



Capacity limit of ensemble perception of multiple spatially intermixed sets

Anna Xiao Luo¹ · Jiaying Zhao^{2,3}

Published online: 8 August 2018
© The Psychonomic Society, Inc. 2018

Abstract

The visual system is remarkably efficient at extracting summary statistics from the environment. Yet at any given time, the environment consists of many groups of objects distributed over space. Thus, the challenge for the visual system is to summarize over multiple groups. The current study investigates the capacity and computational efficiency of ensemble perception, in the context of perceiving mean sizes of multiple spatially intermixed groups of circles. In a series of experiments, participants viewed an array of one to eight sets of circles. Each set contained four circles in the same colors, but with different sizes. Participants estimated the mean size of a probed set. The set that would be probed was either known before onset of the array (pre-cue condition) or afterwards (post-cue condition). By comparing estimation error in the pre-cue and post-cue conditions, we found that participants could reliably estimate mean sizes for approximately two sets (Experiment 1). Importantly, this capacity was robust against attention bias toward individual objects in the sets (Experiment 2). Varying the exposure time to stimulus arrays did not increase the capacity limit, suggesting that ensemble perception could be limited by an internal resource constraint, rather than the speed of information encoding (Experiment 3). Moreover, we found that the visual system could not encode and hold more individual items than ensemble representations (Experiment 4). Taken together, these results suggest that ensemble perception provides an efficient way of information processing but with constraints.

Keywords Perceptual organization · Visual working memory · Selective attention

Faced with the complex information in the environment, our visual system is capable of extracting statistical summaries, including the mean, the variance, or global correlations of objects (Alvarez, 2011). This process is termed ensemble perception. One form of ensemble perception, estimation of the average, has been shown to operate accurately and quickly in many feature domains, such as size (Ariely, 2001, 2008; Chong & Treisman, 2003, 2005a), orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), location (Alvarez & Oliva, 2008), motion (Watamaniuk & McKee, 1998), and facial expression (Haberma & Whitney, 2009).

Ensemble perception is suggested to occur in an automatic manner, with little requirement of focal attention, as it can be computed from crowded objects presented in the peripheral vision (Parkes et al., 2001), or when attention is engaged by a concurrent task (Alvarez & Oliva, 2008). Extracting ensemble representations can occur rapidly: The mean size of two arrays of circles could each be accurately estimated within 50 ms of exposure (Chong & Treisman, 2003), and the mean emotion of faces could be accurately perceived within 500 ms of exposure (Haberma & Whitney, 2009). Ensemble perception can also be formed without precise representation of the individual objects constituting the group (Whitney, Haberma, & Sweeny, 2014). For instance, Parkes et al. (2001) found that although observers could not accurately estimate orientations of Gabor patches presented in the peripheral vision due to crowding, their judgment of a Gabor in the foveal vision was nonetheless biased by the average orientation of the peripheral Gabors. Similarly, Sweeny, Haroz, and Whitney (2013) showed that estimation of the average walking direction of a crowd was more accurate than the estimation of the walking direction of an individual person in the crowd, or a

✉ Anna Xiao Luo
annaxluo@gmail.com

¹ Department of Psychiatry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

² Department of Psychology, University of British Columbia, Vancouver, Canada

³ Institute for Resources, Environment and Sustainability, University of British Columbia, Vancouver, Canada

foveally presented person. Moreover, studies also show that mean size can be computed for sequentially presented circles, suggesting ensemble perception is a dynamic process actively involving working memory (Albrecht & Scholl, 2010).

Capacity limit of ensemble perception

Ensemble perception extracts the gist of an otherwise complex scene. Yet within any natural scene, there can be multiple spatially intermixed object groups, defined by similarity or proximity in one or more feature dimensions. How does the visual system extract summary statistics for multiple spatially intermixed groups of objects at the same time? Answers to this question can help us understand the interplay of selective attention, perceptual organization, and working memory in determining our perceptual experience with limited cognitive resources (Norman & Bobrow, 1975).

Earlier studies suggest that ensemble perception is likely to be a capacity-limited cognitive process. Halberda, Sires, and Feigenson (2006) examined the capacity of one specific type of summary statistics, enumeration, of one to six spatially intermixed sets of circles. They derived the capacity limit of multiset enumeration by comparing the percentage error in the pre-cue condition (where participants knew which set to enumerate prior to seeing the multiset display) and the post-cue condition (where participants did not know which set to enumerate until the multiset display was shown and disappeared). On average, participants could enumerate up to two sets, along with enumerating the whole display (the superset). Chong and Treisman (2005a) were among the first to explore ensemble perception of multiple sets of objects. In their study, circles in two different colors were spatially intermixed, and participants were asked to judge which of the two sets had a larger average size. Performance did not differ whether participants knew which set would be probed before seeing the two sets, suggesting that participants were able to compute the mean size for both of the two sets. Im and Chong (2014) examined the capacity limit of visual working memory for mean sizes of two to five sets of circles and found that approximately 2.5 sets could be remembered when the sets were spatially intermingled. These studies suggest the possibility of extracting ensemble representations for multiple spatially intermixed sets.

However, there is still a lack of systematic characterization of ensemble capacity limit with respect to attention modulation and exposure durations, and detailing the relationship between capacity of ensemble perception and that of individual object perception. The current study sought to provide further insight into these questions.

The role of attention in determining the capacity of ensemble perception is multifaceted. On one hand, the visual system is found to be sensitive to ensemble information and statistical

regularities in the stimuli with little demand of focused attention (Alvarez & Oliva, 2008, 2009; Turk-Browne, Scholl, Chun, & Johnson, 2009). On the other hand, attention toward an individual object is shown to bias estimation of summary statistics of the set. For instance, de Fockert and Marchant (2008) directed participants' attention to an individual circle in a set while asking them to estimate mean size of the set and found their estimation systematically biased toward the attended item. Chong and Treisman (2005b) also showed that mean size estimation was more accurate when distributed attention to the set, rather than focused attention to individual items, was employed. However, it remains to be examined whether biasing attention to individual items would affect how many sets could be summarized. Therefore, one of our objectives in the current study to investigate whether the capacity for ensemble perception was the same when attention is deliberately biased toward an individual object in a set.

Another objective of our study is to determine how capacity of ensemble perception would be a function of stimulus exposure. Ensemble perception has been shown to operate as rapidly as after 50–200 ms of exposure for features such as orientations and sizes (Ariely, 2001; Chong & Treisman, 2003; Dakin, 2001), yet its precision seems to improve with increased exposure (Chong & Treisman, 2003; Whiting & Oriet, 2011). Meanwhile, the Boolean map theory of attention suggests that features in the same dimension (e.g., different colors) can only be attended to one at a time (Huang & Pashler, 2007; Huang, Treisman, & Pashler, 2007; Morales & Pashler, 1999; Woodman & Luck, 2003). Thus, given a multiset display, the capacity of ensemble perception might be constrained by both (a) the rate of sampling and integrating individual objects in a set, which follows a temporal course at the millisecond scale, and (b) the speed of attending to and summarizing each set in a sequential manner, which requires shifts of attention that are suggested to have a time constant of ≤ 150 ms (Grubert & Eimer, 2016; Jenkins, Grubert, & Eimer, 2018; Woodman & Luck, 2003). In estimating the capacity of ensemble perception, effects of encoding speed need to be distinguished from effects of internal resource capacity.

Finally, we aim to compare the capacity of ensemble perception and that of individual object perception. There is a rich body of literature showing that ensemble perception can be formed even when the representation of individual objects at an equivalent precision is not available due to crowding, or brief exposure (Alvarez, 2011; Haberman & Whitney, 2009; Parkes et al., 2001). These findings speak to the efficiency of ensemble perception. However, as these studies primarily focused on comparing the precision of ensemble perception and individual object perception, it remains unknown how these two perceptions differ in terms of their capacities. A direct comparison of their capacities would allow testing of the efficacy of ensemble perception.

Measuring capacity limit

To determine the capacity limit of ensemble perception, we used a pre-cue and post-cue paradigm, which has been widely used in visual perception and working memory studies. In a typical pre-cue condition, the visual cue is presented prior to the onset of the stimulus array, such that attention can be directed to task-relevant information in the array (Griffin & Nobre, 2003; Woodman, Vecera, & Luck, 2003). In a post-cue condition, the cue is presented only after the offset of the stimulus array, such that attention needs to be distributed among all items in the array during encoding (Griffin & Nobre, 2003; Hollingworth & Maxcey-Richard, 2013). Halberda et al. (2006) and Poltoranski and Xu (2013) used the pre-cue and post-cue paradigm to examine the capacity of multiset enumeration. Following these studies, in our experiments, the capacity limit is measured as the minimum number of sets of items where task performance is reliably different between the pre-cue and the post-cue conditions. The major advantage of measuring capacity limit by comparing pre-cue and post-cue performance is that it maintains the same perceptual load between conditions and allows manipulation of attention.

