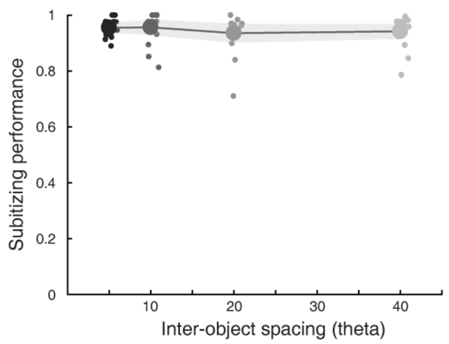
Supplementary Results

S1: Experiment 1

S1.1: Subitizing performance

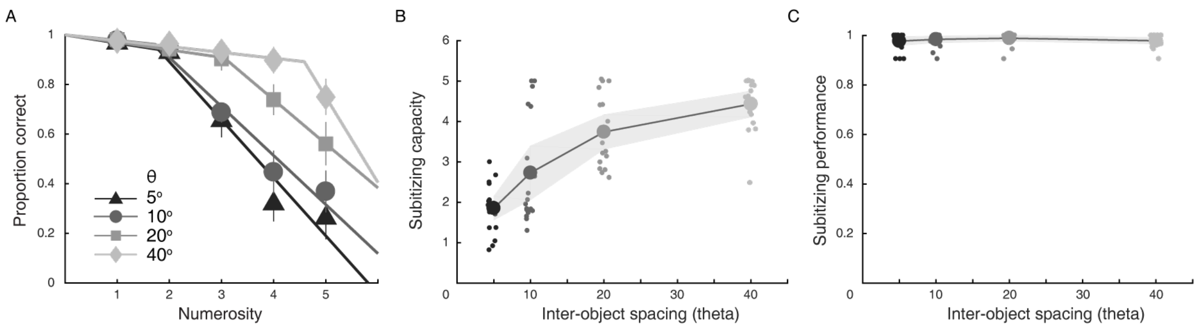
Subitizing performance is the highest accuracy that the participant can achieve for a small number of objects. Supplementary Figure S1 shows that subitizing performance was high and comparable across all inter-object spacings (F (3, 51) = 59.37, p < 0.001, $_p\eta^2$ = 0.67), indicating that enumeration was efficient at all inter-object distances.



Supplementary Figure S1: Experiment 1 results. Mean and individual (one symbol per participant) subitizing performance are plotted as a function of inter-object spacing. Shaded region represents 95% CI. Subitizing performance remains high and similar across spacings, indicating that small numerosities could be accurately reported.

S1.2: Bilinear fits with varying slopes

We estimated subitizing capacity and performance using bilinear fits. For the results described in the main text (section 2.2), we fixed the slope of the first line to zero while allowing the slope of the second to vary. Here, we allowed the slopes of both lines to vary (Supplementary Fig S2A). The results were similar in both approaches, with a modest increase in estimated capacity in the latter method. Subitizing performance (Supplementary Fig S2C) was high and comparable across all spacing conditions ($F(2.6, 44) = 1.03$, $p = 0.381$, ${}_{p}\eta^2 = 0.06$; Huyn-Feldt correction applied, based on violation of sphericity). On the other hand subitizing capacity (Supplementary Fig S2B) increased with spacing ($F(3, 51) = 37.02$, $p < 0.001$, ${}_{p}\eta^2 = 0.67$). Pairwise comparisons, reported in Supplementary Table S1, between the spacing conditions showed that capacities were low and comparable to each other when the objects were within Bouma's bound of each other, but higher beyond it.



Supplementary Figure S2: A. Bilinear fits to the same data as presented in the main text, but with the slopes of both lines in the fit allowed to freely vary. Error bars are 95% CI **B.** Subitizing capacities estimated with this method as a function of spacing between the objects. Shaded areas represent 95% CI and each dot represents one participant. **C.** Subitizing performance estimated with the latter method as a function of spacing.

