Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pecp21

Ensemble statistics as units of selection

Hee Yeon Im[ab], Woon Ju Park[b] & Sang Chul Chong[bc]
[a] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA
[b] Graduate Program in Cognitive Science, Yonsei University, Seoul, Korea
[c] Department of Psychology, Yonsei University, Seoul, Korea
Published online: 08 Dec 2014.

Journal of Cognitive Psychology, 2015
Vol. 27, No. 1, 114–127, http://dx.doi.org/10.1080/20445911.2014.985301

Routledge
Taylor & Francis Group

Ensemble statistics as units of selection

Hee Yeon Im[1,2], Woon Ju Park[2], and Sang Chul Chong[2,3]

[1]Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA
[2]Graduate Program in Cognitive Science, Yonsei University, Seoul, Korea
[3]Department of Psychology, Yonsei University, Seoul, Korea

The current study investigated whether attentional mechanisms operate on ensembles as higher-order units for selection. In Experiment 1, we presented sets of circles and asked participants to compare the mean sizes of the sets while concurrently detecting a small probe appearing at a centroid of one of the sets. We found that, both with and even without the task instruction to favour larger mean sizes, people's mean size judgement was more accurate for the sets with larger mean sizes. In addition, detection of the probe appearing in the set with the largest mean size was facilitated by a matching task instruction. However, when the task instruction favoured smaller mean sizes, mean size judgement became more accurate for the sets with smaller mean sizes. These results suggest that attentional selection can be based on ensembles. In Experiment 2, we found further evidence that attention was directed towards the centroid of an ensemble, rather than towards an individual member of the ensemble. Together, these results suggest that attentional modulation can operate at the level of ensembles instead of selecting individuals separately and that the centroid of an ensemble can be the locus of selection based on an ensemble.

Keywords: Ensemble statistics; Mean size comparison; Probe detection; Selective attention; Task setting.

The human visual system has evolved to process structured environments efficiently (Barlow, 1972). As the visual world is highly structured and predictable, one way to accomplish this efficiency is to form general descriptions of groups of similar items. In our everyday visual experience, we use this ability to represent the global properties of a set of similar objects as an ensemble. Imagine that you are buying apples at the fruit market. To select which apples you wish to buy, you may use many different visual aspects of the groups of apples quickly without scrutinising the individual apples too much, for example, which container has the most apples (at least approximately), or which container has the largest apples on average. Ensemble representation provides summary information about groups of objects.

Previous studies have shown that people can extract ensemble representations very rapidly and efficiently. For example, human observers can represent the mean size of a set of circles from a very brief visual array lasting as little as 50 ms (Chong & Treisman, 2003; but see also Whiting & Oriet, 2011). Observers are also remarkably accurate at extracting the mean size (Ariely, 2001; Chong & Treisman, 2003), mean orientation (Ariely, 2001; Dakin & Watt, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), average

Correspondence should be addressed to Sang Chul Chong, Department of Psychology, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Korea. E-mail: scchong@yonsei.ac.kr

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology [grant number 2011-0025005].
No potential conflict of interest was reported by the authors.

location of objects (Alvarez & Oliva, 2008) and the average emotion of faces in a crowd from visual images (Haberman & Whitney, 2007). Furthermore, previous studies have shown that the precision of ensemble representations of multiple objects was better than that for single individuals (Sweeny, Haroz, & Whitney, 2013). It has been suggested that the better precision of the representation of ensembles than of individual objects is due to the fact that the internal noise that limits the processing of ensemble representation is lower than that for a single object (Im & Halberda, 2012). Presumably, the internal noise for an ensemble is lower because of the averaging process in which the noise of multiple individual measurements cancels each other out (for review, see Alvarez, 2011). Ensemble representation has been found to be precisely extracted even when visibility of the items was impaired (Choo & Franconeri, 2010), and under conditions of reduced attention in which observers were asked to perform an additional task simultaneously (Alvarez & Oliva, 2009; Joo, Shin, Chong, & Blake, 2009). Ensemble representations are therefore robust and efficient codes that summarise the global visual properties of a group of multiple items.

