Supplementary materials

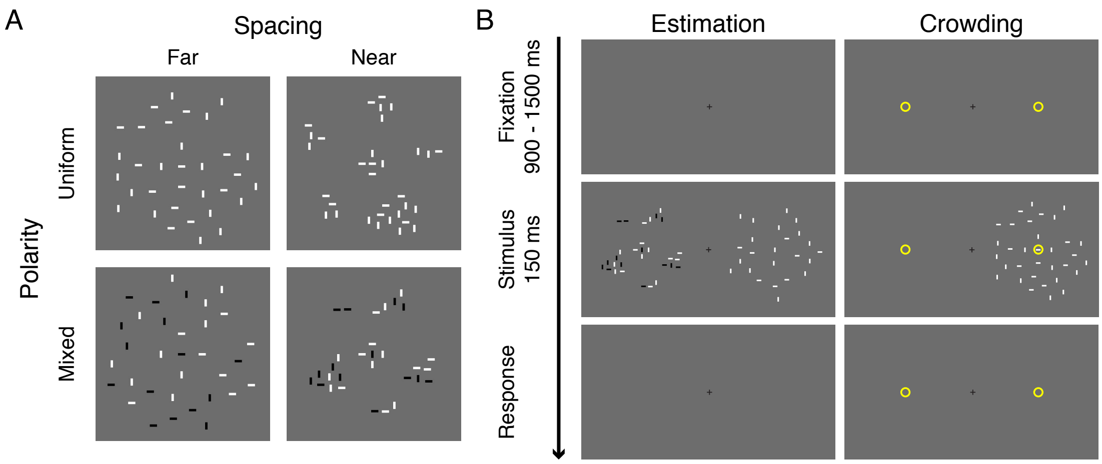
S1. Supplementary Experiment 1

In the main study, we compared performance in two tasks, identification and enumeration, within the same participants. Here, we tested separate sets of participants on the numerosity estimation and crowding tasks, respectively. We also used a different set of stimuli to assess the generality of our findings. We presented participants with a patch of oriented rectangles and asked them to either estimate their numerosity (relative to a reference patch) or identify the orientation of the central line. As in the main experiments, we manipulated two factors in the patch. To test the effect of clustering, we varied the minimal spacing between the rectangles (near or far). Second, to test the effect of similarity, we varied the contrast polarity of the rectangles: all rectangles had the same colour (black *or* white) or roughly half of them were black and the other half were white. We assessed if these two factors affected the two tasks in the same way.

Methods

Participants

Twenty-two observers participated in the numerosity estimation task and a different set of 22 observers participated in the crowding task. We excluded two participants from the estimation task, since their PSE estimates were more than 100 elements, and one participant from the identification task, since they were at chance. We conducted a subsequent crowding experiment as a control (see below for the rationale of this control) with an additional 22 observers. All participants provided informed consent. The study was approved by the Psychology Ethics committee at the University of Aberdeen.



Supplementary Figure 1: Stimulus configurations and presentation sequence. A) Configurations of the probe patch. A patch of 15-33 rectangular elements was presented on each trial. The rectangles were either vertical or horizontal. The polarity (uniform or mixed) of these elements and spacing between them (near or far) was manipulated to obtain four conditions. B) Stimulus sequence for the numerosity estimation experiment (left) and the crowding experiments (right). After a variable fixation period of 900 – 1500 ms, the stimulus was presented for 150 ms followed by a blank screen until response. Participants reported the patch (left/right) with the greater number of elements in the estimation task and the orientation of the rectangle (vertical/horizontal) within the yellow circle in the crowding task.

Materials and stimuli

The materials were the same as in the main experiments. Elements were black and white rectangles presented on a grey background. The rectangles were 0.2° by 0.7° and oriented either vertically or horizontally. Supplementary Figure 1 provides examples of stimuli and illustrates the procedure. Elements were presented within a circular region (diameter 15°) centred at 10° eccentricity along the horizontal axis. In the numerosity experiment, two such patches were presented simultaneously. The *reference* patch had 24 rectangles (density = 0.14 items/deg^2). The numerosity of the *probe* patch varied from 15 to 33 items in steps of 3.