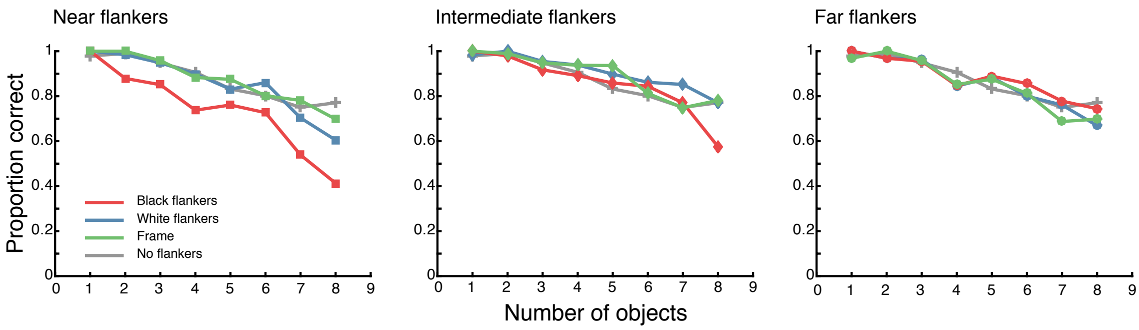
Supplementary Table S1. Pairwise t-tests (Bonferroni corrected for multiple comparisons) of subitizing capacities between the four spacing conditions, when the slopes of both lines in the bilinear fit could vary.

Spacing	5°	10°	20°	40°
5°	*1.86 ± 0.1*			
10°	$t(17) = 2.49$; $p = 0.12$	*2.73 ± 0.3*		
20°	$t(17) = 7.83$; $p < 0.001$	$t(17) = 5.27$; $p < 0.001$	*3.74 ± 0.2*	
40°	$t(17) = 12.49$; $p < 0.001$	$t(17) = 5.25$; $p < 0.001$	$t(17) = 3.13$; $p = 0.024$	*4.43 ± 0.2*

S2: Experiment 2

S2.1 Accuracy data

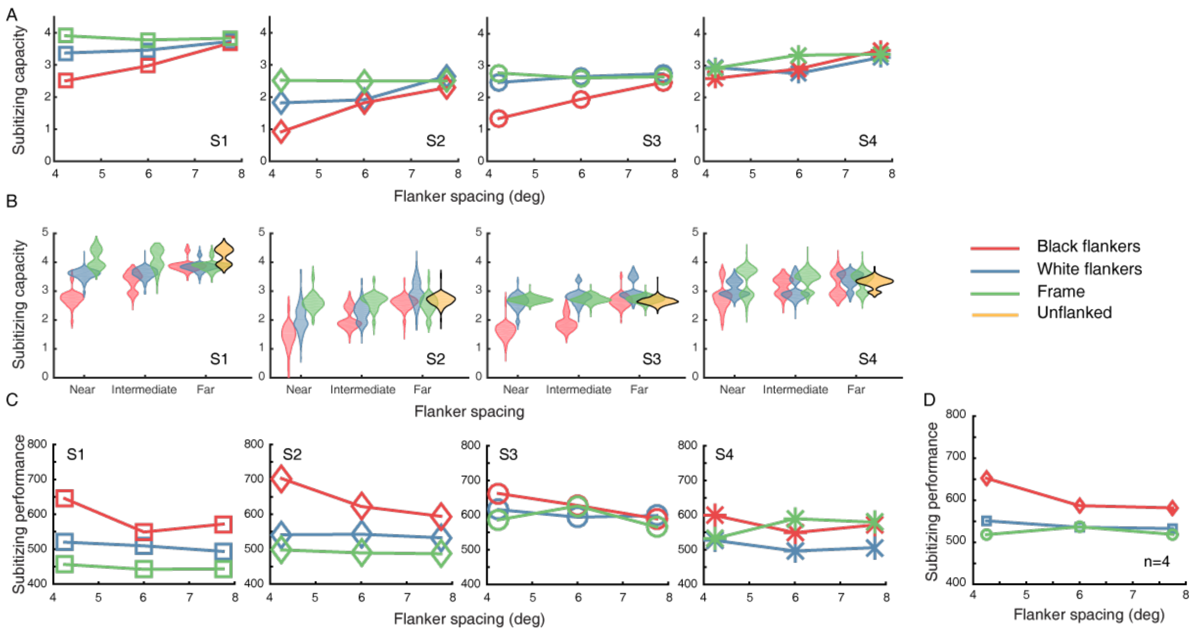
Supplementary Figure S3 plots the accuracy data averaged across the four experienced participants at three different spacings, for each of the flanker conditions. Accuracy remained reasonably high even at the largest numerosity (8) and the closest spacing for all flanker types. Note that accuracy was comparable for almost all flanker types at all spacings, with the exception of black flankers at the closest spacing, where performance was impaired. This supports the findings on reaction times, reported in the main text, that similar flankers within Bouma's bound impair enumeration performance, as can be expected from crowding of subitizing.



