Limited midlevel mediation of visual crowding: Surface completion fails to support uncrowding

Cathleen M. Moore Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

Qingzi Zheng Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

Visual *crowding* refers to impaired object recognition that is caused by nearby stimuli. It increases with eccentricity. Image-level explanations of crowding maintain that it is caused by information loss within early encoding processes that vary in functionality with eccentricity. Alternative explanations maintain that the interference is not limited to two-dimensional image-level interactions but that it is mediated within representations that reflect three-dimensional scene structure. *Uncrowding* refers to when adding stimulus information to a display, which increases the noise at an image level, nonetheless decreasing the amount of crowding that occurs. Uncrowding has been interpreted as evidence of midlevel mediation of crowding because the additional information tends to provide an opportunity for perceptually organizing stimuli into distinct and therefore protected representations. It is difficult, however, to rule out image-level explanations of crowding and uncrowding when stimulus differences exist between conditions. We adapted displays of a specific form of uncrowding to minimize stimulus differences across conditions, while retaining the potential for perceptual organization, specifically perceptual surface completion. Uncrowding under these conditions would provide strong support for midlevel mediation of crowding. In five experiments, however, we found no evidence of midlevel mediation of crowding, indicating that at least for this version of uncrowding, image-level explanations cannot be ruled out.

Introduction

The fidelity of peripheral vision is worse than that of central vision. This is in part because spatial acuity decreases with eccentricity (e.g., Anstis, 1974), but an even greater limitation is that interference from nearby stimuli, or visual clutter, increases steeply with eccentricity (see Rosenholtz, 2016; Strasburger, Rentschler, & Jüttner, 2011 for reviews). A manifestation of this interference is v*isual crowding*, which refers to impaired recognition of stimuli (*targets*) when there are other stimuli (*flankers*) nearby (see Levi, 2014; Pelli, Palomares, & Majaj, 2004; Whitney & Levi, 2011 for reviews). *Critical spacing* is the distance by which flankers must be separated from a target to avoid interference, and it increases at a rate of approximately 0.5 × eccentricity (Bouma, 1970). Because the natural world is rife with visual clutter, visual crowding contributes substantially to limitations of peripheral vision under natural viewing conditions (Rosenholtz, 2016; Vater, Wolfe, & Rosenholtz, 2022; Xia, Manassi, Nakayama, Zipser, & Whitney, 2020).

Pooling models of visual crowding maintain that it is caused by the integration of information that falls within fixed sampling regions of the visual field at early image-encoding stages of processing. According to these models, when a sampling region is larger than a to-be-identified target, any additional information (i.e., clutter) in the region is sampled along with the target, adding noise to the output representation of this early stage of processing. In pooling models, critical spacing is determined by the size of the sampling regions, which increase with eccentricity. High dimensional pooling models (e.g., Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Keshvari & Rosenholtz, 2016) reflect some of the known channel properties of early visual processes and have been successful at capturing many findings in the crowding literature (e.g., Keshvari & Rosenholtz, 2016; Rosenholtz, Yu, & Keshvari, 2019), though not all (Bornet et al., 2021; Doerig et al., 2019; Rosenholtz et al., 2019). Whether simple or high-dimensional, pooling models are examples of what we will refer to in this article as *image-level* explanations of crowding. The source of crowding is noise that is incurred at image-encoding stages of visual processing. The degree of noise is determined by the size of fixed sampling regions that are defined within the two-dimensional (2D) visual field and, critically, are blind with regard to the three-dimensional (3D) surface

Citation: Moore, C. M., & Zheng, Q. (2024). Limited midlevel mediation of visual crowding: Surface completion fails to support uncrowding. *Journal of Vision*, *24*(1):11, 1–16, https://doi.org/10.1167/jov.24.1.11.

https://doi.org/10.1167/jov.24.1.11 Received March 28, 2023; published January 31, 2024 ISSN 1534-7362

structure of the scene that produced the image. An implication of image-level explanations of crowding is that the perceived 3D surface structure of a scene should be irrelevant to the amount of crowding that occurs.

A known property of crowding has raised the possibility that, contrary to pure image-level explanations, crowding may be mediated by representations that reflect the structure of the scene from which the image derived. Specifically, crowding is reduced by feature differences between the targets and flankers such as, color and contrast polarity (Kooi, Toet, Tripathy, & Levi, 1994; Manassi, Sayim, & Herzog, 2012; Rosen & Pelli, 2015), orientation (Felisberti, Solomon, & Morgan, 2005; Huckauf, Heller, & Nazir, 1999; Sayim & Cavanagh, 2013), and shape (Nazir, 1992), among other features. These target-flanker similarity effects suggested the possibility that if flankers can be perceptually grouped and represented as distinct objects (perceptual units), separate from the target, then target processing can be relatively protected from flanker interference (e.g., Francis, Manassi, & Herzog, 2017; Herzog, Sayim, Chicherov, & Manassi, 2015; Manassi et al., 2012). We will use the term *midlevel* mediation to refer to this kind of explanation of crowding. Mid-level explanations, in general, hypothesize that flanker interference is mediated by representations that include scene-level properties such as the relative depths of distinct surfaces, the extension of surfaces behind other surfaces, and 3D object shape. None of that information is explicit in the image but must instead be abstracted from it through what we refer to in this paper as midlevel organization processes, or perceptual organization processes.[1] Although target-flanker similarity effects are consistent with midlevel mediation of crowding, they are also consistent with pooling models which are purely image-based explanations. This is because the more dissimilar two stimuli are, the less overlap there will be in feature-specific processing channels that encode them and therefore, the less loss of discriminating information there will be.

A phenomenon referred to as *uncrowding* presents a greater challenge to image-based models of crowding than do simple target-flanker similarity effects. Uncrowding refers to when *adding* stimulus energy (or noise) to a display—often in locations that are beyond regions defined by critical spacing—*reduces* crowding (e.g., Levi & Carney, 2011; Manassi et al., 2012; Manassi, Sayim, & Herzog, 2013; Manassi, Lonchampt, Clarke, & Herzog, 2016). Because uncrowding effects compare conditions in which the target and flanker relationships are locally identical or nearly identical, and yet different amounts of crowding occur, they are not as easily explained in image-based terms, such as channel overlap in early encoding processes.

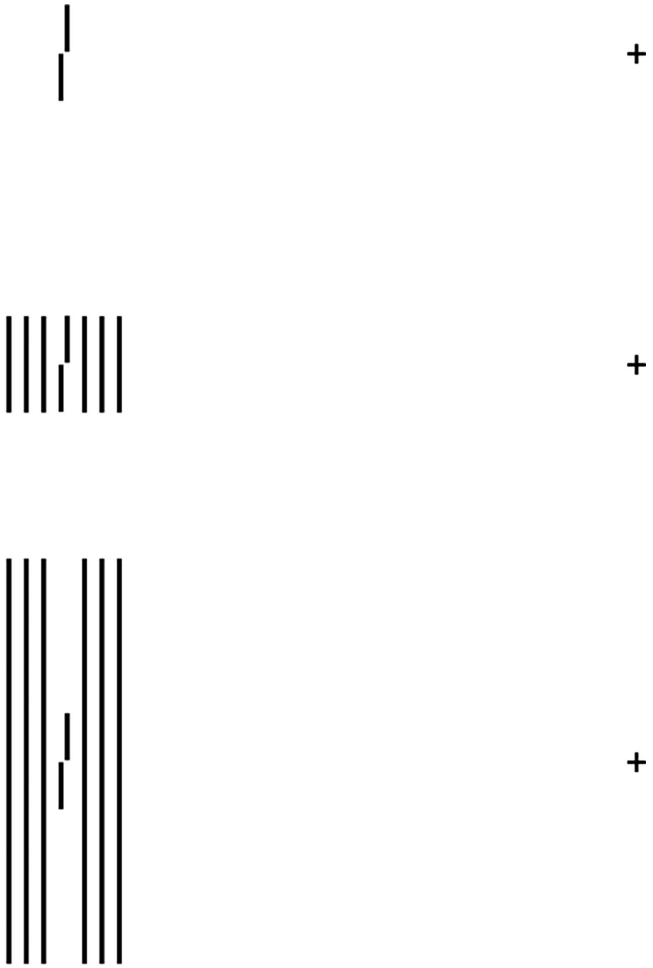

Figure 1. Illustration of crowding and uncrowding with stimuli similar to those used by Manassi et al. (2012) and Chicherov et al. (2014). The top row shows an uncrowded Vernier target, which, when viewed peripherally (fixate the plus sign), is relatively easy to identify as having a top right offset. The middle row shows the Vernier target flanked by equal-length vertical lines, making it difficult to discriminate peripherally (*crowding*). The bottom row shows that elongating the flankers reduces crowding, even though locally the target is still flanked by nearby stimuli and in fact more contrast energy has been added to the display (*uncrowding*).