Although many studies have demonstrated human observers' remarkable ability to extract ensemble representations, some empirical questions remain to be addressed: how can ensemble representations enhance visual cognition? And how do ensemble representations interact with other cognitive processes such as attention? Previous findings have provided evidence that ensemble representations may be treated as higher-order units of visual processing (e.g. visual working memory), in the same manner as individual objects. For example, the visual system can extract and maintain ensemble representations from up to three or four sets of items at once (Halberda, Sires, & Feigenson, 2006; Im & Chong, 2014), but no more than that. This is similar to what has been observed for individual objects, which suggests that representations of ensembles and of individual objects may both be similarly constrained by the capacities of visual attention and working memory (for review, see Feigenson, 2011).

If ensembles function as separate entities for the attention and visual working memory in the same way as individual objects, then it is hypothesised that an ensemble of multiple objects will interact with attentional mechanisms in the same way as an individual object does. In the current study, we investigated whether attentional mechanisms could be modulated at the level of ensembles, just as they can be at the level of individual objects. In the domain of individual objects, previous studies have shown that an object can be selected in a bottom-up manner based on certain visual features, such as the onset or offset of an item (Abrams & Christ, 2003; Yantis & Jonides, 1984) or its featural singularity (Theeuwes, 1992). These properties are prioritised because they are likely to contain important information about changes in the environment. The size of an object is also one of the visual domains that modulate stimulus-driven attention. For example, a large object has been observed to capture bottom-up attention (Proulx, 2010), and a larger target among smaller distractors is detected much more rapidly than in the opposite case (search asymmetry; Treisman & Gormican, 1988). These results show that a larger object can be prioritised during the selection process. It has been speculated that larger objects capture the attention because they are often ecologically more important and advantageous than smaller ones (e.g. Proulx, 2010).

Attentional selection can also be modulated according to the current goal or the task set (Green & Anderson, 1956). The object that is relevant to the task goal attracts top-down attention, being prioritised and selected for further processing by the participant. For example, a target object that is relevant to the current goal is prioritised over other salient items in a visual array (e.g. Folk, Remington, & Johnston, 1992; Folk & Remington, 1998; Most, Scholl, Clifford, & Simons, 2005).

We tested how these different attentional mechanisms interacted with ensemble representations of multiple objects. Specifically, we predicted that the ensemble with the largest mean size would be prioritised over other ensembles, and that the selection of ensembles would also be modulated by the goal of the current task.

In Experiment 1, we measured how accurately people could judge the mean size of multiple ensembles, as well as the response time (RT) for detecting a probe appearing at a centroid of one of the ensembles. Four ensembles were sown in total, and the members of a given ensemble were spatially intermixed with members of other ensembles, ensuring selection based on ensemble rather than on individual elements. To measure the accuracy in mean size judgement, participants were asked to first extract the mean sizes of all four ensembles and then compare only two of the sets at the end of each trial. Because attention enhances the processing of an attended stimulus, the attended stimulus is expected to be detected and identified more precisely and more rapidly than it would without direct attention (e.g. Sagi & Julesz, 1987; Urbach & Spitzer, 1995). Since object-based attention is drawn directly to the location of the stimulus (Egly, Driver, & Rafal, 1994), in our study participants' attention was expected to be directed to the location of the ensemble. Thus, if the ensemble with the larger mean size attracted bottom-up attention, then the accuracy of mean size judgement would be greater for the ensemble with the larger mean size, and the RT for probe detection would be faster at the location of the ensemble with the larger mean size.

We defined the location of an ensemble as the average location (i.e. centroid) of the elements belonging to the ensemble. The centroid can serve as the focus of ensemble-based attention; when it is unnecessary or impossible to process all the different locations of individual objects, the most logical way to yield the best performance is to use the average location. Previous studies have indeed shown that participants are very accurate in representing the centroid of multiple objects even when they could not accurately state the location of each element (Alvarez & Oliva, 2008). In addition, when looking at a stimulus composed of multiple random dot clusters, participants' saccades landed with high precision near the average location of the dots (Melcher & Kowler, 1999). Therefore, we expected to see attentional modulation based on the centroids of the ensembles as attentional foci.