Supplementary Figure S3: Accuracy in reporting numerosity in Experiment 2. Each panel plots mean performance when flankers were placed at different distances from the target. The plots depict proportion correct as a function of numerosity for different flanker types.

S2.2: Individual (reaction time) data and bootstrapping

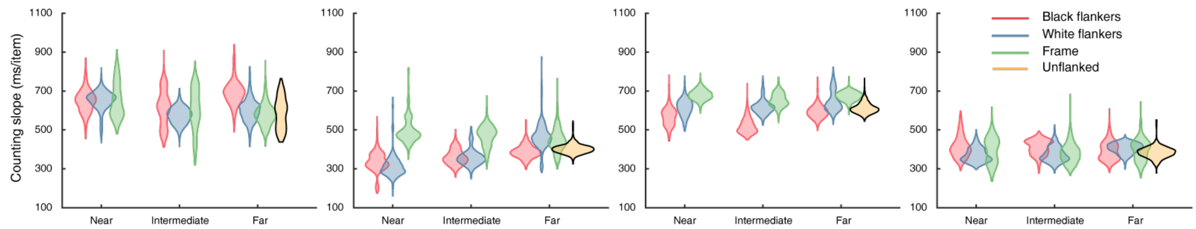
Subitizing capacity and performance for each participant in Experiment 2 is plotted in Supplementary Figure S4. We also conducted a bootstrap analysis on the reaction time data. In this analysis, we sampled, with replacement, correct reaction times for each numerosity, for each target-flanker spacing and flanker type. We estimated subitizing capacity from this resampled reaction time data using bilinear fits. We repeated this analysis 1000 times. The bootstrapped distributions of subitizing capacity are shown in Supplementary Figure S4B as violin plots.



Supplementary Figure S4: Individual data for Experiment 2. A: Subitizing capacity for each participant is plotted as a function of flanker spacing for each flanker-type. Three out of the four participants show the same pattern: Black flankers reduce subitizing capacity when they are within Bouma's bound, whereas other flankers and target-flanker spacings do not. Crowding is weak in the fourth participant. **B.** Violin plots of bootstrapped subitizing capacities at each target-flanker spacing in the presence of the three flanker types, plotted separately for each participant. **C**: Subitizing performance is plotted in the same manner as in panel A. **D**. Average subitizing performance across the four participants. Reaction times are slower for black flankers than for the other flanker-types, particularly at the closest spacing.

S2.3: The effect of flankers on counting (enumerating larger numerosities)