An example of uncrowding that was reported by Manassi et al. 2012; (see also Chicherov, Plomp, & Herzog, 2014) is illustrated in Figure 1 and is the focus of the current study. The target is a Vernier stimulus for which the task is to report whether the top line segment is shifted to the left or to the right of the bottom one (top row). Adding vertical-line flankers adjacent to target causes substantial crowding (middle row). However, *increasing* the length of the flankers so that they extend beyond the vernier target *reduces*

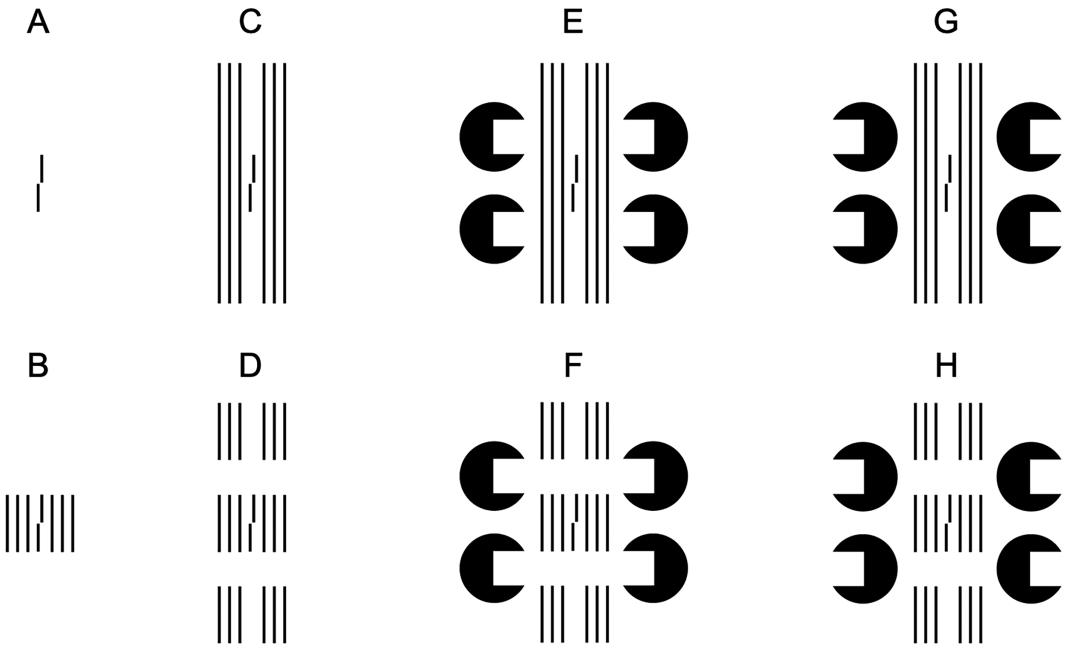

Figure 2. Illustration of stimuli used in this study. (**A**) Base Vernier target. (**B**) Equal-length flankers. (**C**) Elongated flankers. (**D**) Gapped flankers. (**E**) Elongated flankers with aligned inducers. (F) Gapped flankers with aligned inducers. (**G**) Elongated flankers with misaligned inducers. (H) Gapped flankers with misaligned inducers.

the amount of crowding that occurs (bottom row). Despite adding more contrast energy to the display, less interference occurs. One interpretation of this uncrowding effect is that it is a result of midlevel mediation of crowding. The assertion is that the elongated flankers are perceptually grouped as a distinct object or surface, thereby protecting the target from interference (Francis et al., 2017; Herzog et al., 2015). Although the effect is consistent with that interpretation, whenever there are stimulus differences across the conditions from which crowding is measured, it is difficult to rule out an effect of early image-level encoding processes. In this case, the line-terminators of the flankers are in close proximity to those of the target in the equal-length flankers condition, in which crowding occurs, but not in the elongated-flankers condition, in which uncrowding occurs. This difference in the spatial-proximity of image-features across conditions, and their different effects on early encoding processes, could be the source of the difference in crowding.

In the current study, we tested the hypothesis that the uncrowding effect illustrated in Figure 1 reflects mediation of crowding by midlevel representations. Our stimuli and logic are illustrated in Figure 2. In addition to the original no-flanker and equal-length flankers conditions (Figures 2A & 2B), we compared conditions in which flankers were explicitly elongated in the image (Figures 2C, 2E, 2G) to conditions in which the elongated flankers had gaps in them (Figures 2D, 2F, 2H) so that at the image-level, the Vernier target was locally flanked by equal-length flankers. In some conditions, pacman-like inducers were added to the display. In the aligned conditions (Figures 2E & 2F), the inducers supported the perception of illusory rectangular surfaces that were occluding solid disks and were aligned with the location of the gaps when present. This would support the perceptual completion of the gapped flankers into elongated flankers. In the misaligned conditions (Figures 2G & 2H), no completion of the flankers with gaps was supported. Together these conditions allowed us to compare uncrowding effects under conditions in which flankers were explicitly elongated in the image to conditions in which they could only have been elongated within midlevel perceptually organized representations (i.e., as output of surface completion processes that abstracted the spatially extended surfaces from image-level information). If significant uncrowding occurs in conditions in which the flankers require surface completion to be represented as elongated, it would provide strong evidence of midlevel mediation of visual crowding. Across five experiments, however, we did not find this evidence. There was robust crowding in conditions with equal-length flankers as well as substantial reduction in crowding when the flankers were explicitly elongated in the image (i.e., uncrowding). However, we found no evidence of reduced crowding when elongation of the flankers required surface completion.

General method

Protocols for all experiments were approved by the University of Iowa Institutional Review Board.

Participants

Participants were students at the University of Iowa who received credit for fulfilling a requirement in an introductory Psychology course. All participants reported normal or corrected-to-normal visual acuity and were naïve with respect to the purpose of the experiment. We report data from 24 participants in each of the experiments. No individual participated in more than one experiment. The choice of sample size was based on a power analysis of the basic uncrowding effect reported in the elongated flanker condition of Chicherov et al. (2014). That analysis indicated a sample size of 3 to achieve 0.8 power. A similar analysis of an effect in Manassi et al. (2012, Experiment 1) indicated a sample size of 12 to achieve 0.8 power. To be conservative, we tested 24 participants because our measure and design were not identical to either of the previous studies, and because our main predictions concerned the interaction of uncrowding with display conditions. Participants with overall error rates that were not reliably different from chance (50%) were replaced. This resulted in the replacement of 9 participants in Experiment 1, 10 in Experiment 2, and 2 in Experiment 3. In addition, three participants in Experiment 2 and 2 participants in Experiment 3 were replaced because of incomplete sessions. This relatively high rate of chance performance and incomplete sessions is likely due to the fact that those experiments were conducted during COVID shutdown and were therefore run online. Finally, a fifth experiment was conducted later when it was possible to test participants in the lab using the eye tracker. We intended to test 24 participants, but because of overscheduling, we tested 29. We report only the data from the first 24 participants to match the previous experiments. We did conduct all of the same analyses on the full set of 29 participants, and no patterns of effects were different.

Apparatus

Experiments 1 through 4 were programmed in OpenSesame (v 3.3.8) and were run online using OSWeb (Mathôt, Schreij, & Theeuwes, 2012) and a JATOS server (Lange, Kühn, & Filevich, 2015). Participants completed the task on their own computers, most of which were 13-inch laptops. Experiment 5 was programmed in E-Prime 2 and was conducted using an EyeLink 1000 eye tracker to monitor fixation. Stimuli were presented on a 24-inch VIEWPixx/3D LCD monitor with resolution 1920 × 1080 and a 100 Hz refresh rate. Responses were entered using a standard keyboard.

Stimuli

Stimuli are illustrated in Figure 2. Sizes and distances for Experiments 1 through 4 are reported as approximations, based on an assumed 13-inch laptop screen at an estimated viewing distance of 55 cm. Sizes and distances for Experiment 5 are provided in parentheses. Targets (Figure 2A) were ∼1° (1.8°) Vernier stimuli drawn with a 2-pixel pen width. They consisted of one ∼0.5° (0.9°) vertical line segment positioned above a second ∼0.5° (0.9°) vertical line segment, offset by ∼0.2° (0.2°) to the left or to the right. Targets were presented ∼8° (Experiments 1 and 2), ∼7° (Experiments 3 and 4) or 10.5° (Experiment 5) to the left or right of fixation, which was a 0.33° x 0.33° black cross (0.4 radius black circle) at the center of the screen. When present, flankers were *equal-length*, *elongated*, or *gapped*. Equal-length flankers consisted of 3 ∼1° (1.8°) vertical line segments (i.e., equal in length to the target), spaced ∼0.2° (0.4°) apart, on either side of the target (Figure 2B). Elongated flankers were the same as equal-length flankers except that they were ∼4.5° (8.5°) tall (Figure 2C). Finally, gapped flankers were the same as elongated flankers except that they had two ∼1° (1.25°) gaps in them starting at the top and bottom of the target so that locally they were identical to the equal-length-flankers condition (Figure 2D). In some conditions four solid-filled square-mouthed pacman-like inducers (diameter ∼1°) (2.25°) were added to the displays. The "mouths" of the inducers were the same size as the gaps in the gapped flankers, and they were horizontally aligned with the gaps. In the *aligned* conditions (Figures 2E & 2F), the "mouths" of the inducers faced each other. In the *misaligned* conditions (Figures 2G & 2H), they faced in opposite directions. All stimuli were presented on a gray (Hex Code no. BFBFBF) background (106 cd/m^2). Targets and flankers were black (Hex Code no. 000000) (0.5 cd/m^2). In Experiments 1 and 3, the inducers were also black. In Experiments 2 and 4 the inducers were red (Hex Color no. B51700) and were desaturated by setting opacity to 40%. In Experiment 5 they were gray (84 cd/m^2).

Task

The task in all experiments was to report whether the target's top-line segment was shifted to the left or to the

right of the bottom-line segment by pressing the "F" or "J" key with the left or right index fingers, respectively. Participants were asked to respond as accurately as possible without worrying about the speed of their responses.