The probe patch was constructed as in the main experiments: one rectangle was placed at the centre of the patch. This element was surrounded by four flanking rectangles, one in each cardinal direction (top, bottom, left, right). The remaining rectangles (10-28) were distributed according to an algorithm. First, five *seed* rectangles were placed at equidistant points along the edge of the patch (7.5° from the centre of the patch). Then, each of the remaining elements was added to the patch in turn as described in Experiment 1A.

The elements could be either *near* or *far*: the minimal spacing between elements was either 1° or 2°, respectively. Second, all elements could share the same contrast polarity (*uniform* polarity condition; all white or all black elements) or could be of *mixed* polarity: the central rectangle had a different polarity relative to its four surrounding flankers (white rectangle surrounded by black flankers, or vice versa). Each of the remaining elements was randomly assigned white or black colour. The orientation of each rectangle was randomly assigned.

We conducted two crowding experiments. The stimuli used in the first (Crowding1) were an exact match for the estimation experiment. The second (Crowding2) had additional conditions, but tested fewer numerosities. In both, only one patch, with a variable number of elements, was presented. This patch was constructed according to the procedure for the probe patch. Here, unlike in the main experiments, the target is not distinguishable from the flankers. Therefore, we also presented two yellow rings of diameter 1.2°, one on either side of fixation, at the centre of the patches (10° eccentricity) throughout the experiment to indicate the location of the target. They were larger than the target rectangle and hence there was no spatial overlap. These rings should not cause any crowding since they had a different colour and shape relative to the target (Kooi, Toet, Tripathy, & Levi, 1994) and, importantly, they were present throughout. It is well established that distractor preview diminishes crowding (e.g., Scolari, Kohnen, Barton, & Awh, 2007).

Procedure

Estimation: Each probe numerosity (15-33) was tested on four conditions (2 spacing: near or far x 2 polarity: uniform or mixed). There were 40 trials for each numerosity and condition combination. The conditions were randomly intermixed within blocks of 84 trials. Observers were trained on one block of 56 practice trials before the main experiment.

Each block began with a key press. Two patches appeared in the periphery for 150 ms. Participants were asked to report, via an appropriate key, which patch had more elements. No feedback was provided. The locations of the probe and reference patches were randomised. The next trial began 900 – 1500 ms after the response.

Crowding1: The same range of numerosities used in the estimation experiment were tested. The sequence was the same, except that a single patch was presented either to the left or the right of fixation. Observers reported the orientation (horizontal or vertical) of the target rectangle (the one within the yellow ring) with a key press.

Crowding2: Only numerosities 15, 21, 27 and 33 were tested. However, there was an additional baseline condition with only 5 elements (central target + four flankers) to assess crowding for our stimuli in a typical crowding setup.

Results

Estimation

Supplementary Figure 2A plots the proportion of cases in which the probe patch was reported to have higher numerosity than the reference patch, as a function of numerosity. As expected, this *proportion more* response increased with probe numerosity. As in Experiment 1A (main text), we had intended to fit psychometric curves to these data to estimate the point of subjective equality (PSE). However, performance in the near spacing conditions did not reach halfway (0.5) of the response range (Supp. Fig. 2A, filled symbols). Since the ascending part of a typical psychometric curve is approximately linear, we fit straight lines to the *proportion more* and extracted PSE estimates from these fits.