We then tested whether a task goal also modulated the deployment of attention towards ensembles by simply changing the task instructions for mean size judgement. Previous studies have shown that task rules or instructions can modulate the deployment of attention for single objects (e.g. Folk et al., 1992; Ravizza & Carter, 2008). Based on the task demands, observers were able to narrow their selective attention by focusing on the information relevant to the current task instructions, and to prevent the cognitive system from processing any information that did not serve the current goal (Ravizza & Carter, 2008). With this in mind, we switched our mean size judgement task instructions from "which ensemble has a larger mean size" to "which ensemble has a smaller mean size".

When there are more objects than the visual system can process at any given time, the attentional mechanisms select only a few objects that are more important than the others in the visual image (two to four objects; Luck & Vogel, 1997; Pylyshyn, 1989; Rensink, 2002; Scholl & Pylyshyn, 1999). We expected that, due to the large number and complexity of items presented in our visual stimuli, changing the task instructions in this manner would bias the participants' top-down selection process towards prioritising the ensembles demanded by the task (i.e. the sets with smaller mean sizes).

In Experiment 2, we further tested whether prioritisation of ensembles with the largest mean sizes indeed operated on the ensembles themselves, rather than on individual members of the ensembles (e.g. the largest member). When participants were only performing the probe detection task, the probe appeared either at the location of the largest individual member of an ensemble, or at the location of the centroid of the ensemble. We hypothesised that if the mean size of an ensemble attracted bottom-up attention towards the entire ensemble rather than towards an individual member of that ensemble, then participants' probe detection would be faster when the probe appeared at the centroid of the ensemble than when it appeared at the location of the largest individual.

EXPERIMENT 1

Method

Participants. Eighty-eight undergraduate students of Yonsei University participated in this study, either for course credit or for monetary compensation. All the participants had normal or corrected-to-normal vision and were unaware of the purpose of the study. The experimental protocol was approved by The Institutional Review Board of Yonsei University, and signed informed consent forms were obtained from all participants.

Apparatus and stimuli. The stimuli were created using MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and presented on a linearised Samsung 21 monitor with a frame rate of 85 Hz. The participants were positioned approximately 90 cm from the display, with their heads positioned on a forehead-and-chin rest. At this distance, one pixel on the display was subtended at a visual angle of 0.016°.

The stimulus display contained 20 circles grouped as four sets of five circles with different

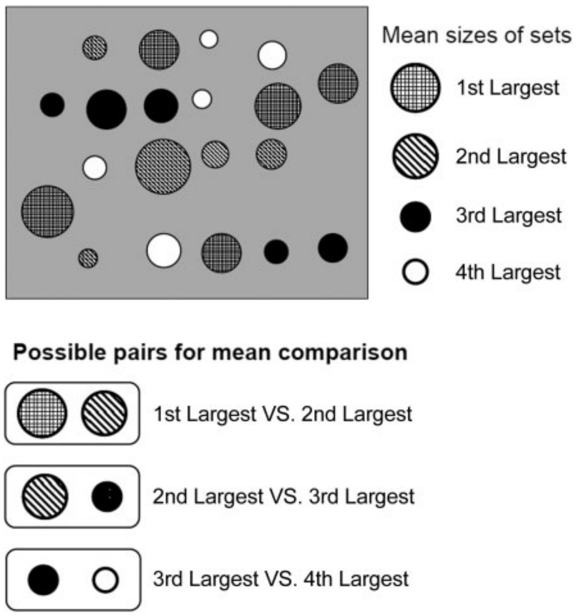

Figure 1. Example of a visual array and illustration of possible pairs of sets for the mean size judgement task in that array. The pair of sets for the mean size judgement task always included two sets adjacent in their mean sizes.