Experiments 1 through 4 procedure

Experiments 1 through 4 were conducted online. Each participant completed a single session that lasted approximately one hour. After an informed consent process, participants were provided with a set of written instructions with step-by-step illustrations of the task. Participants then completed two blocks of 32 practice trials each. For the first practice block, the stimulus display was up until a response was made, which allowed participants to become familiar with the task. For the second practice block, the stimulus display was presented for 120 milliseconds (ms), which was the same duration as that used during the main part of the experiment. Trial-by-trial feedback was provided during practice: a smiling or frowning cartoon face was shown for 500 ms to indicate correct or incorrect responses, respectively. After practice, participants completed nine blocks of 64 trials each, with the first full block of trials considered practice and not included in final analyses. Between blocks, feedback was provided in the form of mean response time and accuracy for the preceding block. Participants could rest as much as they liked between blocks and continue to the next block by pressing the C key. At the end of the final block, a written message indicated that the experiment was complete, and a brief explanation of the experiment was provided. Trials began with the presentation of a black fixation cross (+) at the center of the screen for 750 ms. The stimulus display was then presented randomly to the left or the right of fixation for 120 ms. After a blank 1000 ms intertrial interval following a response, the fixation cross for the next trial was presented. If no key press was detected after 10 seconds, the next trial would begin without a response.

Experiment 5 procedure

Experiment 5 was conducted in-person in the lab but was otherwise similar to Experiments 1 through 4. After an informed consent process, participants were tested in an individual room. After reading instructions describing the task, they were situated comfortably in a chin and head rest, and a nine-point calibration procedure was used to calibrate the eye tracker. Participants then completed 2 brief demonstration blocks of eight trials each (one block with stimuli presented to the left and another with stimuli presented to the right). It was then explained to them that they now had to keep their eyes fixated on the central fixation circle during a trial, and that if their eyes moved away from it, the stimuli would disappear. They completed two demonstration blocks (one left and one right) of eight trials each getting used to the gaze-contingent conditions. Finally they completed two longer (27 trials each) practice blocks (one left and one right) during which they received feedback when they made an incorrect response. Following this set of demonstration and practice blocks, participants completed 8 blocks of 48 trials each from which data were recorded. Feedback was not provided in these blocks. Participants rested, leaning back from the head and chin rest, between blocks, and the eye-tracker was recalibrated as needed after breaks. Trials began with the presentation of a central fixation circle. A research assistant, who was in the room with the participant throughout the experiment viewing output from the eye-tracker on a separate monitor, initiated each trial after confirming that the participant was fixating the central marker. At 500 ms after the initiation of a trial, the stimuli were presented. They appeared consistently to the left or to the right within a block, and they remained present until the participant made a response with the "F" key or the "J" key to indicate a left or right response, respectively. If fixation shifted away from center marker by 1.5° or more, the stimuli were removed from the display. The stimuli were presented again as soon as the participant returned their gaze to the central marker. Trials would end if no response was made within 10 seconds, but otherwise, observers could view the stimuli as long as they wanted.

Design, analyses, and predictions

In Experiments 1 through 4, the eight conditions (Figure 2) were all presented equally often in a pseudorandom order eight times each (four to the left and four to the right of fixation) per block. Data were collected from eight blocks, resulting in 64 observations per condition for each subject. In Experiment 5, stimuli were presented on the same side (left or right) in a given block six times each in a pseudorandom order. Data were collected from eight blocks (four left and four right), resulting in 48 observations per condition for each subject.

The main dependent measure was proportion correct (PC). Alpha was set at .05 throughout, and two-tailed tests were used for specific comparisons. Effect sizes are reported as *adjusted partial eta-squared* ($adj\ \hat{\eta}_p^2$), an estimate of partial eta-squared that adjusts for positive bias (Mordkoff, 2019).

Data were analyzed in three steps. First, base crowding and uncrowding effects were assessed by comparing error rates in no flankers, equal-length

flankers, gapped flankers, and elongated flankers conditions in a single one-way analysis of variance (ANOVA) followed by specific planned comparisons. Equal-length flankers versus no flankers provided a test of the base crowding effect. Equal-length flankers versus elongated flankers provided a test of uncrowding. Gapped flankers versus equal-length flankers assessed whether the extra line segments in the gap flankers altered the base crowding effect. We then conducted two separate 2 × 2 within-subjects ANOVAs, the designs of which are illustrated in in Figure 3.

The factors of the first 2 × 2 analysis (Figure 3A) were Level of Representation (image-level, midlevel) and Gap (gap, no gap). Level of representation refers to whether the elongation of flankers was explicit in the image (image level) or required perceptual completion to be represented as such (midlevel). For the image-level conditions, the flankers were either explicitly elongated (no gap) or they had a gap (gap) leaving the target flanked by equal-length flankers locally. For the midlevel conditions there were inducers in the displays, and the gap and no-gap conditions were defined on the basis of whether the induces were misaligned (gap) or aligned (no-gap), respectively. The prediction for this analysis is that if uncrowding can be mediated by midlevel representations, then accuracy should be higher in both of the no-gap conditions compared to the gap conditions, regardless of whether the elongation is explicit in the image (image level) or exists only at midlevels of representation following perceptual completion (midlevel). If, however, uncrowding depends on the flankers being elongated explicitly in the image, then accuracy should be higher in the image-level no-gap condition compared to the gap condition, but not in the midlevel no-gap condition compared to the gap condition. Thus overall, midlevel mediation of uncrowding predicts no significant interaction between Level-of-Representation and Gap in this analysis, whereas uncrowding based exclusively on image-level information predicts a significant interaction.

The factors of the second 2 × 2 analysis (Figure 3B) were Inducer Alignment (misaligned, aligned) and Gap (gap, no gap). For the misaligned conditions, the flankers were either explicitly elongated (no gap) or they had a gap (gap), leaving the target flanked locally by equal-length flankers. For the aligned conditions, even though there is a gap at the image level in the gap condition, following perceptual completion (i.e., within midlevel representations), the flankers should be represented as elongated even in the gap condition. The prediction for this analysis is that if uncrowding can be mediated by midlevel representations, then accuracy should be higher in the no-gap condition than in the gap condition when inducers are misaligned but not when they are aligned because at midlevels of representation, both conditions (gap and no-gap) will have elongated flankers to support uncrowding.

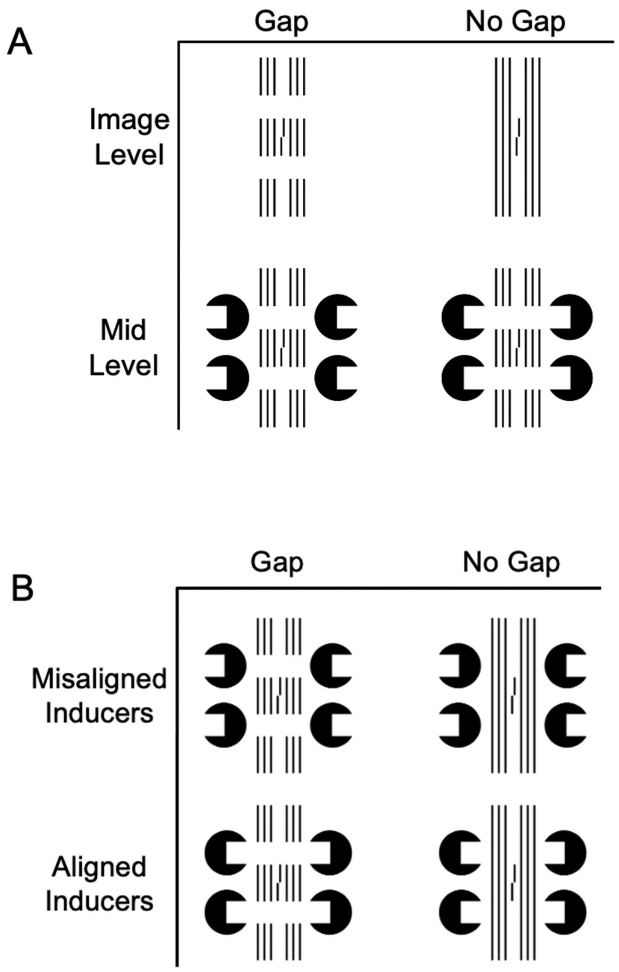

Figure 3. Designs of two 2 × 2 analyses conducted for each experiment. They both tested whether uncrowding due to elongation of the flankers requires explicit elongation in the image or is also supported by midlevel representation of elongation. (**A**) Level of representation (image-level, midlevel) × Gap (gap, no gap): Mid-level mediation of crowding predicts no modulation of the effect of gap by level of representation (i.e., no interaction), whereas exclusively image-level crowding predicts an interaction. (**B**) Inducer Alignment (misaligned, aligned) × Gap (gap, no gap): Mid-level mediation of crowding predicts a modulation of the effect of gap by whether or not the inducers are aligned (i.e., an interaction), whereas exclusively image-level crowding predicts no interaction.

If, however, uncrowding depends on the flankers being elongated explicitly in the image, then accuracy should be low in both the misaligned and aligned gap conditions and higher in both the misaligned and aligned no-gap conditions. Overall, therefore, this analysis makes the opposite predictions of the previous one. Specifically, midlevel mediation of uncrowding predicts a significant interaction between Inducer Alignment and Gap, whereas uncrowding based only on image-level information predicts no interaction.