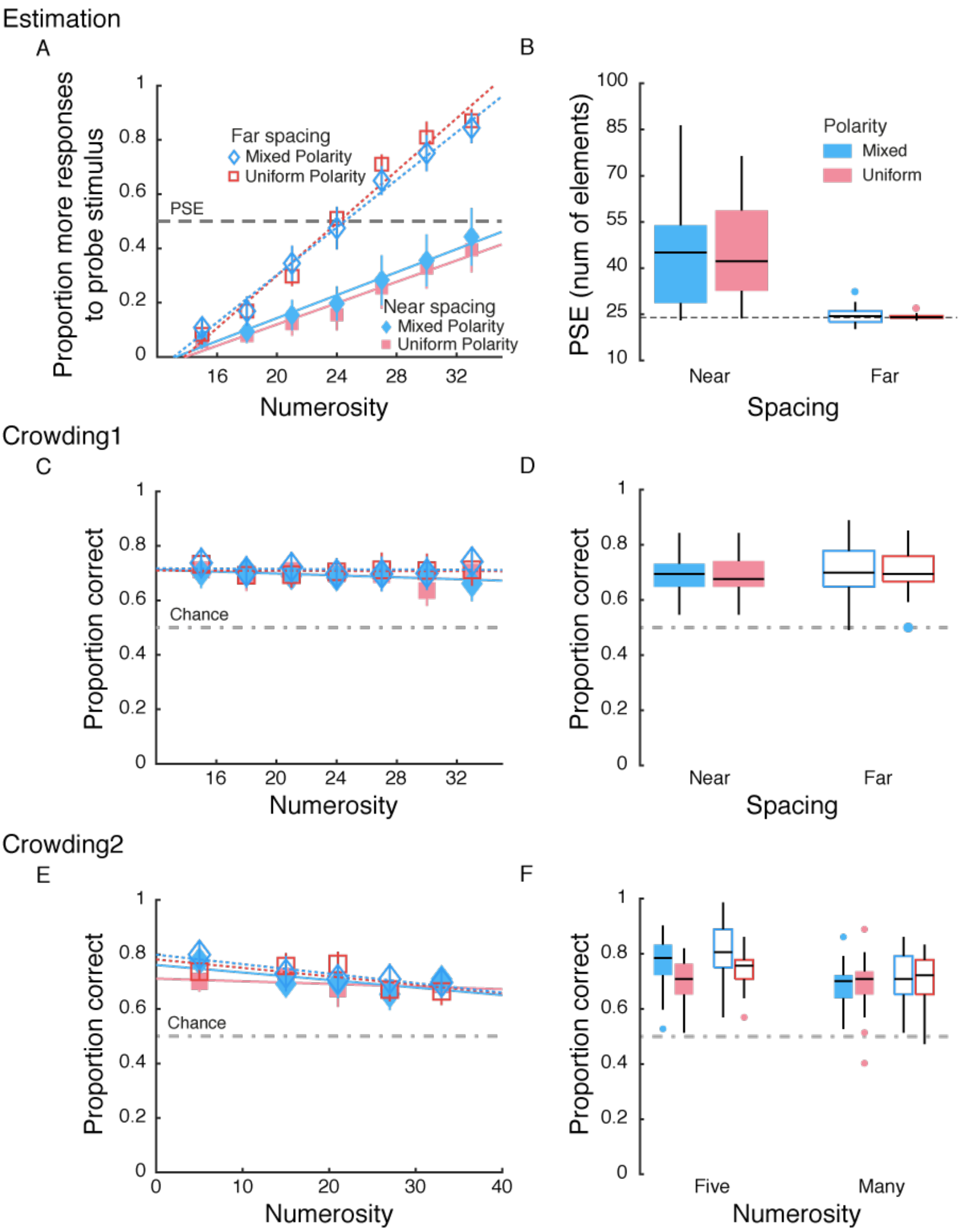
The PSEs were higher, roughly twice as high, in the *near* condition (45.5±3.3 items) compared to the *far* condition (24.5±0.4), suggesting that there was substantial underestimation of clustered elements ($F(1,19)=44.5$, $p<0.0001$, ${}_p\eta^2=0.7$). Further, there was no effect of polarity on the PSEs ($F(1,19)=0.004$, $p=0.95$, ${}_p\eta^2<0.001$), also replicating previous findings (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011). There was no interaction between the two factors ($F(1,19)=0.11$, $p=0.74$, ${}_p\eta^2=0.006$).

Crowding1

Supplementary Figure 2C plots accuracy in identifying the rectangle orientation in each condition as a function of numerosity. Supplementary Figure 2D summarises the data by collapsing across numerosities. There was surprisingly no effect of spacing ($F(1,21)=2.63$, $p=0.12$, ${}_p\eta^2=0.11$) on target identification, with performance at closer spacing (0.69±0.01) very similar to that at far spacing (0.71±0.02). Similarly, participants identified the target just as well when it was surrounded by same polarity flankers as when it was surrounded by opposite polarity flankers ($F(1,21)=0.1$, $p=0.75$, ${}_p\eta^2=0.005$). There was also no interaction between the two factors ($F(1,21)=0.09$, $p=0.76$, ${}_p\eta^2=0.004$). In short, our manipulations did not alter performance much with this stimulus set. However, this was not just a floor effect, as performance was higher than chance (0.5) in all conditions. In fact, performance here was better than in Experiment 2B, which was also used a 2AFC identification task, where we did observe the effects of our manipulations. Hence, we can surmise that the pattern of results observed in the estimation task here is different from the one observed in the identification task, with the same configuration of elements.