The current study

The goal of the current study is to investigate the capacity limit of ensemble perception. The first three experiments characterized the capacity of ensemble perception by exploring whether there was a fixed upper limit. Experiment 1 examined the maximum number of sets over which the visual system could summarize. To examine the robustness of the capacity limit, Experiment 2 deliberately biased participants' attention to one circle in each set, while asking them to compute the set means. In Experiment 3, we manipulated exposure durations of the stimulus arrays to examine whether the capacity limit found in Experiment 1 was due to information processing speed. Experiment 4 compared the capacity of ensemble and individual object perception by eliminating any requirement of mean computation.

General method

Participants

Participants in all experiments were undergraduate students at the University of British Columbia (UBC), and were recruited through the Human Subjects Pool in the Department of Psychology. All participants had normal or corrected-to-normal vision, provided informed consent, and received

course credits as compensation. All experiments were approved by the UBC Behavioral Research Ethics Board.

Statistical power

The required sample size was calculated a priori assuming a Type I error level of $\alpha = 0.5$ and a medium effect size ($\eta^2 = 0.1$) for a two-way repeated-measures ANOVA. Based on the power analysis, our sample size should be at least 19. Thus, 22 participants were recruited for each of the four experiments.

Apparatus

In all experiments, participants were seated 60 cm from a computer monitor (13.3 in.; 1,280 pixels \times 900 pixels; refresh rate = 60 Hz). Stimuli were presented using MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox (<http://psyctoolbox.org>).

Stimuli

Each stimulus array consisted of one to eight sets of colored circles presented against a gray background (see Fig. 1). Hereafter in our study, we define "set size" as the number of sets in a stimulus array, which should be distinguished from the magnitude of a set (i.e., how many items a set contains), and from the total or mean area covered by circles in a set.

Each set contained four circles in the same color (in Experiments 1–3), or only one circle in one color (in Experiment 4). The color was unique for each set and randomly selected from a pool of eight colors (red, green, blue, yellow, cyan, magenta, orange, and black). The circles were randomly assigned to cells in an invisible 5×7 grid (subtending $16.04^\circ \times 22.46^\circ$ in visual angle) on the screen. Within each cell, the circle was placed at the center with random jittering of up to 0.17° .

To determine the circle diameter for experiments with four circles per set (Experiments 1–3), the grand mean (i.e., the mean of all circles in the array) was first determined for trials where the set size S was an even number ($S = 2M$, $M = 1, 2, 3, 4$). The individual set means were evenly spaced around the grand mean by steps of 0.21° . For trials where the set size S was an odd number ($S = 2M - 1$, $M = 1, 2, 3, 4$), the grand mean of $2M$ sets was first determined, and the individual set means were evenly spaced around the grand mean by steps of 0.21° . We removed one of the two set means closest to the grand mean to obtain $2M - 1$ sets. Within each set, the diameters of individual circles were evenly spaced around the set mean by steps of 0.21° . The range of circle diameters was between 0.31° and 2.90° . In Experiment 4 with only one circle per set, the set means themselves were used as individual circle sizes.

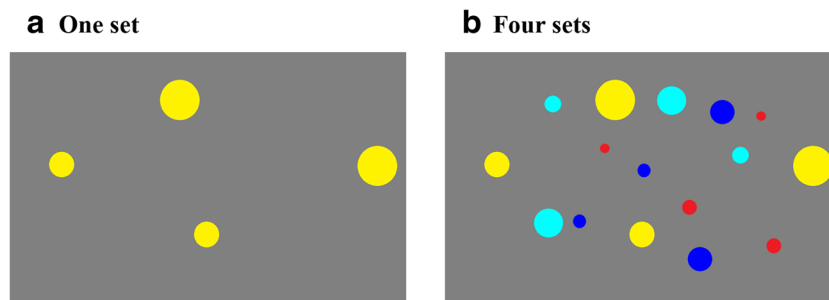


Fig. 1 Examples of stimulus array for the ensemble task. **a** One set of stimuli. **b** Four sets of stimuli. In Experiments 1–3, each set contained four circles as the figures show here. In Experiment 4, each set contained only one circle. (Color figure online)

All four experiments used a direct report task, in which the participants reported a set mean or a circle size by adjusting the size of a probe circle with the mouse. The probe circle was placed at the screen with a starting diameter of 0.21° , and its color indicated the set or the circle being probed. As this starting diameter was always smaller than all possible circle sizes (0.31° to 2.90°) in the encoding array, participants would not be able to make a categorical judgment about the target (e.g., smaller than the probe). The size of the probe circle was bounded between 0.04° and 3.52° . No feedback was provided. Participants were given unlimited time to make a response in each trial, though they were instructed to respond as quickly and accurately as possible.

Data analysis

Trials with response times greater than three standard deviations from the participant's mean in each condition were removed. We used the root mean square error (RMSE; in square visual degrees) as the measure of error, which is computed for each experimental condition c as:

$$RMSE = \sqrt{\frac{\sum_i^{n_c} \theta_i^2}{n_c}}, \quad (1)$$

where $\theta_i = u_i - R_i$ is the signed error in trial i , with u_i being the participant's estimate of set mean, and R_i the true set mean probed in that trial. n_c is the number of trials in a given cue and set size condition c . Alternative RMSE computations were used in Experiment 2.

In all experiments, we inferred the capacity limit based on the difference of RMSE between the pre-cue and post-cue conditions (Halberda et al., 2006).

Experiment 1

In Experiment 1, we sought to determine the capacity limit of ensemble perception using the pre-cue/post-cue paradigm. We hypothesized that mean size perception might have a two-set or three-set capacity constraint, similar to the capacity of other

summary statistics such as multiset enumeration (Halberda et al., 2006; Poltoratski & Xu, 2013).

Participants and procedure

Twenty-two undergraduate students (14 females, $M_{\text{age}} = 20.1$ years, $SD = 1.9$ years) participated in the experiment. Experimental procedures, stimuli, and apparatus followed the general method. There were two within-subject conditions: pre-cue and post-cue (see Fig. 2). In each trial, participants first fixed a central cross for 1,000 ms. In the pre-cue condition, the fixation cross was filled with a stimulus color, signaling that participants would be asked to report the mean size of the set in that particular color. In the post-cue condition, the cross was hollow with a white outline, so participants did not know which set would be later probed. After fixation, an array with one to eight sets of circles were presented for 1,000 ms, followed by a white-noise mask for 500 ms, and then a probe circle in a stimulus color. The intertrial interval was 500 ms. The experiment began with 10 practice trials, followed by a total of 320 experimental trials. There were 20 trials in each Cue \times Set Size condition.

Results and discussion

The RMSE is shown in Fig. 3. A 2 (cue condition: pre-cue vs. post-cue) \times 8 (set size: one to eight) repeated-measures ANOVA revealed a significant main effect of cue condition, $F(1, 21) = 48.40$, $p < .001$, $\eta_p^2 = 0.70$, and RMSE was lower in the pre-cue condition, $t(175) = 10.32$, $p < .001$, $d = 0.78$, 95% CI [0.45, 0.66]. A main effect of set size, $F(7, 147) = 6.38$, $p < .001$, $\eta_p^2 = 0.23$, and an interaction effect, $F(7, 147) = 5.12$, $p < .001$, $\eta_p^2 = 0.20$, were also observed. A one-way ANOVA revealed significant effect of set size on post-cue RMSE, $F(1, 147) = 7.94$, $p < .001$, $\eta_p^2 = 0.27$, but not on pre-cue RMSE, $F(7, 147) = 1.17$, $p = .322$, $\eta_p^2 = 0.05$. Paired-sample t tests showed that post-cue RMSE was significantly greater than the pre-cue RMSE at set sizes ≥ 3 , $t(21) = 0.85$, $p = .405$, $d = 0.18$ at set size one; $t(21) = 1.76$, $p = .092$, d

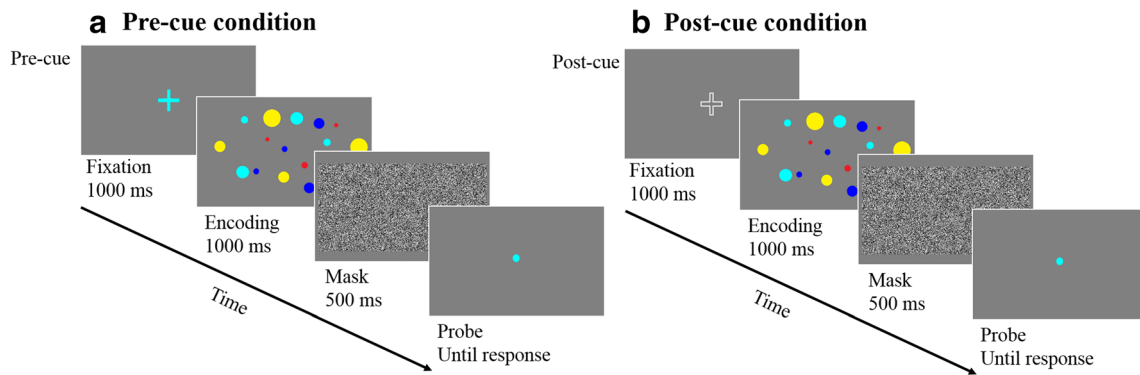


Fig. 2 Experiment 1 procedure. Example trial in the pre-cue condition (a) and in the post-cue condition (b). In each trial, the array contained one to eight sets of circles, each set containing four circles in the same color. In the pre-cue example trial, the cyan fixation cross indicated the mean size

of the four cyan circles would be probed later; in the post-cue example trial, the fixation cross was hollow and the set to be probed was not known in advance. (Color figure online)

= 0.38 at set size two; and $ts(21) \geq 3.70$, $ps \leq .001$, $ds \geq 0.79$ for set sizes three to eight. Thus, beyond set size two, mean size estimation was less accurate in the post-cue condition than in the pre-cue condition. The two-set capacity limit we observed here was consistent with the two-set to three-set capacity for enumeration (Halberda et al., 2006; Poltoratski & Xu, 2013) and visual working memory of mean size of circles (Im & Chong, 2014).