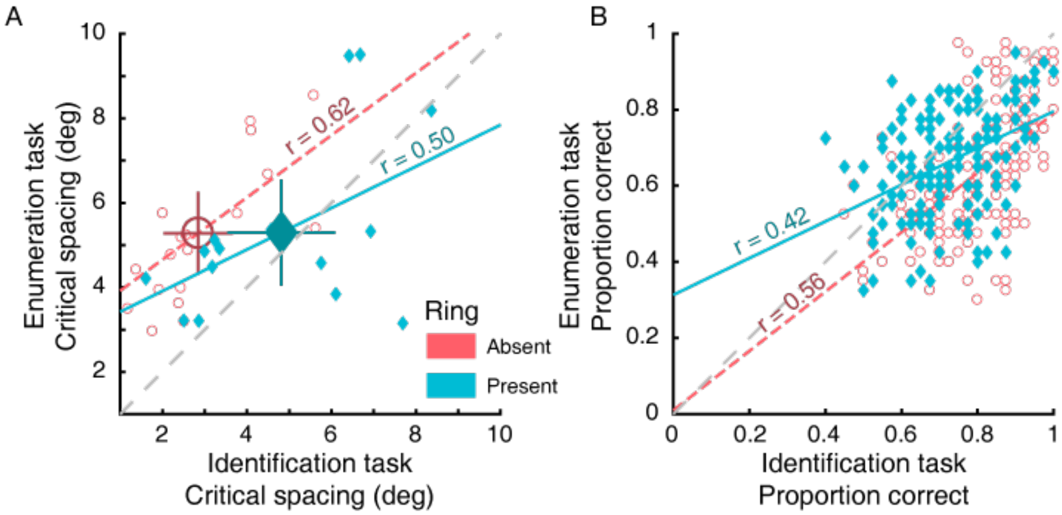
In this study we tested numerosities beyond the subitizing range (5-9), which allowed us to explore if enumeration of large numerosities, called counting, which requires attention, can be impaired by the presence of flankers. Recall that bilinear fitting involves varying the intercept of one (flat) line and varying the slope and intercept of the other. To address the question about impairment of counting, we examine the slopes of the second line, which indicates the time required to process each additional object in the counting range. It appears that flanker type and target-flanker distance do not modulate the slopes. Slopes in the presence of flankers are about the same as when no flankers are presented.



Supplementary Figure S5: Counting slopes for each participant. Violin plots of bootstrapped counting slopes at each target-flanker spacing in the presence of the three flanker types, plotted separately for each participant. The orange violin with black border is the distribution of counting slopes in the absence of any flankers.

S3: Experiment 4: Analysis of data without bootstrapping

We fit psychometric curves to individual participants' data (accuracy as a function of target-flanker distance) and extracted critical spacing from these fits (see main text for details). Some participants had poor fits and we could not estimate a critical spacing in at least one condition. Estimates from other participants were far beyond the range of tested target-flanker distances (> 10 deg) and these participants were removed from the current analysis. 10 participants were excluded due to these two constraints. The bootstrapped analysis allowed us to include all participants; hence, we believe it provides a much better estimate of population parameters and we report that in the main text. Here, we report the data analysis on the remaining 14 participants. We correlated critical spacing in the two tasks, enumeration and identification, for ring absent and present conditions separately. The data show the same relationships as reported in the main manuscript. Critical spacing (Supplementary Fig S6A) and proportion correct (Supplementary Fig S6B) are strongly correlated across the two tasks



Supplementary Figure S6: Results for Experiment 4. A. Scatterplot of critical spacing in the two tasks estimated from individual psychometric curves. Red circles represent critical spacing estimates when no ring was presented, and blue diamonds are critical spacing estimates when a ring was presented. The mean and within-subject 95% CI are shown as larger and darker symbols with error bars. Linear fits are also shown along with correlation coefficients. The dashed grey line represents the equality line, with a slope of 1. **B.** Scatter plot of proportion correct values in the two tasks at all target-flanker spacings. Best fitting straight lines and corresponding correlation coefficients are shown.

The critical spacing estimates from the curve fits were then analysed in a 2-way repeated measures (2 tasks x 2 ring presence conditions) ANOVA. Critical spacing was lower in the identification task than in the enumeration task ($F(1, 13) = 11.43$, $p = 0.005$, ${}_p\eta^2 = 0.47$). Critical spacing was lower in the absence of a ring compared to when a ring was presented ($F(1, 13) = 19.83$, $p = 0.001$, ${}_p\eta^2 = 0.6$). Interestingly, there was an interaction between task and ring presence ($F(1, 13) = 11.65$, $p = 0.005$, ${}_p\eta^2 = 0.47$). Critical spacing was lower in the identification task compared to the enumeration task when no ring was present (identification = 2.72 deg ± 0.33 versus enumeration = 5.21 ± 0.47; $t(13) = 10.72$, $p < 0.001$), but not when a ring was presented (identification = 4.56 ± 0.53 versus enumeration = 5.1 ± 0.54; $t(13) = 0.9$, $p = 0.38$).