One explanation of these results might be that in dense displays our manipulations do not produce standard effects. If this were the case, then we should not have observed the effects of spacing on estimation either. Further, the experiments described in the main text have demonstrated that these manipulations can modulate performance even in dense displays, although more weakly than in the standard crowding configuration. An alternative explanation derives from the fact that the current task, unlike in the other experiments, required only coarse (orientation) discrimination. It has been argued that coarse discrimination might not be susceptible to crowding (Pelli, Palomares, & Majaj, 2004). If that were the case, it would raise further objections to the crowding hypothesis, since if the elements cannot be crowded we should not observe underestimation, whereas here we do.

To test if these stimuli might not have been susceptible to crowding and to compare performance in dense displays and standard crowding configurations, we conducted a follow-up identification experiment.



Supplementary Figure 2: Results of Supplementary Experiment 1. A) Performance in the estimation task as a function of numerosity. Also shown are straight line fits. Error bars represent 95% CI. B) Box plots of PSEs extracted from the fits in the estimation task. C) Performance (proportion correct responses) in the crowding task as a function of numerosity. Note that chance here is 0.5. Error bars represent 95% CI. D) Box plots of performance in the four conditions averaged across numerosities. E) Performance (proportion correct) plotted as a function of numerosity in the second crowding experiment in each of the four conditions. F) Box plots of performance in each of the four conditions. The left set of four boxplots depicts performance in the numerosity five (four-flanker) condition. The right set depicts performance at higher numerosities (collapsed across numerosities 15 – 33).

Crowding2

In this experiment, we assessed crowding in a traditional crowding setup (one target with four flankers) and compared it to crowding in our previous setup. Supplementary Figure 2E plots identification accuracy in the four spacing x polarity conditions at each numerosity (5, 15-33) separately. Supplementary Figure 2F depicts performance when the target was surrounded by either four flankers or by many flankers (collapsed across the higher numerosities).

We first analysed the data from the four-flanker condition. Target identification performance in the mixed polarity condition (0.79±0.02) was higher than in the uniform polarity condition (0.72±0.02; $F(1,21)=40.5$, $p<0.0001$, ${}_p\eta^2=0.66$). Further, there was a small (~3%) effect of spacing on crowding. As expected from the crowding literature, near flankers (0.74±0.02) reduced performance more than far flankers (0.77±0.02; $F(1,21)=12.5$, $p=0.002$, ${}_p\eta^2=0.37$). Finally, there was no interaction between the two factors ($F(1,21)=0.23$, $p=0.64$, ${}_p\eta^2=0.01$).