Experiment 2

How robust would the two-set capacity limit observed in Experiment 1 be if participants were subject to bottom-up attention biases? In this experiment, we deliberately biased participants’ attention to one circle in each set while instructing them to compute the set means. Despite previous studies showing attention bias could undermine precision of ensemble perception (Chong & Treisman, 2005b; de Fockert & Marchant, 2008), we hypothesized that attention bias might not affect its capacity, since capacity and precision could be two separable aspects of a cognitive function (Ma, Husain, & Bays, 2014).

Participants and procedure

Twenty-two new participants (16 females, $M_{age} = 19.7$ years, $SD = 1.5$ years) from UBC took part in the experiment. The stimuli and procedure (see Fig. 4) were identical to those in Experiment 1, except for one critical difference. In each trial, the array appeared on the screen for 300 ms after fixation, and one randomly selected circle in each set briefly disappeared for 200 ms. Afterwards, all circles would be presented for 700 ms. The brief flash of the circles intended to bias attention to one circle in each set and was completely task irrelevant.

As in Experiment 1, we computed the RMSE based on the set mean ($RMSE_S$) and the RMSE based on the flashed circle in the probed set ($RMSE_F$). The former was calculated with Eq. 1, and the latter with:

$$RMSE_F = \sqrt{\frac{\sum_{i=1}^{n_c} \phi_i^2}{n_c}}, \tag{2}$$

where $\phi_i = u_i - R'_i$ is the difference between actual response (u_i) and size of the flashed circle in the probed set (R'_i) for trial i . Variable n_c is the number of trials in a given Cue \times Set Size condition c . Comparison between $RMSE_S$ and $RMSE_F$ would indicate whether participants based their estimation on the true set mean or on the flashed circle.

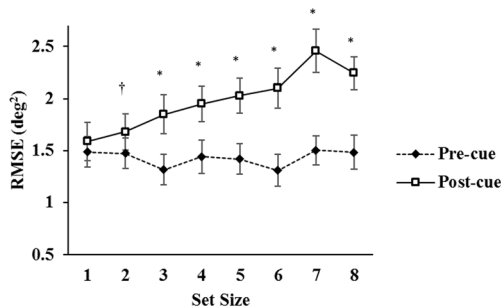


Fig. 3 Experiment 1 results. Estimation accuracy measured in RMSE (in square visual degrees) at each set size in the pre-cue and the post-cue conditions. (* $p < .05$, † $p < .10$, error bars reflect ± 1 SEM)

Results and discussion

The RMSE’s are shown in Fig. 5. A two-way repeated-measures ANOVA on $RMSE_S$ revealed a significant main effect of cue condition, $F(1, 21) = 36.31$, $p < .001$, $\eta_p^2 = 0.63$, with post-cue $RMSE_S$ greater than pre-cue $RMSE_S$, $t(175) = 9.53$, $p < .001$, $d = 0.72$, 95% CI [0.38, 0.58]. A main effect of set size was also observed, $F(7, 147) = 4.26$, $p < .001$, $\eta_p^2 = 0.17$. There was a reliable interaction effect between cue and set size, $F(7, 147) = 7.05$, $p < .001$, $\eta_p^2 = 0.25$, which could be

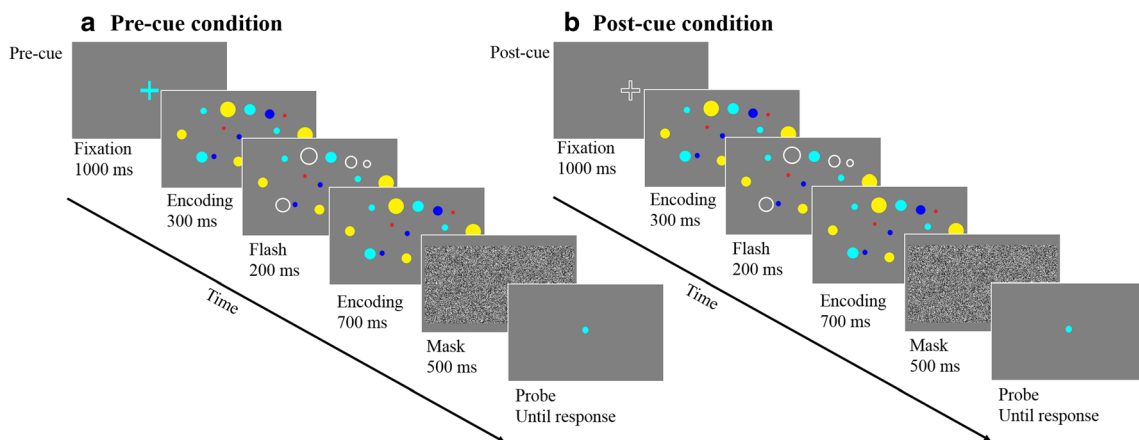


Fig. 4 Experiment 2 procedure. An example trial in the pre-cue condition (a) and in the post-cue condition (b). During the flash display in both conditions, one circle from each set disappeared for 200 ms (white circles

indicate the positions of the circles that disappeared and are not shown in the array). (Color figure online)

explained by a significant effect of set size on post-cue $RMSE_S$, $F(7, 147) = 8.21, p < .001, \eta_p^2 = 0.28$, but not on pre-cue $RMSE_S$, $F(7, 147) = 1.14, p = .344, \eta_p^2 = 0.05$. Paired-sample t tests indicated significant difference between the pre-cue and post-cue $RMSE_S$ at set sizes ≥ 3 , $t(21) = -1.21, p = .241, d = 0.26$ for set size one; $t(21) = 1.20, p = .243, d = 0.26$ for set size two; and $t(21) \geq 3.95, ps < .001, ds \geq 0.84$ for set sizes three to eight. This showed a deterioration of estimation accuracy at set sizes ≥ 3 . The two-set capacity limit here was consistent with Experiment 1 and demonstrated the robustness of ensemble capacity. To examine whether mean size estimation was influenced by flashed circles in the array, we compared $RMSE_S$ and $RMSE_F$ across conditions and set sizes using a three-way ANOVA with a 2 ($RMSE$ condition: $RMSE_S$ and $RMSE_F$) \times 2 (cue condition) \times 8 (set size) design. There was a main effect of $RMSE$ condition, $F(1, 21) = 52.53, p < .001, \eta_p^2 = 0.71$, with $RMSE_F$ being significantly higher than $RMSE_S$, $t(351) = 15.84, p < .001, d = 0.84, 95\% CI [0.17, 0.22]$. This was true in both the pre-cue condition, $F(1, 21) = 47.67, p < .001, \eta_p^2 = .69$, and the post-cue condition, $F(1, 21) = 41.40, p < .001, \eta_p^2 = .66$. Estimation was more likely to be

based on the mean of all circles in the set rather than the flashed circle.

Nonetheless, were the participants still biased toward the flashed circle? A partial correlation test was run on each participant between their estimation and the size of the flashed circle in the probed set, after controlling for the probed set mean. Among the 22 subjects, only five showed significant ($p < .05$) or marginally significant ($p < .10$) partial correlations (see Appendix 1). This observation, together with the ANOVA analysis, suggested a minimal biasing effect. Our results differed from those of de Fockert and Marchant (2008), which suggests mean size computation was biased toward the attended item. One possible explanation would be that in their study, participants were explicitly instructed to attend to either the largest or the smallest circle in the set, whereas our biasing items were randomly chosen from the set, and were presented in a bottom-up manner. Possibly, the salience of the biasing items was stronger in their paradigm than in ours, which led to the different results.