sizes of circle in each set (Figure 1). Sets in each trial were distinguished by being randomly assigned one of five colours (red, blue, yellow, green and cyan). The luminance for the five colours was 20.34 cd/m^2 for red, 10.53 cd/m^2 for blue, 92.87 cd/m^2 for yellow, 73.25 cd/m^2 for green and 83.06 cd/m^2 for cyan. The luminance of the grey background was 56.01 cd/m^2. For each trial, the mean size of the smallest set was chosen randomly from a uniform distribution ranging from 1.06° to 1.17°. The other three mean sizes were then determined in multiplicative steps of a 14% size difference. Multiplicative steps were used because the perceived size of circles follows a power function with an exponent of 0.76 (Teghtsoonian, 1965), and computation of the mean size also follows this function (Chong & Treisman, 2003). This 14% difference in proportion between the mean sizes of the sets yielded a range of 1.06–1.97° for mean size across the four sets. Once the mean sizes of the sets were determined, the sizes of the individual circles in each set were randomly generated within a range of 1.01–2.11° so that their combined average accorded with the given mean size of their set. The variance of elements within a set did not differ systematically across sets, and it ranged from 0.16° to 0.17°.

The array of possible circle locations was specified by an invisible 6 × 6 grid (each cell subtending 3° × 3°). The location of each circle was randomly determined, with the constraint that any two centroids for the four sets should not fall closer to one another than 4.8°. Each circle was presented with a slight spatial jitter within a uniformly distributed range of 0.64° from the centre of the cell. The test indicators, which signified the pairs of sets to be compared at the end of each trial, were 0.48° × 0.96° filled rectangles. Each indicator had a different colour chosen from one of the sets in the original display. Two indicators were presented 11.84° apart on either side of the screen's centre. In one-third of the trials, a probe for the detection task was presented. The probe was a small black square subtending at 0.12° × 0.12°. It appeared at the centroid of one of the sets presented in the visual array. The centroid location for each set was defined as the average location of the five circles belonging to the set, as in Alvarez and Oliva (2008).

Design. The current study had a 3 × 3 × 4 design, with three different task settings (larger, smaller and mixed), three pairs of mean sizes to be compared, and four probe locations (centroids) for the probe detection task. The task settings varied between participants and were blocked across the participants. The other variables were varied between tasks conducted by one participant. In the "larger" condition (n = 30), participants were instructed to judge which of the two indicated colour sets had the larger mean size; in the "smaller" condition (n = 29), they were asked to judge which had the smaller mean size. In the "mixed" condition (n = 29), in each trial a post-cue indicated whether they should judge the larger or smaller mean size. It was emphasised to the participants that they should respond based on the mean sizes of the sets of circles rather than on the individual sizes of circles. The two sets to be compared were always adjacent to one another in terms of their mean sizes, yielding three possible pairs of to-be-compared sets in total (Figure 1). The probe appeared with equal frequency at each of the four centroids, regardless of the sets to be compared. Each participant completed 192 trials, as well as one practice block of 12 trials. The order of the trials within each block was randomly selected. In the mixed condition, the total number of trials was doubled because the participants randomly reported both the larger and smaller mean size. It took participants 20 minutes to complete the larger and smaller conditions and 40 minutes to complete the mixed condition.

We controlled for the possibility that participants would simply pay attention to the largest individual member of a set rather than to the set with the largest mean size when comparing the two indicated sets. Comparing individual elements is a reasonable and natural strategy given that a larger member draws more bottom-up attention than other, smaller elements (Proulx, 2010; Treisman & Gormican, 1988), and because simply selecting the set containing the largest member in the array will be more likely to yield a correct response. To minimise the possibility that the participants would use this sampling strategy, we generated two different types of trial. In half of the trials, of the sets to be compared the set with the larger mean size contained the largest circle ("same" trials), whereas in the other half of the trials, the set with the smaller mean size contained the largest circle ("different" trials). This constraint ensured that the participants could not have recourse to simply comparing the largest circle in each set and judging the mean according to these extreme values; if they responded based on the individual size of the largest circle in each set, their accuracy in the mean size judgement task would only be 50%. During each trial, therefore, the sizes of the individual circles were generated until they met the given mean sizes and satisfied the constraint for either a "same" or "different" trial.