Experiments 1 through 4

Experiments 1–4 differed in the nature of the inducers as illustrated in Figure 4. In Experiment 1, the inducers were black and the horizontal distance from the target to the center of an inducer was ~2.5°. The results of Experiment 1 revealed evidence of crowding from the inducers themselves. To reduce the potential crowding from inducers in Experiment 2, we made them a different color from the target and flankers. In Experiment 3, we increased the horizontal distance from the center of an inducer to the target to ~4.3°. And finally in Experiment 4, the inducers were both a different color and at the further distance from the target and flankers. These changes reduced crowding from the inducers but did not eliminate it. Despite the extra bit of crowding from the inducers, the pattern of results across critical conditions was clear and consistent across all four experiments.

Results and discussion

Robust crowding and uncrowding occurred in all four experiments. There was, however, no evidence that crowding was mediated by midlevel representations. Instead, the pattern of results was consistent with the hypothesis that the uncrowding was due exclusively to image-level differences between conditions. Specifically, in all four experiments, there was a significant interaction between the Level of Representation and Gap condition, indicating that explicit elongation of the flankers in the image was necessary to support uncrowding. Moreover, there was no interaction between Inducer Alignment and Gap condition, indicating that providing support for the perceptual completion of elongated flankers behind perceived surfaces was insufficient to support uncrowding. Statistical analyses summarized in the following paragraphs confirmed these patterns of results in all four experiments.

Basic crowding and uncrowding effects

Figure 5 shows mean PCs for the base crowding and uncrowding conditions in each of the four experiments. Subject means were submitted to separate one-way within-subjects ANOVAs, and revealed a significant effect of condition in all four experiments, Exp 1: $F(3,69) = 97.20$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.800$; Exp 2: $F(3,69) = 64.65$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.726$; Exp 3: $F(3,69) = 110.35$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.820$; Exp 4: $F(3,69) = 68.13$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.737$. Specific planned comparisons confirmed that for all four experiments there was a base crowding effect such that accuracy decreased with equal-length flankers compared to no flankers, Exp 1: mean diff = 0.348 ± 0.028, $t(23) = 12.67$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.869$; Exp 2: mean diff = 0.322 ± 0.028, $t(23) = 11.47$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.845$; Exp 3: mean diff = 0.365 ± 0.026, $t(23) = 14.16$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.893$; Exp 4: mean diff = 0.297 ± 0.022, $t(23) = 13.78$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.887$. In addition, for all four experiments, uncrowding

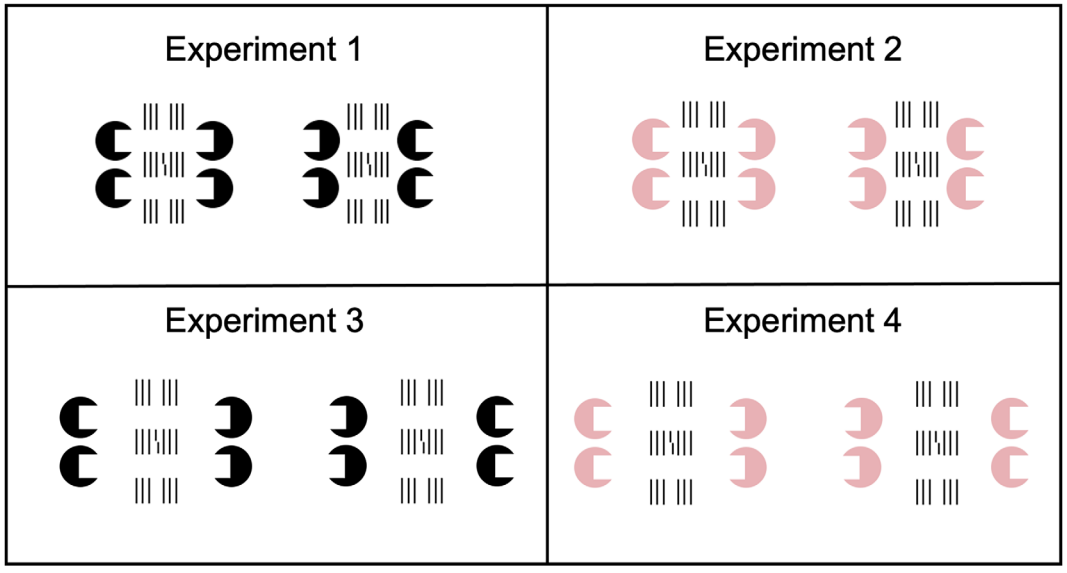

Figure 4. Schematic illustrations of the stimulus differences across experiments. The increased spacing and change in color of the inducers (see text for details) were designed to reduce crowding from the inducers themselves.

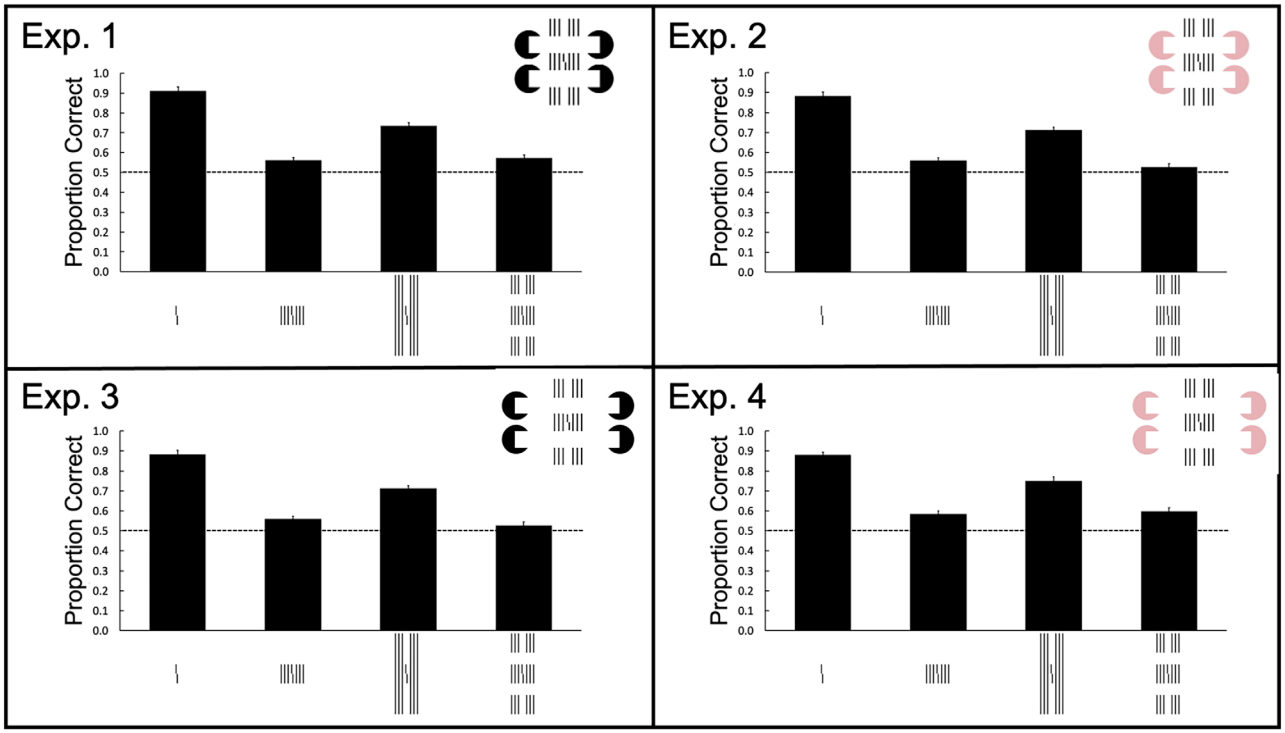

Figure 5. Proportion correct data from all four experiments for the base-crowding and uncrowding conditions. Error bars indicate within-subjects standard errors (Cousineau, 2005; Morey, 2008).

occurred. Responses were more accurate with explicitly elongated flankers than with equal-length flankers, Exp 1: mean diff = 0.17 ± 0.02, $t(23) = 7.18$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.678$; Exp 2: mean diff = 0.151 ± 0.037, $t(23) = 4.09$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.396$; Exp 3: .214 ± 0.028, $t(23) = 7.73$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.710$; Exp 4: mean diff = 0.167 ± 0.028, $t(23) = 5.92$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.587$. And finally, there was no reliable difference in the amount of crowding caused by elongated flankers with gaps in them than that caused by equal-length flankers: Exp 1: mean diff = 0.011 ± 0.017, $t(23) = 0.26$, $p = 0.524$, $adj\ \hat{\eta}_p^2 = -.040$; Exp 2: mean diff = 0.033 ± 0.018, $t(23) = 1.80$, $p = 0.09$, $adj\ \hat{\eta}_p^2 = 0.085$; Exp 3: mean diff = 0.014 ± 0.018, $t(23) = 0.78$, $p = 0.442$, $adj\ \hat{\eta}_p^2 = -.017$[2]; Exp 4: mean diff = 0.014 ± 0.019, $t(23) = 0.73$, $p = 0.473$, $adj\ \hat{\eta}_p^2 = -.020$.

