/10.1037/a0029333>.
- Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306, [https://doi.org/10.1016/S0042-6989\(98\)00114-X](https://doi.org/10.1016/S0042-6989(98)00114-X).
- 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>.