Experiment 3

In an ensemble perception task, performance is likely to be determined by two independent factors: An *encoding limit* that describes the maximum rate at which visual inputs can be sampled, integrated, and encoded (Huang & Pashler, 2007; Huang et al., 2007; Treisman & Gelade, 1980), and a *capacity limit*, which determines the maximum amount of information that can be ultimately encoded. Such a dual source of limit is associated with other high-level cognitive functions, such as working memory storage (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Experiment 3 examined the capacity limit of ensemble perception independently of its encoding limit, by varying the exposure durations of stimulus arrays. If increased exposure durations could not further improve the capacity limit,

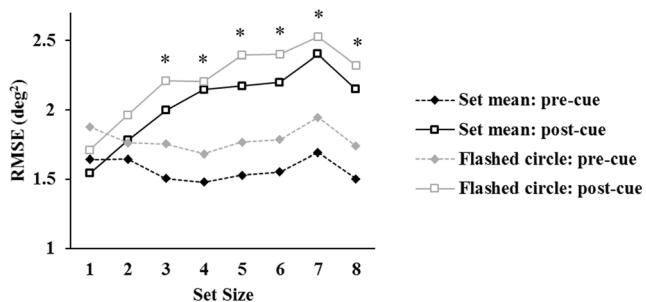


Fig. 5 Experiment 2 results. $RMSE$ based on probed set mean ($RMSE_S$) or on the size of the flashed circle in the probed set ($RMSE_F$) in the pre-cue and the post-cue conditions. (* indicates $p < .05$ between pre-cue and post-cue $RMSE_S$; error bars reflect $\pm 1 SEM$)

then the bottleneck of ensemble perception would more likely be due to an internal capacity limit than to the encoding limit.

Participants and procedure

Sixty-six new participants (42 females, $M_{age} = 20.7$ years, $SD = 2.7$ years) were randomly assigned to three exposure duration conditions (500 ms, 1,500 ms, and 2,000 ms; $N = 22$ in each condition). The procedure was identical to that in Experiment 1, except for changes in exposure duration. Data from Experiment 1 (1,000 ms) was reused for comparison.

Results and discussion

The RMSE is shown in Fig. 6a–d for the 500-ms, 1,000-ms, 1,500-ms, and 2,000-ms conditions. For all exposure conditions, post-cue RMSE was significantly higher than pre-cue RMSE at set sizes ≥ 3 , $t_s(21) \geq 2.64$, $p_s \leq .015$, $d_s \geq 0.56$ for set sizes three to eight in all four exposure conditions. For the

1,000-ms condition (Experiment 1), a marginal difference occurred at set sizes ≥ 2 , $t(21) = 1.76$, $p = .092$, $d = 0.38$. The two-set capacity limit of ensemble perception was reliably observed at different exposure durations. Importantly, extending the exposure duration beyond 1,000 ms did not increase the capacity limit. Its invariance to exposure time suggests that the capacity limit we observed in Experiment 1 (1,000-ms exposure) was not due to the encoding limit but rather stemmed from an internal resource constraint.

Experiment 4

Finally, we investigated the capacity of object perception under the same pre-cue and post-cue paradigm (the “object perception task”), and compared it to the capacity of ensemble perception (Experiment 1, the “ensemble perception task”). Since ensemble perception allows encoding of objects that are inaccessible individually (e.g., due to crowding), hence

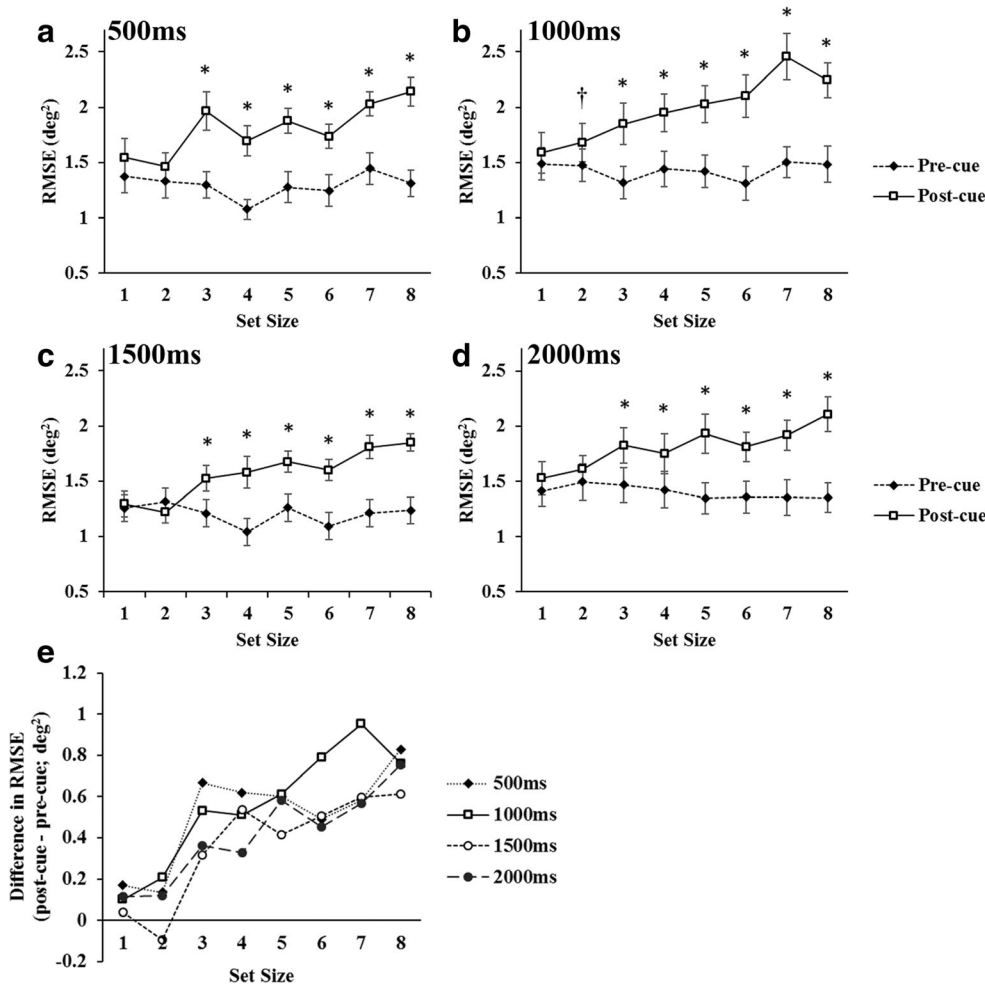


Fig. 6 RMSE in Experiment 3 with different exposure durations. **a–d** RMSE with the 500-ms, 1,000-ms, 1,500-ms, and 2,000-ms exposure durations, respectively. (* $p < .05$, † $p < .10$ for difference between pre-

cue and post-cue RMSE.) **e** Difference between pre-cue and post-cue RMSE (post-cue – pre-cue) in the four exposure conditions

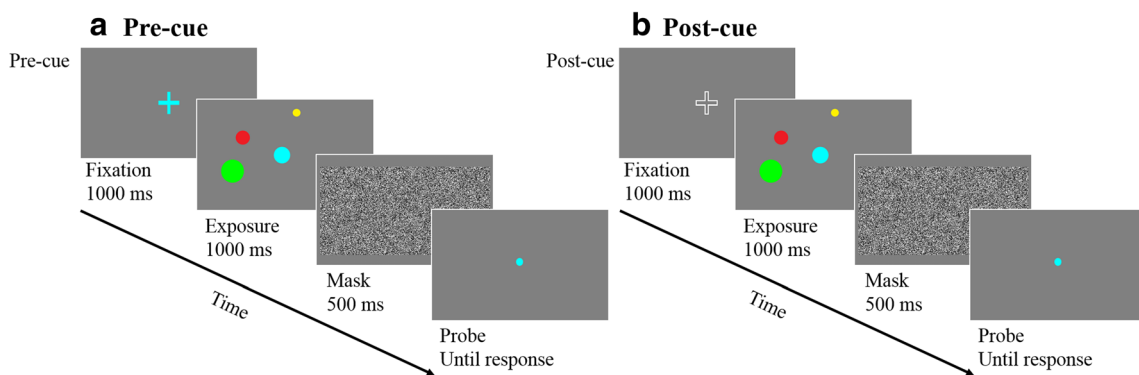


Fig. 7 Experimental procedure of Experiment 4. An example trial in the pre-cue condition (a) and in the post-cue condition (b). In each trial, the array contained one to eight sets of circles, each set containing one circle. In the pre-cue example trial, the cyan fixation cross indicated the mean

size of the cyan circle would be probed later; and in the post-cue example trial, the fixation cross was hollow and the set to be probed was not known in advance. (Color figure online)

allowing a great amount of information about objects to be encoded, we hypothesized that the capacity of individual object perception might not exceed that of ensemble perception.

Participants and procedure

Twenty-two new participants (17 females, $M_{age} = 20.0$ years, $SD = 2.1$ years) completed the experiment. The stimuli were similar to those in Experiment 1, except that each set now contained only one circle, so that the entire display contained between one and eight circles. The circles were randomly assigned to an invisible 5×7 grid. The procedure was identical to that in Experiment 1 (see Fig. 7): Each trial started with a central fixation cross for 1,000 ms, which might be filled with a specific color that indicated which circle would be probed (pre-cue condition), or hollow with a white outline and the probed circle would not be known until after the array disappeared (post-cue condition). After the cue, an array with one to eight circles, each in a unique color, appeared for 1,000 ms, and participants were instructed to remember the size of a specific circle (pre-cue condition) or every circle (post-cue condition). The array was then replaced by a 500 ms white-noise mask, followed by a probe circle in a specific color. Participants reported the size of the circle in that color in the stimulus array. A total of 320 trials were completed by each

participant. Analyses of RMSE followed the same procedure as in Experiment 1.