Level of representation × gap analyses

Figure 6 shows the mean PCs for the conditions used in the Level of Representation (image level, midlevel) × Gap (no gap, gap) analysis for each of the four experiments. In all four experiments, there was a main effect of Level of Representation, Exp 1: $F(1,23) = 39.76$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.618$; Exp 2: $F(1,23) = 25.67$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.507$; Exp 3: $F(1,23) = 37.93$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.606$; Exp 4: $F(1,23) = 64.12$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.725$, as well as a main effect of Gap, Exp 1: $F(1,23) = 34.81$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.585$; Exp 2: $F(1,23) = 34.43$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.582$; Exp 3: $F(1,23) = 41.72$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.629$; Exp 4: $F(1,23) = 21.41$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.460$. Critically, the interaction between Level of Representation and Gap was also significant in all four experiments, Exp 1: $F(1,23) = 18.99$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.428$; Exp 2: $F(1,23) = 18.14$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.417$; Exp 3: $F(1,23) = 64.48$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.726$; Exp 4: $F(1,23) = 17.05$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.401$. Follow-up simple main-effect analyses confirmed that for all four experiments, the difference between gap and no gap conditions was significant for the image-level condition in which the elongation was explicit in the image, Exp 1: mean diff = 0.160 ± 0.025, $t(23) = 6.29$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.616$; Exp 2: mean diff = 0.184 ± 0.033, $t(23) = 5.60$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.558$; Exp 3: mean diff = 0.228 ± 0.029, $t(23) = 7.91$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.720$; Exp 4: mean diff = 0.153 ± 0.031, $t(23) = 4.87$, $p < 0.001$, , $adj\ \hat{\eta}_p^2 = 0.486$, but was not significant for the midlevel condition in which flanker elongation required perceptual completion, Exp 1: mean diff = 0.026 ±

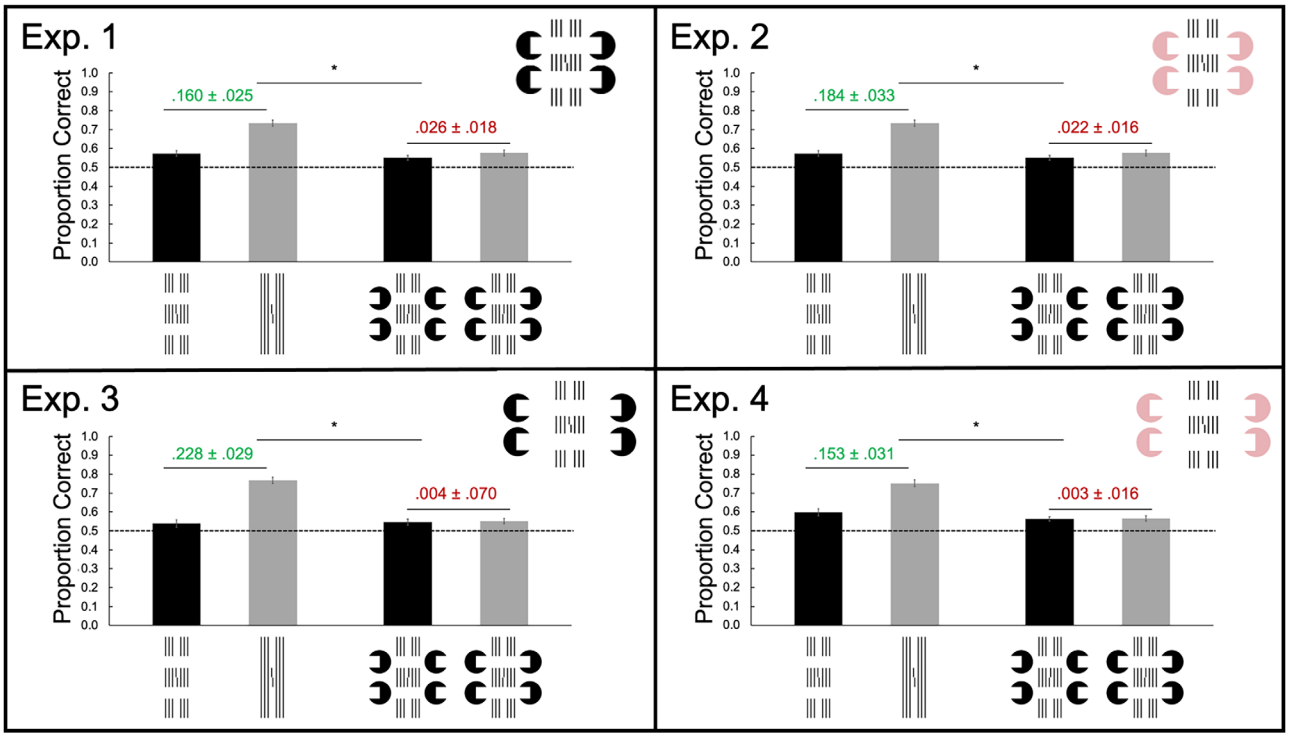

Figure 6. Proportion correct data from all four experiments for the Level of Representation (image level, midlevel) × Gap (gap, no-gap) analyses. Error bars indicate within-subjects standard errors (Cousineau, 2005; Morey, 2008). The green and red numbers indicate mean differences that were significant and non-significant, respectively. The topline and asterisk in each figure indicates the significance of interaction between Level of Representation and Gap.

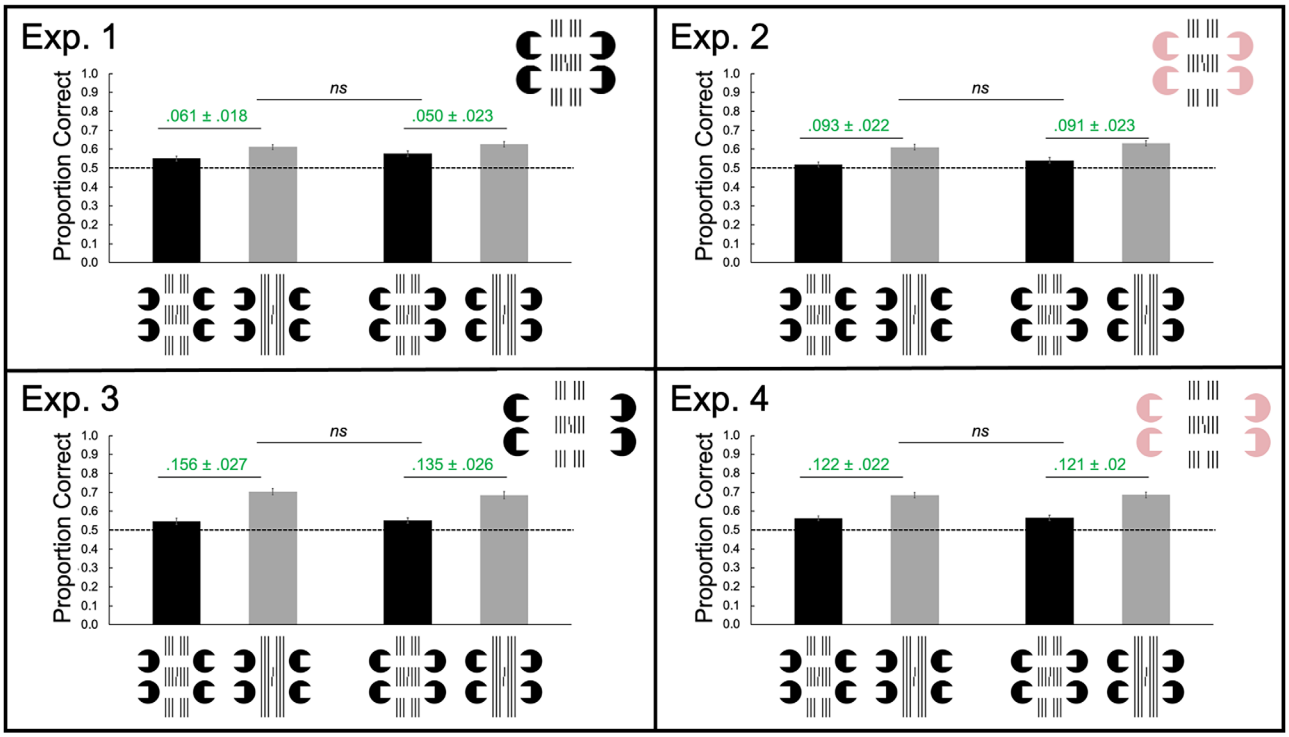

Figure 7. Proportion correct data from all four experiments for the Inducer Alignment (unaligned, aligned) × Gap (gap, no-gap) analyses. Error bars indicate within-subjects standard errors (Cousineau, 2005; Morey, 2008). The green numbers indicate significant mean differences. The topline in each figure indicates that interaction between Inducer Alignment and Gap condition was non-significant (*ns*).

0.018, $t(23) = 1.43$, p = 0.166, $adj\ \hat{\eta}_p^2 = 0.042$; Exp 2: mean diff = 0.022 ± 0.016, t (23) = 1.343, p = 0.192, $adj\ \hat{\eta}_p^2 = 0.032$; Exp 3: mean diff = 0.004 ± 0.070, $t(23) = 0.29$, p = 0.773, $adj\ \hat{\eta}_p^2 = -.040$; Exp 4: mean diff = 0.003 ± 0.016, t (23) = 0.16, p = 0.875, $adj\ \hat{\eta}_p^2 = -.042$.