Procedure. The procedure followed is shown in Figure 2. Each trial began with the presentation of a fixation cross for 500 ms. The stimulus display was then presented for 1000 ms, at which point the participants were instructed to compute the mean sizes of the four sets, which were distinguished by colour. A grey screen with a fixation cross then appeared for 2000 ms. For one-third of the trials, a probe for the detection task was presented for 120 ms during this delay. The probe appeared briefly after the offset of the stimulus display, either after 30 or 60 ms in the "larger" condition, and either after 30, 60 or 90 ms in the "smaller" and "mixed" conditions. We varied the stimulus-onset asynchrony (SOA) between the offset of the stimulus display and the onset of the probe in order to ensure that our participants would not be able to predict the time at which the probe would appear. In response to the probe, the participants had to press the "z" key with their left hand as quickly as possible. If the participants missed the probe, or if they responded when no probe appeared, they heard an error sound. After the response, or after a duration of 2000 ms in the absence of a response, a grey screen with two colour indicators appeared. When there was no probe, the indicators were presented 2000 ms after the offset of the stimulus display. The participants were then asked to judge which of the two sets having the same colours as the indicators had either the larger or the smaller mean size, depending on the top-down task settings. In the mixed condition, an alphabet letter ("L" for larger, "S" for smaller) in the centre of the test display indicated whether they had to choose the set with the larger or the smaller mean size. Note that in every condition, the participants did not know which sets to compare until the indicators appeared. The colour indicators remained until the participants had responded by pressing the number "1" key if they thought the left indicator represented the set with the larger or smaller mean size (depending on the task settings), or by pressing the number "2" key if they thought the indicator on the right represented the required set.

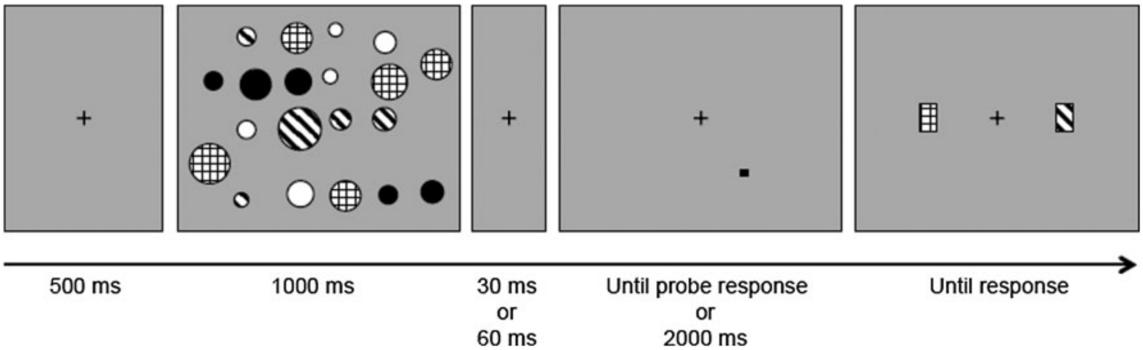

Figure 2. Timeline of trials in Experiment 1.

Auditory feedback was given on both the probe detection task and the mean size judgement task.

Results

Since the different SOAs between the offset of the stimulus and the onset of the probe display did not yield differences in any aspect of the data (all the *p*s were insignificant), we collapsed the data from the different SOAs when analysing accuracy and RT.