Results and discussion

As before, ANOVA on RMSE (see Fig. 8) showed significant main effects of set size, $F(7, 147) = 14.89, p < .001, \eta_p^2 = 0.41$, and cue condition, $F(1, 21) = 145.40, p < .001, \eta_p^2 = 0.87$, as well as an interaction, $F(7, 147) = 16.26, p < .001, \eta_p^2 = 0.44$. Paired-sample t tests showed that post-cue RMSE was reliably greater than pre-cue RMSE at set sizes $\geq 2, t(21) = 0.34, p = .740, d = 0.07$ for set size one; $t(21) \geq 3.00, ps \leq .007, ds \geq 0.64$ for set sizes ≥ 2 .

To compare the performance in the ensemble perception task (Experiment 1) and the object perception task (Experiment 4), we computed the ratio of post-cue RMSE over pre-cue RMSE at each set size (see Fig. 8), and ran a 2 (task condition: ensemble vs. object perception) \times 8 (set size: one to eight) mixed ANOVA. Task condition significantly influenced performance, $F(1, 42) = 5.26, p = .027, \eta_p^2 = 0.11$, which was due to a greater overall RMSE ratio in the object perception task than in the ensemble perception task, $t(338.89) = 2.77, p = .006, d = 0.30, 95\% \text{ CI } [0.06, 0.34]$. Set size had a main effect on performance, $F(7, 294) = 14.05, p < .001, \eta_p^2 = 0.25$. The

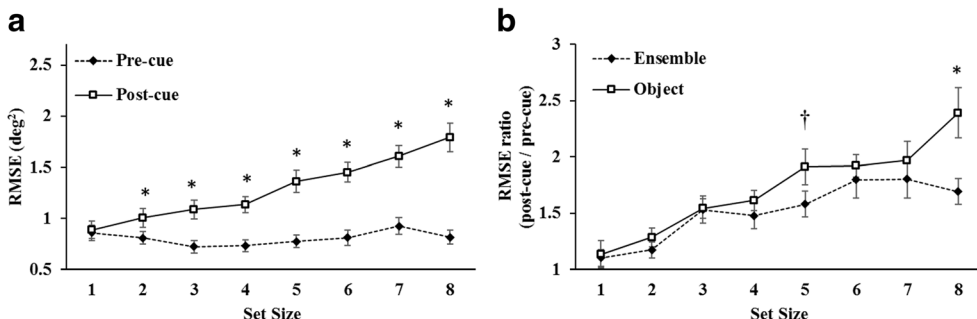


Fig. 8 Results of Experiment 4. **a** RMSE in the pre-cue and the post-cue conditions. **b** RMSE ratio in the ensemble perception task (Experiment 1, dashed line) and in the object perception task (Experiment 4, solid line). (* $p < .05$, † $p = .103$, error bars reflect $\pm 1 \text{ SEM}$)

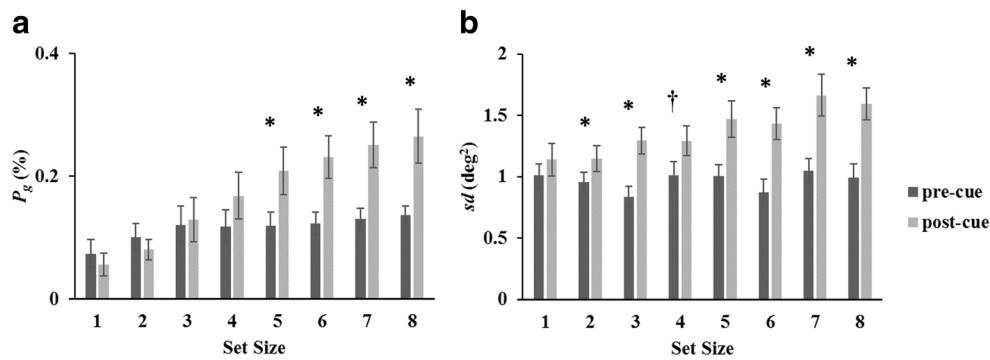


Fig. 9 **a** Guess probability (P_g) and **(b)** standard deviation (SD) at each set size-by-cue condition in Experiment 1. Pre-cue and post-cue P_g differed significantly at set sizes ≥ 5 , showing a four-set capacity for ensemble perception. (* $p < .05$, † $p < .10$, error bars reflect $\pm 1 SEM$)

interaction effect was nonsignificant, $F(7, 294) = 1.65, p = .121, \eta_p^2 = 0.04$. Independent-samples t tests revealed a significant difference between ensemble and object RMSE ratios at set size eight, $t(31.51) = 2.77, p = .009, d = 0.84$, marginally at set size five, $t(38.40) = 1.67, p = .103, d = 0.50$, and nonsignificant at the other set sizes ($ps \geq 0.325, ds \leq 0.30$). One interpretation of these results is that, compared with individual object perception, ensemble perception of a multiset display achieved a lower noise level. Taken together, ensemble perception had a capacity at least as large as that of individual object perception and conferred a noise-reduction mechanism for encoding large stimulus sets.

General discussion

The goal of the current study was to examine the capacity limit of ensemble perception of multiple spatially intermixed sets. Experiment 1 showed that observers could accurately estimate the mean size of circles from maximally two sets. Experiment 2 suggested that their performance was robust against attention bias to individual objects in a set. With varying exposure durations, Experiment 3 showed that the two-set capacity limit was possibly due to internal resource limits, rather than an encoding speed limit. Experiment 4 showed that under our paradigm, the capacity of ensemble perception was at least as large as that of individual object perception.

Different capacity limits

Our finding of a two-set capacity for ensemble perception was in line with the two-set or three-set capacity for multiset enumeration (Halberda et al., 2006; Poltoratski & Xu, 2013) and visual working memory of mean size of circles (Im & Chong, 2014). However, this capacity limit was apparently lower than the three-object or four-object limits for visual working memory (Luck & Vogel, 1997; Zhang & Luck, 2008), and multiple target tracking (Pylyshyn & Storm, 1988; review: Cavanagh & Alvarez, 2005). One possible explanation is due to the different paradigms for measuring capacity in those studies and ours.

To infer the capacity for working memory of objects, Zhang and Luck (2008) assumed the response error (difference between the reported color and the target color) follows a uniform-Gaussian mixture model. That is, if the target value has not been encoded, participants can only make a random guess when it is probed. Thus, their response would follow a uniform distribution, bounded by the maximally and minimally possible response values. In contrast, if the target value has been encoded, participants may then report the target value with Gaussian noise. Using this method, they were able to separate capacity of working memory (in terms of how many objects can be held) and noise in representation.

In additional analyses (see Appendix 2), we fit a uniform-Gaussian model to the response error (reported size – target

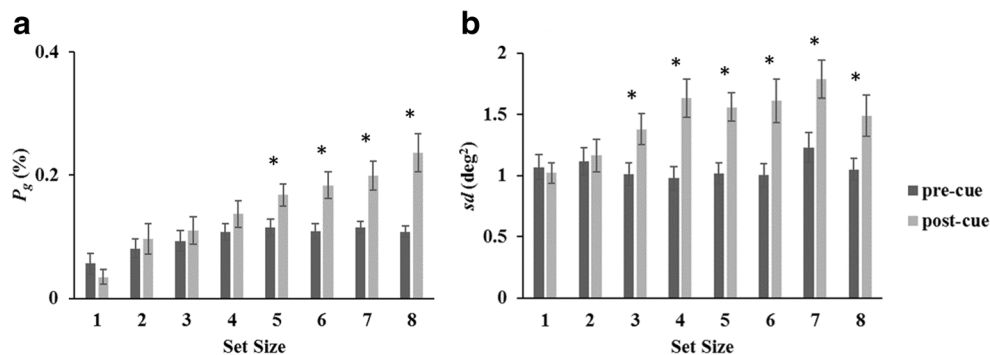


Fig. 10 **a** Guess probability (P_g) and **(b)** standard deviation (SD) at each condition in Experiment 2. Confirming findings in Experiment 1, capacity limit for ensemble perception occurred at set size four. (* $p < .05$, † $p < .10$, error bars reflect $\pm 1 SEM$)