Inducer alignment × gap analyses

Figure 7 shows the results from the Inducer Alignment X Gap analysis for each of the four experiments. For all four experiments, there was a main effect of Gap, Exp 1: $F(1,23) = 11.88$, $p = 0.01$, $adj\ \hat{\eta}_p^2 = 0.312$; Exp 2: $F(1,23) = 22.79$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.476$; Exp 3: $F(1,23) = 34.78$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.585$; Exp 4: $F(1,23) = 34.56$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.583$, but not for Inducer Alignment, Exp 1: $F(1,23) = 2.65$, $p = 0.117$, $adj\ \hat{\eta}_p^2 = 0.064$; Exp 2: $F(1,23) = 3.02$, $p = 0.096$, $adj\ \hat{\eta}_p^2 = 0.078$; Exp 3: $F(1,23) = 0.43$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = -.024$; Exp 4: $F(1,23) = 0.035$, $p = 0.944$, $adj\ \hat{\eta}_p^2 = -.042$. Critically, there was no reliable interaction between Inducer Alignment and Gap in any of the experiments, Exp 1: $F(1,23) = 0.196$, $p = 662$, $adj\ \hat{\eta}_p^2 = -.035$; Exp 2: $F(1,23) = 0.003$, $p = 0.957$, $adj\ \hat{\eta}_p^2 = -.043$; Exp 3: $F(1,23) = 1.12$, $p = 0.173$, $adj\ \hat{\eta}_p^2 = -.036$; Exp 4: $F(1,23) = 0.01$, $p = 0.944$, $adj\ \hat{\eta}_p^2 = -.043$.

Experiment 5

Experiments 1 through 4 were conducted online due to COVID restrictions. Experiment 5 was conducted in the lab to confirm that various the parameter choices that we were forced to make to accommodate online conditions did not determine the pattern of results that we observed. Using an eye-tracker to monitor fixation and gaze-contingent displays, stimulus duration was effectively unlimited, and stimuli appeared on the same side of fixation on every trial within a given block of trials, thereby eliminating uncertainty with regard to where the stimuli would appear.

Results and discussion

As can be seen in Figure 8, the same pattern of results was observed in Experiment 5 as in the online experiments. Robust crowding and uncrowding occurred (Figure 8A). There was, however, no evidence that crowding was mediated by midlevel representations. Specifically, there was an interaction between the Level of Representation and Gap condition, indicating that explicit elongation of the flankers in the image was necessary to support uncrowding (Figure 8B). Moreover, there was no interaction between Inducer Alignment and Gap condition, indicating that providing support for the perceptual completion of elongated flankers behind perceived surfaces was insufficient to support uncrowding (Figure 8C). Statistical analyses confirmed these patterns.

Basic crowding and uncrowding effects

Figure 8A shows the proportion correct for the base crowding condition. A one-way within-subjects ANOVA on the subject means revealed a significant effect of condition, $F(3,69) = 113.17$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.824$. Specific comparisons confirmed that there was a base crowding effect such that more errors were made with equal-length flankers than with no flankers, mean diff = 0.406 ± 0.028, $t(23) = 14.71$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.900$, and that uncrowding occurred in that fewer errors were made with explicitly elongated flankers than with equal-length flankers, mean diff = 0.323 ± 0.035, $t(23) = 9.26$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.779$. And finally, there was no reliable difference in the amount of crowding caused by elongated flankers with gaps in them than that caused by equal-length flankers, mean diff = 0.018 ± 0.021, $t(23) = 0.20$, $p = 0.394$, $adj\ \hat{\eta}_p^2 = -.042$.

Level of representation × gap analyses

Figure 8B shows the mean PCs for the conditions used in the Level of Representation (image level, midlevel) × Gap (no gap, gap) analysis. There was a main effect of Level of Representation, $F(1,23) = 47.85$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.661$, as well as a main effect of Gap, $F(1,23) = 77.65$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.762$. Critically, the interaction between Level of Representation and Gap was also significant, $F(1,23) = 56.56$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.698$. Follow-up analyses of the simple main effects confirmed that the difference between gap and no gap conditions was significant for the image-level condition in which the elongation was explicit in the image, mean diff = 0.305 ± 0.033, $t(23) = 9.31$, $p < 0.001$, $adj\ \hat{\eta}_p^2 = 0.781$. Unlike all of the previous experiments, this difference was small but significant in the midlevel condition in which flanker elongation required perceptual completion as well mean diff = 0.037±.018, $t(23) = 2.10$, p = 0.046, $adj\ \hat{\eta}_p^2 = 0.124$. This uncrowding effect is in the direction predicted by midlevel mediation of crowding. However, it is only about 12% the size of the effect of uncrowding that occurred in the explicitly elongated flankers condition (.037 versus .305) and may reflect a difference in the amount of crowding caused by the different distances from target to nearest edge due to

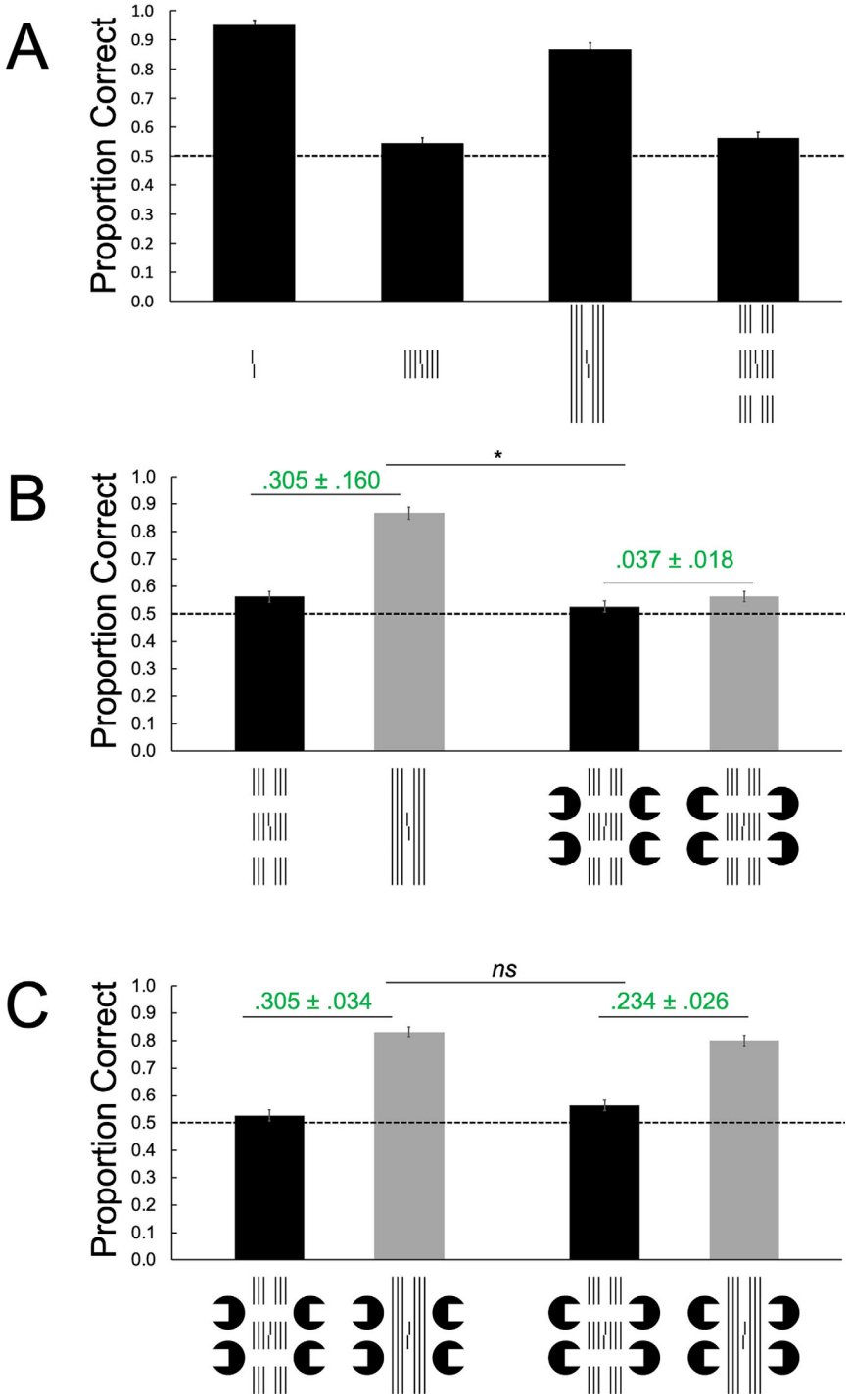

Figure 8. Proportion correct data from Experiment 5 (**A**). Base crowding and uncrowding effects (**B**). Level of Representation (image level, midlevel) × Gap (gap, no-gap) analysis (**C**). the Inducer Alignment (unaligned, aligned) × Gap (gap, no-gap) analysis. Error bars indicate within-subjects standard errors (Cousineau, 2005; Morey, 2008). The green numbers indicate significant mean differences. The topline in each figure indicates the significance (*) or non-significance (*ns*) of the interaction.

the different orientations of the inducers. The small magnitude of this effect, the large difference between it and the effect in the condition with explicit elongated flankers, and the fact that this difference was reliable in only one of five experiments all leads us to conclude that it does not constitute strong evidence of midlevel mediation of crowding.