We examined whether trial type (same: the largest individual circle belonged to the larger set; different: the largest individual circle belonged to the smaller set) systematically affected performance when interacting with the other main factors. We conducted 2×3 mixed design and 2×4 repeated measures of analysis of variance (ANOVA) on the accuracies for the mean size judgement task and on the RT for the probe detection task to examine the interaction between the trial type and the other main factors: the task settings (three levels: larger, smaller, or mixed), the pair of mean sizes to be compared (three levels), or the probe location (four levels). Although the significant main effect of trial type (same vs. different) was on the accuracy for the mean size judgement task (same trials; 61% vs. different trials 58.7%; $F(1,85)$, $p < .01$), we found that the interaction between trial type and task setting was not significant ($F(2, 85) = 1.108$, $p = .335$). The interaction between trial type and the pair of mean sizes to be compared was also not significant ($F(2, 170) = 0.495$, $p = .610$). Moreover, we found that the main effect of trial type on RT was not significant ($F(1,85) = 3.095$, $p = .084$); nor were the interaction between trial type and the task settings ($F(2, 85) = 0.685$, $p = .507$) and the interaction between trial type and the probe's location ($F(3, 255) = 1.449$, $p = .229$) significant. Since trial type did not show any significant interaction with the other main factors under investigation (i.e. task settings, the pair of mean sizes to be compared and probe location), we collapsed the trial types (same and different) in further analyses. In the following analyses, Huynh–Feldt corrected p values were reported when the result of Mauchly's sphericity test was significant and Bonferroni correction was used for multiple comparisons.

We first looked at the overall accuracies for the mean size judgement task and the probe detection task in each of the three conditions (larger, smaller and mixed). All conditions produced an average accuracy that was better than chance. The overall accuracy of the mean size judgement task was 59.8%, which was significantly higher than that of chance ($t(87) = 15.217$, $p < .01$). The overall average accuracy of the probe detection task was 94%.

To determine whether participants were better at distinguishing the set with the larger mean size than the one with the smaller mean size in the different task settings, we compared the accuracies of the mean size judgement task according to the pairs of mean sizes and task settings (Figures 3A

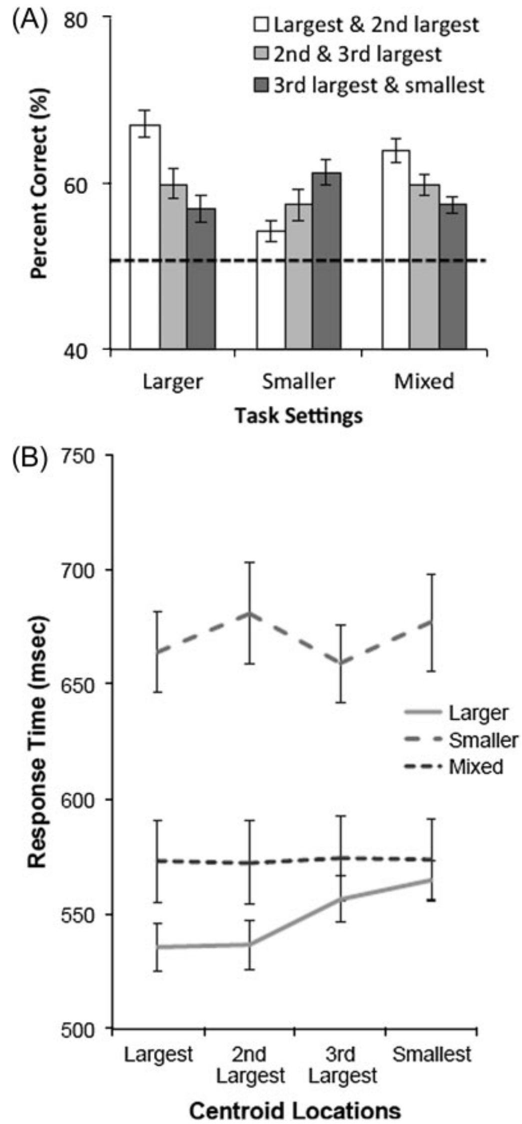

Figure 3. Results of Experiment 1: (A) Mean accuracy of mean size judgement for each pair of sets in the different task settings. The dotted horizontal line indicates the chance level (50%). (B) RT for detecting the probe appearing at the centroid of each set in the different task settings. The error bars indicate the standard error of the mean.