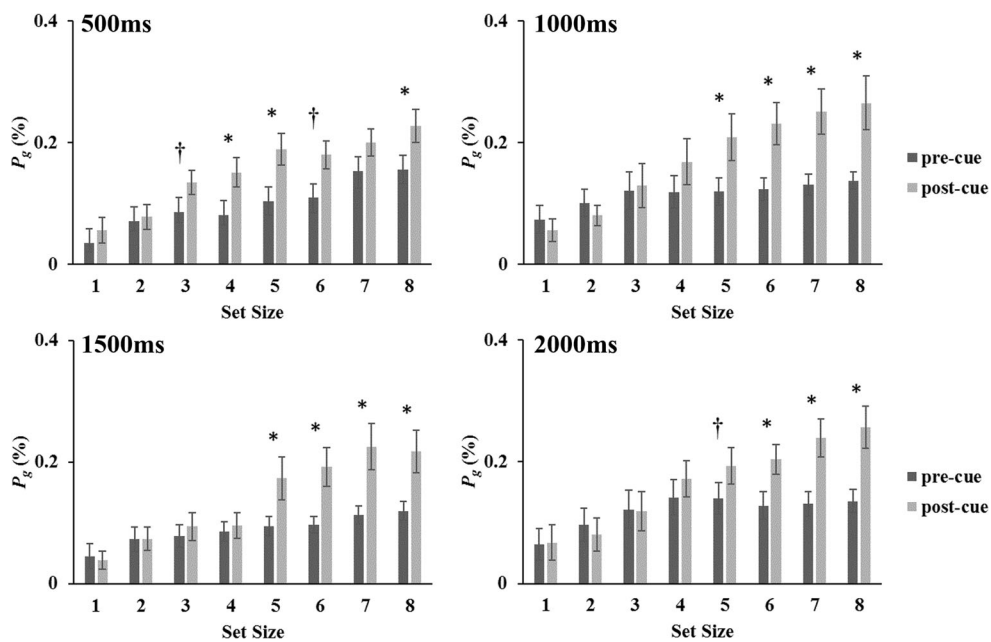


Fig. 11 Guess probability (P_g) at each exposure condition. Capacity limit was reliably four sets with an exposure duration $\geq 1,000$ ms. (* $p < .05$, † $p < .10$, error bars reflect ± 1 SEM)

size) in the ensemble perception tasks (Experiments 1–3) or the object perception task (Experiment 4), for the pre-cue and post-cue conditions independently. We estimated the probability of random guess (probability of the uniform distribution), and the precision of encoded set means (standard deviation of the gaussian distribution). The capacity limit was defined as the maximum number of sets or objects beyond which pre-cue and post-cue guess probabilities would significantly differ. Remarkably, we found a four-set capacity limit for ensemble perception in all Experiments 1, 2, and 3, and a four-object capacity limit for

Experiment 4. This suggests convergence of capacity limits between ensemble perception and a variety of other cognitive functions, such as working memory (Zhang & Luck, 2008).

However, it should be noted that certain assumptions of the uniform-Gaussian model, as outlined in Zhang and Luck (2008), might be violated under our paradigm. Zhang and Luck (2008) had a circular space of response (a 360° color wheel). Each allowable response value is thus equally likely to be chosen if the right answer is not encoded. However, the allowable response in our experiments lies in a linear,

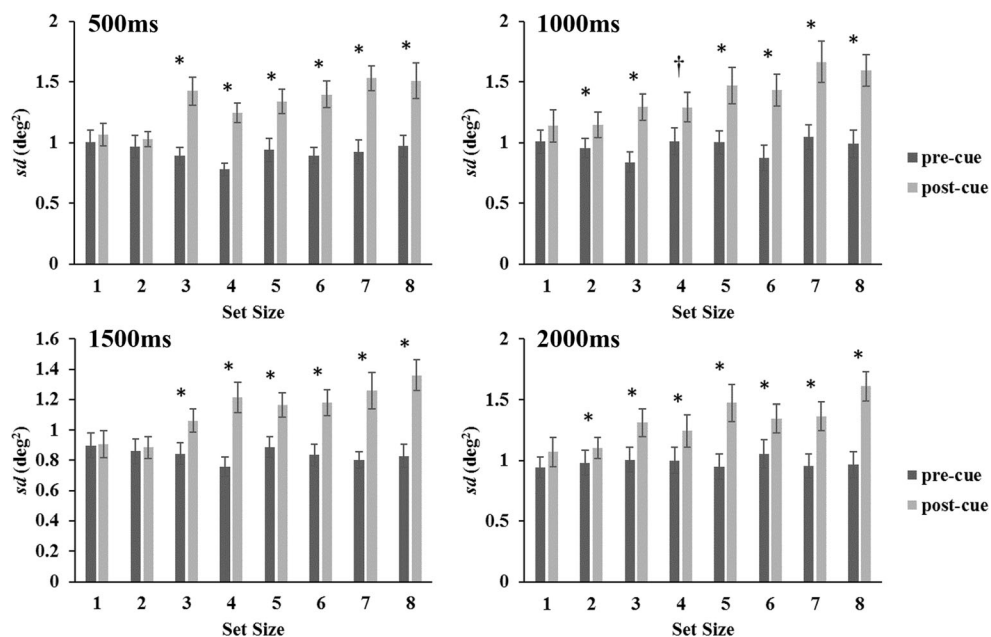


Fig. 12 Standard deviation (SD) at each exposure duration of Experiment 3. (* $p < .05$, † $p < .10$, error bars reflect ± 1 SEM)

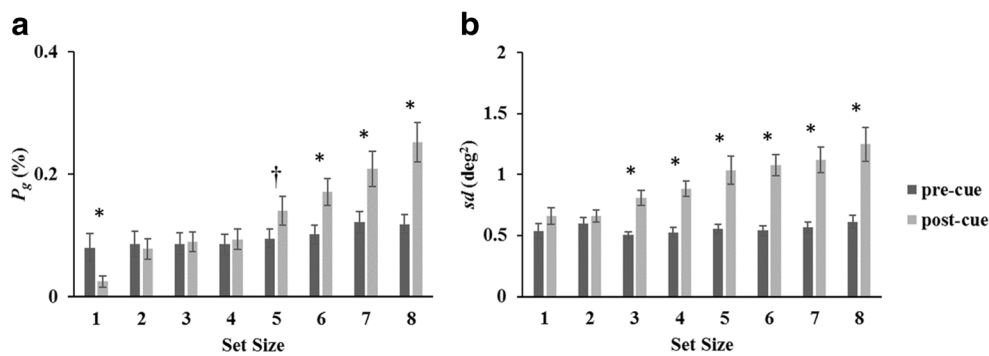


Fig. 13 Guess probability (P_g) and standard deviation (SD) in Experiment 5. Capacity limit for object perception occurred at set size four. ($*p < .05$, $\dagger p < .10$, error bars reflect $\pm 1 SEM$)

bounded space; depending on what value the observers would choose when they missed the right answer, the probability of an allowable response value being chosen might or might not be uniformly distributed.

RMSE, on the other hand, gives us a picture of the capacity limit in terms of how many ensembles (or objects) can be formed, as well as the precision in representations in an integrated manner. It thus seems to be a valid and informative measure to our purpose. Future studies may seek to detail the relationship between the number and precision of encoded ensembles (Im & Chong, 2014; Solomon, 2010).

Robustness of capacity against attention bias

In Experiment 2, we observed biasing participants' attention to one circle in a set did not significantly skew their estimation of the set mean. This is seemingly at odds with previous findings of mean size estimation biased toward the attended individual set member (Chong & Treisman, 2005b; de Fockert & Marchant, 2008), which might predict a greater noise level and hence a lower capacity in Experiment 2. One possible explanation would be the inevitability and automaticity of ensemble perception, despite bottom-up attention bias to individual objects. Its automaticity is evidenced by studies in which summary statistics were not required, but were still computed and biased task performance (Brady & Alvarez, 2011; Parkes et al., 2001). On the other hand, while previous studies often directed attention to a single item (Chong & Treisman, 2005b; de Fockert & Marchant, 2008), our Experiment 2 biased attention to one circle in each set, dispersed across the whole display. This might have in fact encouraged global attention to the whole stimulus array, which is shown to facilitate ensemble perception (Chong & Treisman, 2005b).

Relationship between working memory of features and ensemble perception

Poltoratski and Xu (2013) showed that visual working memory of the feature that defines the set (color in their

case) was a limiting factor for the enumeration task in Halberda et al. (2006). In their experiments, participants were asked to enumerate spatially intermixed sets, and to remember all color types in the sets, either as a separate task or simultaneously with the enumeration task. They found that maximally 2.5 colors could be remembered, and two sets could be enumerated. Since observers must remember the colors before enumerating the sets associated with the colors, the authors argued that working memory capacity for multiple colors constrained the number of sets that could be enumerated. The two-set capacity of ensemble perception in our study is compatible with the 2.5-color capacity in Poltoratski and Xu (2013). Working memory for the color types presented in the array might also be required, hence constraining how many mean sizes the participants could encode and report.

Efficiency of ensemble perception

Under our paradigm, ensemble perception has a greater capacity than individual object perception. Moreover, the ratio of RMSE (post-cue RMSE divided by pre-cue RMSE at each set size) in the ensemble perception task was in fact lower than in the object perception task overall.