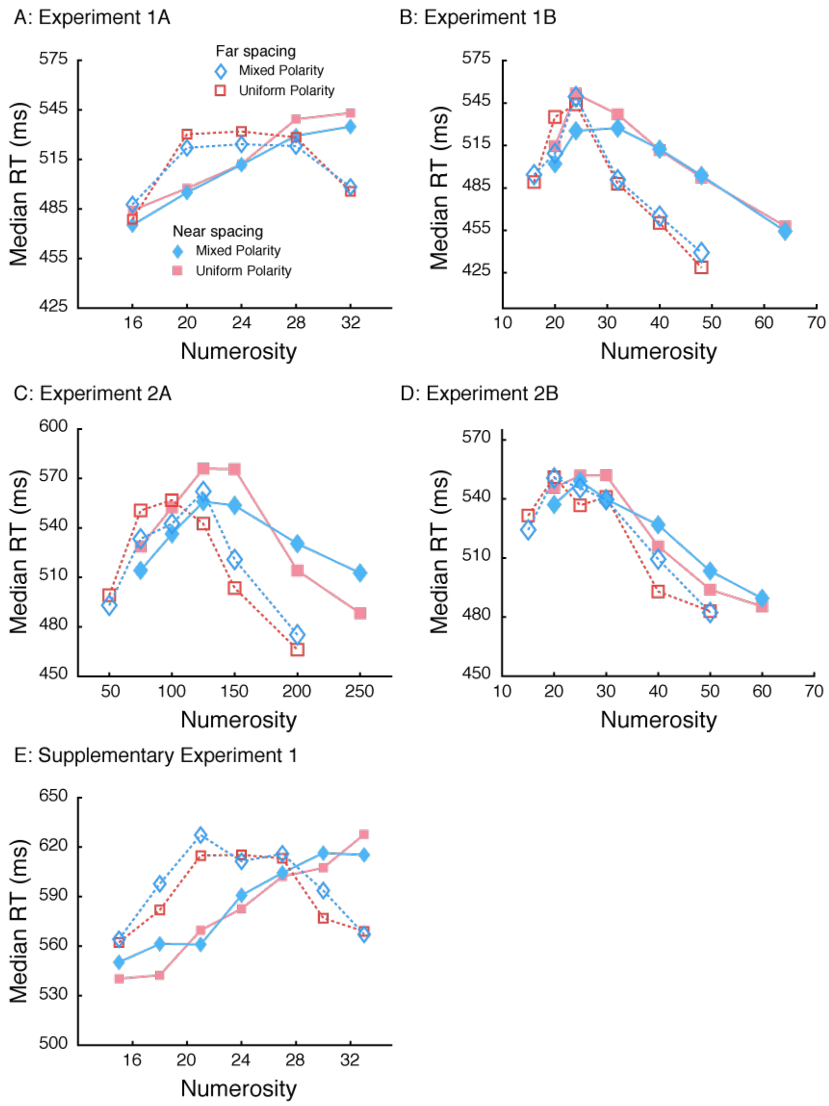
In contrast, the pattern of results was different in the many-flankers condition (with data collapsed across the various numerosities). Unlike in the four-flankers condition, there was no effect of polarity ($F(1,21)=0.03$, $p=0.88$, ${}_p\eta^2<0.01$). However, there was a small effect of spacing. Performance was slightly better when the flankers were farther away (0.71±0.02) compared to when they were closer (0.69±0.02) to the target ($F(1,21)=4$, $p=0.06$, ${}_p\eta^2=0.16$). No interaction was observed between the two factors. Performance in the dense displays (multi-flanker conditions) was worse than in the four-flanker conditions, indicating that the targets did suffer from crowding despite using a coarse discrimination task.

Crowding1 & Crowding2

The results from the two crowding studies taken together show that, in the presence of a large set of elements, factors such as spacing and polarity have, at best, small and inconsistent effects on identification. In the presence of a small number (four) of flankers, these factors modulated crowding in the expected direction, but the effects were rather weak with oriented rectangles. Even when crowding was modulated (four-flanker conditions), contrast polarity was more influential than spacing. However, the pattern of results was noticeably different in the estimation task, where spacing had a substantial effect on estimation but contrast polarity had no effect. These results speak against the crowding hypothesis of cluster-induced underestimation.

S2. Reaction time analysis in the estimation task

In the estimation task, participants were asked to indicate which of two patches, the probe (variable numerosity) or the reference (fixed numerosity), had more elements. We analysed the median reaction times to these responses (Supplementary Figure S3). The results demonstrate a clear difference in response patterns for near (solid lines) and far (dashed lines) spacing conditions, but not a noticeable difference between the similarity conditions. At larger probe numerosities reaction times were slower when elements were clustered than when they were spaced apart. Also, the peak of the RT 'distribution' (the slowest responses) was shifted rightwards in the near spacing conditions. This indicates that the most difficult decision was at a higher reference numerosity in the near spacing condition than in the far spacing condition. These results once again suggest that numerosity estimation was severely impaired by spacing but not much by similarity between objects.



Supplementary Figure S3. Median reaction times in the estimation task in Experiments 1A (Panel A; 24 Squared Thetas), 1B (Panel B; 24 Squared Thetas), 2A (Panel C; 100 Lines), 2B (Panel D; 25 Lines) and Supplementary Experiment 1 (Panel E; 24 Rectangles).

S3. Comparing crowded performance across various flanker numerosities

We compared performance in the standard crowding configurations (target + four flankers in Experiment 1 and target with two flankers in Experiment 2) with that in the multi-flanker displays to test if adding flankers affects performance.