Inducer alignment × gap analysis

Figure 8C shows the results from the Inducer Alignment X Gap analysis. There was a main effect of Gap, $F(1,23) = 11.88$, $p = 0.01$, $adj\ \hat{\eta}_p^2 = 0.312$, but not for Inducer Alignment, $F(1,23) = 2.65$, $p = 0.117$, $adj\ \hat{\eta}_p^2 = 0.064$. Critically, there was no reliable interaction between Inducer Alignment: $F(1,23) = 0.196$, $p = 662$, $adj\ \hat{\eta}_p^2 = -.035$.

General discussion

Target-flanker similarity effects on visual crowding (e.g., Felisberti et al., 2005; Kooi et al., 1994; Nazir, 1992) suggest the possibility that crowding is mediated by midlevel representations such that, for example, flankers can be perceptually grouped as a unit distinct from the target, thereby protecting it from flanker interference. Another possibility, however, is that differences in target-flanker similarity create differences in how much information is lost during early encoding processes because, for example, less similar stimuli will result in less overlap in early feature-specific sensory channels (e.g., Balas et al., 2009; Keshvari & Rosenholtz, 2016; Rosenholtz et al., 2019). Uncrowding effects, which occur when *adding* information to the display *decreases* crowding, seem to provide stronger evidence of midlevel mediation of crowding (Herzog et al., 2015). It is difficult, however, to rule out image-level explanations when stimulus differences remain between conditions. In this study, we modified displays in one version of uncrowding so that they differed in the perceptual organization that they supported while minimizing stimulus differences. In five experiments, we found no strong evidence of uncrowding that could be attributed to perceptually organized components of the displays. This does not preclude the possibility that a different approach could yield evidence of midlevel mediation of crowding, but insofar as the tests that we conducted on this version of uncrowding, the evidence did not support the conclusion that midlevel representations alone were a significant source of it.

The uncrowding effect that is the focus of the current study, as well as other uncrowding effects, may reflect advantages that are due to differences in image segmentation, rather than what we have been referring to as midlevel processes. Image segmentation, by our classification, is distinct from midlevel processes. Image segmentation yields output representations that are still in 2D image-based terms (e.g., mosaics of contrast regions). In contrast, midlevel processes yield output representations that include aspects of 3D scene structure that is not explicit in the image but that must instead be inferred from it (e.g., relative depths of surfaces and the extension of one surface behind another into invisible regions of space). Image segmentation processes can yield a parsing of the image that is based on feature information and that is, like the image itself, 2D and that does not therefore correspond to individuated components of the scene. That parsing will tend to correlate with objects and surfaces in the scene because objects and surfaces in the scene are what determine the nature of the image. The critical distinction, in our view, is that the output of image-segmentation is mute with regard to scene structure, whereas the output of midlevel processes *is* scene structure.

Two studies by Kimchi and Pirkner (Kimchi & Pirkner, 2015; Pirkner & Kimchi, 2017) showing robust configural effects on crowding provide an opportunity to highlight the distinction that we are drawing between image segmentation and midlevel mediation. Using stimuli made up of local elements to form global shapes, they showed that flankers that matched the target globally but not locally caused more crowding than flankers that matched the target locally but not globally (see also Livne & Sagi, 2007; Livne & Sagi, 2010). These are clearly effects of stimulus configuration on crowding, but they can be accounted for based on differences in image processing, such as image segmentation, rather than on midlevel mediation of the interference. To make this possibility clear, Rosenholtz et al. (2019) submitted examples of the stimuli used in Kimchi and Pirkner (2015) to their Texture Tiling Model, which is a purely image-based pooling model, and the target information was lost in the output for the configuration-matched stimuli but was largely retained for the configuration-mismatched stimuli. Image-encoding processes alone, therefore, can account for the difference in performance across the different configuration conditions.

We applied this same distinction between midlevel processes and image segmentation to what appeared to be evidence of midlevel mediation of the flanker-congruence effect (Moore, He, Zheng, & Mordkoff, 2021). The flanker congruence effect (FCE) refers to the observation that responses to the identity of a target stimulus that is presented at a known location (often fixation) is faster and/or more accurate when nearby flanking stimuli are associated with the same response as the target rather than a different response (e.g., Eriksen & Eriksen, 1974; Eriksen & Hoffman, 1972). The FCE is structurally similar to crowding in

that it involves the influence of nearby task-irrelevant stimuli on target identification. It is different, however, in that it is focused on the effect of congruence of stimulus identities, rather than general interference; both the targets and flankers are easily identifiable. It is not eccentricity dependent, and is thought to reflect limitations of selective attention. Similar to crowding, however, the flanker congruence effect is larger when the target and flankers are featurally similar (e.g., all the same color) than when they are different (e.g., Harms & Bundesen, 1983). This finding suggested the possibility that the FCE is mediated by midlevel representations (Baylis & Driver, 1992; Driver & Baylis, 1989). Specifically, it was hypothesized that when the target and flankers are featurally similar, they are perceptually grouped into a single perceptual unit (i.e., an object representation) that is then selected as a whole, leading to the identification and influence of the flanker information along with the target information. When they are dissimilar, they are instead perceptually organized into distinct perceptual units, allowing the target to be selected alone and protected from flanker interference. Notice how similar this explanation is to that of midlevel mediation of crowding. Like similarity effects in crowding, however, when the target and flankers are different along some image feature, they support better image segmentation, and that alone, without appeal to further organization, could result in a better representation (i.e., one with less information loss) of the target. In Moore et al. (2021), we tested this hypothesis by comparing conditions in which the targets and flankers should have grouped or not grouped, respectively, based on color similarity while holding constant differences at the image level. Within that design, the congruence of same-color flankers with the target had no greater effect than those of different-color flankers. As we discussed in that study, those findings do not challenge the claim that selective attention can be object-based (or in the language of the current paper, mediated by midlevel processes); there is substantial evidence that it can be (see Chen, 2012 for a review). Rather, the point is that some effects that appear to reflect midlevel mediation may actually reflect differences in image-level processing.

We cannot conclude from this one study that uncrowding never reflects true midlevel mediation. There are many versions of uncrowding (see Herzog et al., 2015 for a review), and each version would require the development of a targeted strategy analogous to the one used in the current study but specific to the particular version of uncrowding, which is beyond what we can offer here. Another strategy for assessing whether an uncrowding effect can be accounted for in terms of image-level processing has been to submit images of the stimuli to models that include only image-level processes and ask how well performance differences across conditions can

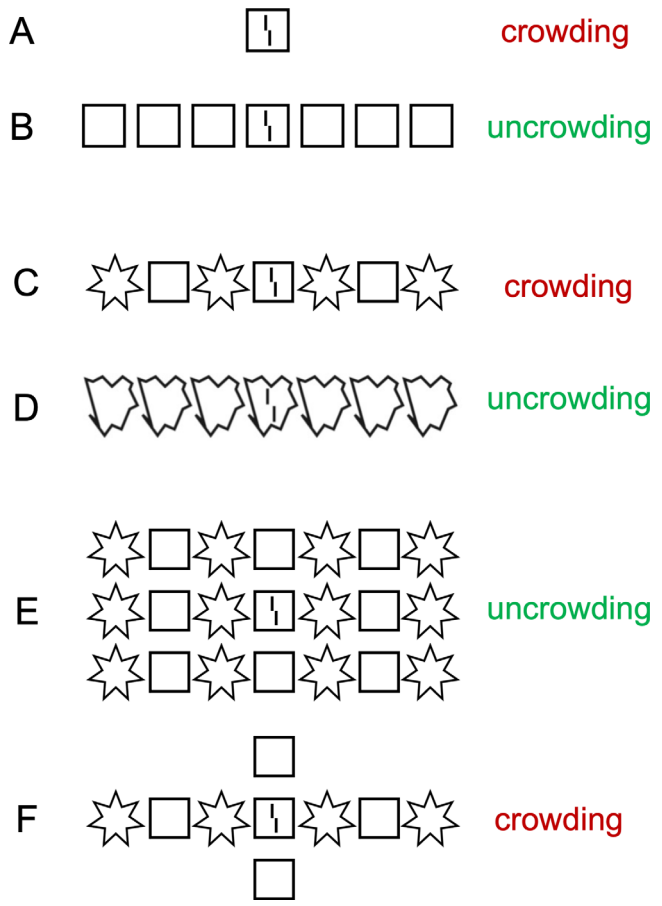

Figure 9. Illustrations of stimulus configurations used in different studies. (**A**) and (**B**) are from Manassi et al. (2013) and **C**–**F** are from Manassi et al. (2016). The labels to the right of each example indicate whether human participants exhibited crowding or uncrowding with those stimulus configurations.

be accounted for based on the output (e.g., Bornet et al., 2021; Doerig et al., 2019; Rosenholtz et al., 2019). This is the strategy described above with regard to the stimulus-configuration effects on crowding. This approach is limited, however, by the specific assumptions of the models, and by detailed decisions that must be made to implement them.