One implication of these findings is that, although ensemble perception involves greater noise with respect to the target per se within the limit of its capacity (compared with object perception), it provides a mechanism for encoding and storing information about more objects (a greater capacity), and scales better for perception of large stimulus sets beyond the capacity limit (a lower post-to-pre RMSE ratio). This finding is in line with previous studies showing that when individual objects are not accessible for precise encoding due to crowding, diversion of attention, or brief exposure, ensemble representations can still be encoded and reported with a better precision than individual objects (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2005a).

Another way to characterize the relationship between ensemble and individual object perception is to see

whether they share the same pool of cognitive resources, which can be done by including both ensemble stimuli and single-circle stimuli in a trial. This would allow us to examine whether ensemble representations are indistinguishable from individual object representations.

Limitations and future directions

One limitation of the current study is that, the comparisons between exposure durations in Experiment 3 were performed using a between-subjects design; similarly, comparisons between individual and ensemble representations in Experiment 4 were also between subjects. Although using within-subjects designs may lend these experiments additional power, we derived the capacity limits from within-subject difference between the pre-cue and post-cue conditions in each experiment, which minimized the random between-subjects effects. Therefore, our current between-subjects comparisons could still offer valid results.

It should also be noted that the perceived mean size may or may not be the arithmetic linear mean, but rather on a power scale, and the exponent for the power function varies for different types of features (Chong & Treisman, 2003; Da Silva, 1985; Teghtsoonian, 1965). Readers should thus be careful extending our results to other feature dimensions.

Finally, we also note that using a mask comprising of random colored dots would be a better option than a grayscale mask with white noise, which could have interfered with the estimation of capacity limit.

Future studies can examine whether the capacity limit of ensemble perception observed in the current study is generalizable to other feature domains, using other probing methods (e.g., discrimination tasks). Since ensemble perception could occur in a hierarchical manner (i.e., different levels of summaries are extracted in parallel; Emmanouil & Treisman, 2008; Haberman, Brady, & Alvarez, 2015), and spatial organization of items interferes with perceptual efficiency and memory storage (Alvarez & Cavanagh, 2005; Im & Chong, 2014), future studies can seek to illustrate how these factors would expand or constrain the capacity limit for ensemble perception.

Conclusion

The current study contributes to the literature on ensemble perception in four important ways. First, we extended beyond past studies that primarily focused on single-ensemble perception and examined the limits of perceiving multiple ensembles. Second, we demonstrated a two-set capacity limit of ensemble perception of multiple spatially intermixed sets, and showed its robustness against bottom-up attention bias to an individual object in a set. Third, we showed that

ensemble perception was limited by an internal resource constraint, rather than the speed of encoding. Finally, comparing ensemble and individual object perception suggests ensemble perception might provide a greater cognitive capacity and a noise-reduction mechanism for large amounts of inputs. Taken together, our results demonstrated that ensemble perception is a resource-limited cognitive process that enables efficient processing of visual information. Understanding properties of ensemble perception helps reveal how the visual system extracts and abstracts goal-related information from a complex environment.

Acknowledgements We thank Ru Qi Yu and Yu Luo for assistance in data collection. For helpful conversations, we thank Jim Enns, Lisa Feigenson, and the Zhao Lab. This work was supported by NSERC Discovery Grant (RGPIN-2014-05617 to J.Z.), the Canada Research Chairs program (to J.Z.), and the Leaders Opportunity Fund from the Canadian Foundation for Innovation (F14-05370 to J.Z.).

Appendix 1 Supplemental statistics for Experiment 2

Table 1 Partial correlation in Experiment 2. Second column shows the partial correlation of estimation and size of the flashed circle in the probed set, after controlling for size of the probed set mean. (* $p < .05$, † $p < .10$)

Subject	Partial correlation	p
1	−0.03	.614
2	0.05	.358
3	0.30	<.001*
4	−0.02	.679
5	0.00	.950
6	0.04	.449
7	−0.09	.127
8	0.01	.929
9	0.04	.526
10	0.10	.069†
11	−0.03	.647
12	−0.01	.925
13	−0.02	.766
14	−0.03	.604
15	0.03	.592
16	0.02	.711
17	0.12	.034*
18	0.15	.008*
19	−0.07	.208
20	−0.07	.213
21	0.46	<.001*
22	0.01	.920

Appendix 2 Mixture modeling results

Here we adopted the uniform-Gaussian mixture modeling in Zhang and Luck (2008) to the analysis of capacity limit for ensemble perception in four experiments.

Method

We assumed the signed error in each trial i follows a uniform-normal mixture distribution:

$$p(\theta_i) = P_g * \mathcal{U}(a, b) + (1 - P_g) * \mathcal{N}(\mu, \sigma^2). \quad (3)$$

In this model, P_g is the probability of guessing at random. A random guess follows a uniform distribution $\mathcal{U}(a, b)$ with a lower-bound a and an upper-bound b . Since values of a and b differ from trial to trial depending on the true set mean, they are heuristically estimated using the smallest and the largest signed errors respectively, for each participant. With a probability of $(1 - P_g)$, the probed set mean can be successfully reported with noise, captured by a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$. Since the expected error in this case is zero, so μ is fixed at zero. Precision of response is indicated by $SD = \sigma$. Therefore, for each participant and each experimental condition c , there are two parameters to be estimated (P_g^c and sd^c) by minimizing the negative log likelihood:

$$\operatorname{argmin}_{P_g^c, sd^c} \sum_{i=1}^{n_c} -\log p(\theta_i), s.t. P_g^c \geq 0, sd^c \geq 0, \quad (4)$$

where n_c is the number of trials in experimental condition c , and $p(\theta_i)$ is defined by Eq. 2. Optimization of Eq. 3 is done using the Nelder–Mead algorithm (Nelder & Mead, 1965) over a range of different initial values in search for the global minimum within a reasonable value range for each parameter. To empirically evaluate the goodness of fit for each model, we resampled n_c data points from the fitted mixture distribution, refitted the mode, and computed the minimized negative log likelihood (hereafter termed “simulated negative log likelihood”). This procedure was repeated 200 times, and the p value indicating the goodness of fit was computed as the proportion of the simulated negative log likelihoods that were greater than the negative log likelihood for the actual data (Kingdom & Prins, 2010). A p value of less than .05 indicates that the actual data significantly deviate from the fitted mixture distribution.

Best-fitting parameters for the mixture models offer an alternative metric for identifying the capacity limit. Zhang and Luck (2008) measured the probability that an item is remembered (P_m), and defined the capacity of working memory as the maximum number of objects S^* beyond which P_m would drop significantly. Analogous to this, in our experiments, we defined the capacity limit of ensemble perception as the

maximum set-size S^* beyond which the guess probability P_g in the pre-cue and post-cue conditions would be reliably different under a paired-samples t test (two-tailed).

Results

Experiment 1

We estimated the guess probability P_g and the standard deviation SD for each participant in each Cue \times Set Size condition, using the signed error. Figure 9a–b show the average P_g and SD , respectively. The uniform-normal mixture model provided good fit to the data ($.270 \leq ps \leq .960$). A 2 (cue condition) \times 8 (set size) repeated-measures ANOVA was run on P_g . A main effect of cue condition was observed, $F(1, 21) = 5.24, p = .033, \eta_p^2 = .20$, which was due to a higher P_g in the post-cue condition, $t(175) = 4.50, p < .001, d = 0.34, 95\% \text{ CI } [0.03, 0.08]$. A main effect of set size, $F(7, 147) = 8.42, p < .001, \eta_p^2 = .29$, and an interaction, $F(7, 147) = 5.63, p < .001, \eta_p^2 = .21$, were also observed. Paired-sample t tests revealed significantly higher post-cue P_g at set sizes $\geq 5, ts(21) \leq 1.44, ps \geq .165, ds \leq 0.31$, for set sizes up to four; $ts(21) \geq 2.18, ps \leq .041, ds \geq 0.46$, for set sizes five to eight. Hence, there was a higher probability of random guess in the post-cue condition, when the stimulus array contained more than four sets.

For the estimation precision SD , a 2 (cue condition) \times 8 (set size) repeated-measures ANOVA revealed a main effect of cue condition, $F(1, 21) = 46.62, p < .001, \eta_p^2 = .69$, which was driven by a higher post-cue $SD, t(175) = 9.51, p < .001, d = 0.72, 95\% \text{ CI } [0.33, 0.50]$. A main effect of set size, $F(7, 147) = 3.76, p < .001, \eta_p^2 = .15$, and an interaction, $F(7, 147) = 2.94, p = .007, \eta_p^2 = .12$, were observed. There were significant differences in SD between the two cue conditions at most set sizes, $ts(21) \geq 2.37, ps \leq .027, ds \geq 0.51$, except set size one, $t(21) = 1.38, p = .181, d = 0.29$, and marginally at set size four, $t(21) = 1.9, p = .066, d = 0.41$. The analyses of P_g and SD suggest that within the four-set limit, the reduced accuracy at set sizes two, three, and four measured by RMSE (see Experiment 1) was driven by a lower encoding precision, rather than by a higher guess probability.