Experiment 1A

We performed a three-way repeated measures ANOVA with spacing (near or far), similarity (uniform or mixed polarity) and numerosity (few or many) as factors and identification performance as the dependent variable. Interestingly for our comparison, there was an effect of numerosity ($F(1,19)=21.8$, $p<0.0001$, $_p\eta^2=0.53$): performance in the four-flanker condition was higher than in the many-flankers condition. In other words, adding flankers increased crowding. As expected, inter-element spacing modulated identification performance ($F(1,19)=39.9$, $p<0.0001$, $_p\eta^2=0.68$). There was also an effect of similarity ($F(1,19)=34.5$, $p<0.0001$, $_p\eta^2=0.65$) on target identification. Performance in the uniform polarity condition was worse than in the mixed polarity condition.

None of the interactions were significant (all *p*s > 0.55), except the interaction between similarity and spacing ($F(1,19)=4.7$, $p=0.042$, $_p\eta^2=0.2$). This finding suggests that the effect of similarity (better performance in the mixed polarity condition compared to uniform polarity condition) was stronger in the far condition than in the near condition.

Experiment 1B

We conducted the same three-way ANOVA in Experiment 2B. Once again, the number of flankers affected identification performance ($F(1,18)=9.4$, $p=0.007$, $_p\eta^2=0.34$): performance was worse when there was a large number of flankers compared to the standard crowding paradigm. As expected, increased spacing between elements led to better performance ($F(1,18)=32.5$, $p<0.0001$, $_p\eta^2=0.64$). Further, similarity between elements also modulated performance, with similar flankers leading to worse identification than dissimilar flankers ($F(1,18)=38.6$, $p<0.0001$, $_p\eta^2=0.68$).

There was an interaction between numerosity and spacing ($F(1,18)=7$, $p=0.017$, $_p\eta^2=0.28$), where the effect of spacing was stronger in the many flanker condition than in the standard crowding condition. On the other hand, there was an interaction between numerosity and similarity ($F(1,18)=5.5$, $p=0.03$, $_p\eta^2=0.24$), which was driven by a stronger effect of similarity in the standard crowding condition than in the many flanker condition. No other interactions were significant.

Experiment 2A

As in previous experiments, adding flankers to the standard crowding configuration further reduced identification performance ($F(1,20)=65$, $p<0.0001$, $_p\eta^2=0.77$). As expected, near flankers reduced performance relative to far flankers ($F(1,20)=14.5$, $p=0.001$, $_p\eta^2=0.42$). Similar flankers caused more crowding than dissimilar flankers ($F(1,20)=25.9$, $p<0.0001$, $_p\eta^2=0.57$). None of the interactions were significant except between numerosity and similarity ($F(1,20)=14.2$, $p=0.001$, $_p\eta^2=0.41$). As in Experiment 1B, this interaction indicates that the effect of similarity (better performance in the presence of dissimilar flankers compared to similar flankers) was stronger in the standard crowding setup (2 flankers) compared to when there were many flankers. This is in line with previous findings of weakened similarity effect when there were multiple flankers of alternating polarity (Manassi, Sayim, & Herzog, 2012; Rosen & Pelli, 2015).

Experiment 2B

A three-way repeated measures ANOVA indicated that the number of flankers affected identification: performance worsened with a larger number of flankers ($F(1,14)=38.8$, $p<0.0001$, ${}_p\eta^2=0.73$). Closely spaced elements caused more crowding than distant flankers ($F(1,14)=12.1$, $p=0.004$, ${}_p\eta^2=0.46$). Similarity between elements also affected performance: similar flankers led to more crowding than dissimilar flankers ($F(1,14)=19.3$, $p=0.001$, ${}_p\eta^2=0.58$). There was an interaction between numerosity and spacing ($F(1,14)=5.4$, $p=0.035$, ${}_p\eta^2=0.28$) where the effect of spacing was stronger in the standard crowding setup (two flankers) than in the multi-element setup. As in Experiment 1A, there was an interaction between spacing and similarity ($F(1,14)=6.37$, $p=0.024$, ${}_p\eta^2=0.31$), where the effect of similarity was much stronger at far spacing than at near spacing. The other interactions were not significant, although the three-way interaction was marginal ($F(1,14)=3.7$, $p=0.075$, ${}_p\eta^2=0.21$), and was presumably driven by the substantially high performance when the flankers were of the opposite polarity to the target, and were separated from the latter by far spacing and when their number was only two.

All four experiments showed that adding flankers increased crowding, as expected from standard theories of crowding. The additional flankers also appeared to modulate the effect of similarity and spacing in certain configurations.

References:

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279.

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

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

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

Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision*, *7*(2), 7–7.