We end with a discussion of a particularly compelling and robust form of uncrowding that has been a focus of the debate surrounding midlevel mediation of crowding because it has so far defied explanation in terms of any implemented feedforward image-based model (Bornet et al., 2021; Doerig et al., 2019; Rosenholtz et al., 2019). Figures 9A and 9B illustrate the original version that was reported by Manassi et al. (2013). Enclosing a Vvernier target in a square reduces discriminability compared to when the target is alone, which is a basic crowding effect (Figure 9A). However, adding more squares improves performance relative to just the single square (Figure 9B). That's uncrowding. A recent study compared the ability of a wide range of different

models to account for this uncrowding effect (Doerig et al., 2019). The only model that yielded uncrowding for displays like that in Figure 8B was LAMINART (Francis et al., 2017), which has a recurrent rather than a feedforward organization, allows stimuli well beyond critical spacing to influence a stimulus' representation, and most important for the current discussion, includes a grouping process. It is the only model that was tested that has a grouping process. Essentially, LAMINART representationally connects the multiple surrounding squares into a separate representational layer, analogous to a surface, based on the alignment of their top and bottom edges; they become connected through illusory contours. Because the target does not share that alignment, it is represented separately on its own layer. A single square without the support of aligned surrounding squares is insufficient to segment the target and square into separate representational layers. A template-matching process, which is applied at the level of layers, yields better performance when target and flankers are represented on separate layers than when they are not.

Although image-based explanations of crowding that are based on local pooling regions cannot, in any obvious way, account for the basic uncrowding effect in Figure 9B, variations of these displays demonstrate that no current account of it, including LAMINART or less formal accounts in terms of Gestalt grouping principles (e.g., Herzog et al., 2015) can account for it. LAMINART, for example, yields uncrowding for the version shown in Figure 9C (Doerig et al., 2019), whereas humans show crowding (Manassi et al., 2016; see also Manassi et al., 2012; Rosen & Pelli, 2015). Conversely, LAMINART yields crowding for the version shown in in Figure 9D, whereas humans show uncrowding (Manassi et al., 2016). LAMINART has not been applied to the two variations in Figures 8E and 8F because it is not designed for those stimuli, but describing these displays in terms of traditional Gestalt grouping, it would seem that the vertical boxes should be equally well grouped with each other, and away from the target, in the two versions. Yet, 9E yields uncrowding, and 9F yields crowding (Manassi et al., 2016). Finally, the original uncrowding effect for displays like those in 9B fails to yield uncrowding when the rows are oriented obliquely instead of vertically or horizontally (Choung, Bornet, Doerig, Herzog, 2021). It is not obvious why oblique orientations would fail to group while vertical and horizontal ones do; illusory contours are experienced regardless of orientation. These and related findings (e.g., Choung, et al., 2021; Manassi et al, 2016; Rosen & Pelli, 2015) represent a complex set of results that so far seem to us anyway to defy simple explanations in terms either midlevel or image-level processes. It seems clear that a recurrent architecture and influence from stimuli beyond the range defined by critical spacing are necessary for a full accounting of the range of effects that have been reported, but neither of those attributes—recurrence nor global influence—necessitates midlevel mediation of crowding.

In summary, the results of five experiments that were designed to provide evidence of midlevel mediation of visual crowding failed to yield evidence of it. We draw a distinction between midlevel mediation which involves representations of 3D scene structure and image segmentation which is limited to representations of 2D image-level information (see also Moore et al., 2021). This work is not offered as evidence against the possibility of midlevel mediation of crowding more generally because it is limited in scope to the one uncrowding effect assessed. It does, however, contribute to the balance of evidence concerning the possibility that visual crowding is caused mostly, if not entirely, by properties of image-level visual processes, separate from the establishment of midlevel representations.

Footnotes

[1]The term *object level* has also been used in the literature, but that term often connotes a broader scope than we intend (i.e., to include semantic and categorical aspects of representation).
[2]When partial eta squared is very small, adjusted partial eta squared can become negative because it adjusts for a positive bias.

Keywords: crowding, uncrowding, midlevel processing, perceptual organization, surface completion

Acknowledgments

Supported in part by NIH grant R21 EY029432 and NSF grant BCS 2319133 to CMM.

Commercial relationships: none.
Corresponding author: Cathleen Moore.
Email: cathleen-moore@uiowa.edu.
Address: Department of Psychological and Brain Sciences, G60 PBSB, University of Iowa, Iowa City, IA, 52242.

References

Anstis, S. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research, 14*, 589–592, https://doi.org/10.1016/0042-6989(74)90049-2.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral

vision explains visual crowding. *Journal of Vision, 9*(12):13, 1–18, https://doi.org/10.1167/9.12.13.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics, 51*(2), 145–162, https://doi.org/10.3758/BF03212239.

Bornet, A., Choung, O.-H., Doerig, A., Whitney, D., Herzog, M. H., & Manassi, M. (2021). Global and high-level ef- fects in crowding cannot be predicted by either high-dimensional pooling or target cueing. *Journal of Vision, 21*(12):10, 1–25, https://doi.org/10.1167/jov.21.12.10..

Bouma, H. (1970). Interactional effects in parafoveal letter recognition. *Nature, 226*, 177–178, https://doi.org/10.1038/226177a0.

Chen, Z. (2012). Objecg-based attention: A tutorial review. *Attention Perception & Psychophysics, 74*, 784–802, https://dx.doi.org/10.3758/s13414-012-0322-z.

Chicherov, V., Plomp, G., & Herzog, M. (2014). Neural correlates of visual crowding. *Neuroimage, 93*, 23–31, https://doi.org/10.1016/j.neuroimage.2014.02.02.

Choung, O.-H., Bornet, A., Doerig, A., & Herzog, M. H. (2021). Dissecting (un)crowding. *Journal of Vision, 21*(10):10, 1–20, https://doi.org/10.1167/jov.21.10.10.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45, https://doi.org/10.20982/tqmp.01.1.p042.

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.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception & Performance, 15*(3), 448–456, https://doi.org/10.1037/0096-1523.15.3.448.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149, https://doi.org/10.3758/BF03203267.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12*(2), 201–204, https://doi.org/10.3758/BF03212870.

Felisberti, F. M., Solomon, J. A., & Morgan, M. J. (2005). The role of target salience in crowding. *Perception, 34*(7), 823–833, https://doi.org/10.1068/p5206.

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.

Freeman, J., & Simoncelli, E. P. (2011) Metamers of the ventral stream. *Nature Neuroscience, 14*, 1195–1201, https://doi.org/10.1038/nn.2889.

Harms, L., & Bundesen, C. (1983). Color segregation and selective attention in a nonsearch task. *Perception & Psychophysics, 33*(1), 11–19, https://doi.org/10.3758/BF03205861.

Huckauf, A., Heller, D., & Nazir, T. A. (1999). Lateral masking: Limitations of the feature interaction account. *Perception & Psychophysics, 61*, 177–189, https://doi.org/10.3758/BF03211958.

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

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

Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous features provides a unifying account of crowding. *Journal of Vision, 16*(3):39, 1–15, https://doi.org/10.1167/16.3.39.

Kimchi, R., & Pirkner, Y. (2015). Multiple level crowding: Crowding at the object parts level and at the object configural level. *Perception, 44*(11), 1275–1292, https://doi.org/10.1177/0301006615594970.

Lange, K., Ku¨hn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE, 10*(6). e0130834, https://doi.org/10.1371/journal.pone.0130834.

Levi, D. M. (2014). Visual crowding. In J. S. Werner, & L. M. Chalupa (Eds.), *The New Visual Neurosciences*. Cambridge, MA, USA: MIT Press.'.

Levi, D. M., & Carney, T. (2011). The effect of flankers on three tasks in central, peripheral, and amblyopic vision. *Journal of Vision, 11*(1):10, 1–23, https://doi.org/10/1167/11.1.10.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision, 7,* 4.1–12, https://doi.org/10.1167/7.2.4.

Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision, 10,* 1.1–1.14, https://doi.org/10.1167/10.3.1.

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., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision, 16*(3):35, 1–13, https://dx.doi.org/10.1167/16.3.35.

Mathoˆt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open- source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324, https://dx.doi.org/10.3758/s13428-011-0168-7.

Moore, C. M., He, S., Zheng, Q., & Mordkoff, J. T. (2021). Target-flanker similarity effects reflect image segmentation not perceptual grouping. *Attention, Perception, & Pscyhophysics, 83*, 658–675, https://doi.org/10.3758/s13414-020-02094-z.

Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared (adj ηp2). *Advances in Methods and Practices in Psychological Science, 2*(3), 228–232. http://doi.org/10.1177/2515245919855053.

Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap resolution in central and peripheral vision. *Vision Research, 32*, 771–777, https://doi.org/10.1016/0042-6989(92)90192-l.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64, https://doi.org/10.20982/tqmp.04.2.p061.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12), 1136–1169, https://doi.org/10.1167/4.12.12.

Pirkner, Y., & Kimchi, R. (2017). Crowding and perceptual organization: Target's objecthood influences the relative strength of part-level and configural-level crowding. *Journal of Vision, 17*(11):7, 1–18, https://doi.org/10.1167/17.11.7.

Rosen, S., & Pelli, D. G. (2015). Crowding by a repeating pattern. *Journal of Vision, 15*(6):10, 1–9, https://dx.doi.org/10.1167/15.6.10.

Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science, 2*(1), 437–457, https://doi.org/10.1146/annurev-vision-082114-035733.

Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision, 19*(7):15, 1–25, https://doi.org/10.1167/19.7.15.

Sayim, B, & Cavanagh, P (2013) Grouping and crowding affect target appearance over different spatial scales, *PLoSONE, 8*(8), e71188, https://dx.doi.org/10.1371/journal.pone.0071188.

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

Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic Bulletin & Review, 29*, 1531–1557, https://doi.org/10.3758/s13423-022-02117-w.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15,* 160–168, https://doi.org/10.1016/j.tics.2011.02.005.

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