Experiment 2

Experiment 2 deliberately biased participants’ attention to an individual circle in each set by briefly flashing the circle. It aimed to examine how robust the capacity limit of ensemble perception would be, against attention bias.

As in Experiment 1, we modeled the set mean-based signed error with the uniform-normal mixture distribution (goodness-of-fit p values were between 0.230 and 0.940), and examined the guess probability P_g (Fig. 10a) and precision SD (Fig. 10b). For P_g , a repeated-measures ANOVA revealed a main effect of cue condition, $F(1, 21) = 5.45, p = .030, \eta_p^2 = .21$, with the

pre-cue P_g being significantly lower than the post-cue P_g , $t(175) = 4.69$, $p < .001$, $d = 0.35$, 95% CI [0.03, 0.07]. A main effect of set size, $F(7, 147) = 17.93$, $p < .001$, $\eta_p^2 = .46$, and an interaction, $F(7, 147) = 5.61$, $p < .001$, $\eta_p^2 = .21$, were also observed. Paired-sample t tests showed greater post-cue P_g at set sizes ≥ 5 , $ts(21) \geq 2.35$, $ps \leq .029$, $ds \geq 0.50$, but not set sizes ≤ 4 , $ts(21) \leq 1.04$, $ps \geq .310$, $ds \leq 0.22$, suggesting a four-set limit in ensemble perception. These results were consistent with those in Experiment 1, despite the deliberate attention bias.

A two-way repeated-measures ANOVA on SD showed a main effect of cue condition, $F(1, 21) = 49.92$, $p < .001$, $\eta_p^2 = .70$, with pre-cue SD being lower than post-cue SD , $t(21) = 9.54$, $p < .001$, $d = 0.72$, 95% CI [0.31, 0.48]. There was also a main effect of set size, $F(7, 147) = 7.32$, $p < .001$, $\eta_p^2 = .26$, and an interaction, $F(7, 147) = 7.18$, $p < .001$, $\eta_p^2 = .25$. Paired-sample t tests revealed significant differences in SD between the two cue conditions at set sizes ≥ 3 , $ts(21) \leq 0.45$, $ps \geq .353$, $ds \leq 0.20$ for set sizes one and two; $ts(21) \geq 3.54$, $ps \leq .002$, $ds \geq 0.76$ for set sizes ≥ 3 . Consistent with Experiment 1, within the four-set limit the reduced accuracy at set sizes three and four seemed to be driven by lower encoding precision, rather than by a higher guess probability. Experiment 2 confirms the capacity limit in Experiment 1, and suggested ensemble capacity was robust against attention to individual objects in a set.

Experiment 3

In this experiment, we examined how varying exposure durations of the encoding array would affect the capacity limit of ensemble perception. Four exposure durations were tested: 500 ms, 1,000 ms (reusing data from Experiment 1), 1,500 ms, and 2,000 ms.

The uniform-normal model fitted well to the data at all exposure durations (500 ms: $0.285 \leq ps \leq 0.995$; 1000 ms: $0.270 \leq ps \leq 0.960$; 1,500 ms: $0.280 \leq ps \leq 0.930$; 2,000 ms: $0.175 \leq ps \leq 0.970$). A repeated-measures ANOVA was conducted on guess probabilities P_g (Fig. 11). Main effects of cue condition and set size were observed for all exposure durations, cue condition: $Fs(1, 21) \geq 5.21$, $ps \leq .033$, $\eta_p^2 \geq .20$; set size: $Fs(7, 147) \geq 8.42$, $ps < .001$, $\eta_p^2 \geq .29$.

In Experiment 1, when set sizes increased to five and beyond, P_g in post-cue trials reliably differed from that in pre-cue trials (Fig. 11, 500-ms condition). In the current experiment, when exposure duration decreased to 500 ms, post-cue P_g was at least marginally different (at $p < .100$) from pre-cue P_g at set sizes ≥ 3 , $ts(21) \geq 1.87$, $ps \leq .075$, $ds \geq 0.40$, except at set size seven, $t(21) = 1.32$, $p = .200$, $d = 0.28$. Post-cue P_g was not different from pre-cue P_g at set sizes ≤ 2 , ($ts(21) \leq 0.96$, $ps \geq .346$, $ds \leq 0.20$). Possibly, 500 ms was not sufficient to reach the maximum ensemble representation capacity. This bottleneck could be driven by either the speed of set individuation, or that of mean computation, or a combination of both.

On the other hand, when the exposure duration increased to 1,500 ms, post-cue P_g significantly differed from pre-cue P_g at set sizes ≥ 5 , $ts(21) \geq 2.17$, $ps \leq .042$, $ds \geq .46$, but not at set sizes ≤ 4 , $ts(21) \leq 0.60$, $ps \geq .556$, $ds \leq 0.13$ (see Fig. 11, 1,500-ms condition). Increasing the exposure duration to 2,000 ms did not alter this trend, except that the difference at set size five now become marginally significant, $t(21) = 1.75$, $p = .094$, $d = 0.37$; the difference was significant at set sizes six to eight, $ts(21) \geq 3.13$, $ps \leq .005$, $ds \geq 0.67$, but not at set sizes one to four, $ts(21) \leq 1.02$, $ps \geq .321$, $ds \leq 0.22$ (see Fig. 11, 2,000-ms condition). In addition, difference between pre-cue and post-cue SD was observed for set sizes three to eight at all exposure durations (see Fig. 12).

Experiment 4

The uniform-normal model provided good fit to the data ($0.170 \leq ps \leq 0.965$). Analyses of the guess probability P_g (see Fig. 13a) showed that pre-cue and post-cue P_g were significantly different at set sizes ≥ 6 , $ts(21) \geq 2.37$, $ps \leq .028$, $ds \geq 0.50$, and marginally different at set size five, $t(21) = 1.76$, $p = .094$, $d = 0.37$, but not at set sizes two to four, $ts(21) \leq 0.38$, $ps \geq .706$, $ds \leq 0.08$. Curiously, pre-cue/post-cue difference at set size eight was significant, $t(21) = -2.17$, $p = .041$, $d = 0.46$. This was, however, driven by a low post-cue P_g , which did not suggest a capacity limit was reached. Despite this, since a higher probability of random guess in the post-cue condition was still consistently observed at set sizes ≥ 5 , we concluded that individual representation had a four-item capacity limit. This limit coincided with those in Experiments 1 to 3, and was consistent with previous studies that suggested a capacity of three or four items for visual working memory (Luck & Vogel, 1997). We also examined the standard deviation SD (see Fig. 13b). Similar to the previous experiments, difference between post-cue and pre-cue SD occurred at set sizes three to eight, $ts(21) \geq 4.58$, $ps < .001$, $ds \geq 0.98$, but not at set sizes one or two, $ts(21) \leq 1.51$, $ps \geq .146$, $ds \leq 0.26$.

References

- Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567.
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.
- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*, 7345–7350.

- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157–162.
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, 70, 1325–1326.
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage and reallocation of visual working memory. *Journal of Vision*, 11(10): 1–15.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science* 22(3), 384–392.
- Cavanagh P., & Alvarez G.A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9, 349–354.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393–404.
- Chong, S. C., & Treisman, A. (2005a). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, 45(7), 891–900.
- Chong, S. C., & Treisman, A. (2005b). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, 67(1), 1–13.
- Da Silva, J. A. (1985). Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. *The American Journal of Psychology*, 98(1), 119–144.
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, 18(5), 1016–1026.
- De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, 70(5), 789–794.
- Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, 70(6), 946–954.
- Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience*, 15(8), 1176–1194.
- Grubert, A., & Eimer, M. (2016). The speed of serial attention shifts in visual search: Evidence from the N2pc component. *Journal of Cognitive Neuroscience*, 28(2), 319–332.
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, 144(2), 432.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, 17(7), 572–576.
- Hollingworth, A., & Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 39(4), 1047–1058.
- Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, 114(3), 599–631.
- Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, 317(5839), 823–825.
- Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, 43(7), 663–676.
- Jenkins, M., Grubert, A., & Eimer, M. (2018). The speed of voluntary and priority-driven shifts of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 44(1), 27–37.
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. London: Academic Press.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356.
- Morales, D., & Pashler, H. (1999). No role for colour in symmetry perception. *Nature*, 399(6732), 115–116.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7(4), 308–313.
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744.
- Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, 13(8): 1–11.
- Pylshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179–197.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14): 1–16.
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337. <https://doi.org/10.1037/a0028712>
- Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, 78(3), 392–402.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136.
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21(10), 1934–1945.
- Watamaniuk, S. N., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics*, 60(2), 191–200.
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast!. *Psychonomic Bulletin & Review*, 18(3), 484–489.
- Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Wener and L. M. Chalupa (Eds.), *The New Visual Neuroscience* (pp. 695–710). Cambridge, MA: MIT Press.
- Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 121.
- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1), 80–87.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235.