and 3B). A 3 × 3 mixed ANOVA was performed, with the pair of mean sizes to be compared (largest vs. second largest, second and third largest, and third largest and smallest pair) as the within-subjects variable and the three task settings (larger, smaller and mixed) as the between-subjects variable. The main effect of the task setting was not significant ($F(2, 85) = 2.991$, $p = .056$, $\eta^2 = 0.07$), but accuracy in the larger and mixed conditions was generally higher than that in the smaller condition. The main effect of the mean size pairs was significant ($F(2, 170) = 5.357$, $p < .01$, $\eta^2 = 0.07$). Specifically, the accuracy of the mean size judgement tasks was higher when the pair of sets with the two largest mean sizes were compared than it was for when other pairs were compared (all ps < .05). Importantly, the interaction between the task setting and the pairs compared was also significant ($F(4, 170) = 12.706$, $p < .01$, $\eta^2 = 0.23$).

To understand the nature of this interaction, we separately analysed the effect of each task setting on the accuracy of mean size judgement for three pairs of mean sizes. In the larger and mixed conditions, mean size comparison of the two sets with the largest mean sizes was the most accurate (larger condition: $F(2, 58) = 12.342$, $p < .01$, $\eta^2 = 0.30$; mixed condition: $F(2, 56) = 9.179$, $p < .01$, $\eta^2 = 0.25$). In other words, the mean size of the sets with the largest mean sizes was more accurately represented both in and outside the setting which encouraged participants to focus attention on these sets. However, in the smaller condition, the accuracy was higher for the pair of sets with the two smallest means than for the pair of sets with the two largest means ($F(2, 56) = 8.421$, $p < .01$, $\eta^2 = 0.23$), suggesting that the top-down task setting reversed the pattern of the results. It is important to note that better accuracy for larger mean sizes was also found in the mixed condition. Because the participants were not informed which task to perform until they had seen the ensembles, in this condition selecting only the larger sets was not the optimal strategy. In the mixed condition, participants did not simply choose one of the strategies that might have been used in the larger or smaller conditions. Most of the participants in the mixed group had a higher accuracy for larger sets, and the rest did not clearly perform better for smaller sets. Importantly, if an inherent bias towards larger mean sizes did not exist (so that the participants in the mixed group randomly chose a strategy favouring either "larger" or "smaller" sets), then we would expect the accuracies to be similar for all pairs in the mixed group. Thus, while the task instruction was a powerful source of attentional modulation, the fact that larger mean sizes were more accurately represented in the mixed condition suggests possible asymmetry in the representation of mean size which informs stimulus-driven modulations.

We next analysed the participants' RTs in the detection task in relation to the location of the probe and the task setting (Figure 3B). We selected only the trials featuring correct probe detection for further analysis of the RT. For each participant, we removed the outliers that were more than three standard deviations from the mean of all RTs for that participant. As a result, an average of 0.8% of the trials was excluded. A 4 × 3 mixed ANOVA analysis was performed, with a within-subjects variable of the probe locations (the centroid of the largest, second largest, third largest and smallest set) and a between-subjects variable of the task setting. The main effects of probe location and of the task setting were both significant (probe location: $F(3, 255) = 2.675$, $p < .05$, $\eta^2 = 0.15$; task setting: $F(2, 85) = 17.750$, $p < .01$, $\eta^2 = 0.26$). Specifically, probe detection was faster at the centroids of the two sets with the largest mean sizes than at the centroids of the two sets with the smallest mean sizes (all ps < .05). Moreover, probe detection was faster in the larger and mixed conditions than in the smaller condition (all ps < .01). It is important to note that in the mixed condition, the task setting changed between almost every trial, yet participants' RT in the mixed condition was significantly faster than in the smaller condition. Thus, participants in the mixed condition seemed to perform the task in a similar manner to the larger condition, presumably by prioritising the sets with the largest mean sizes, as if prioritisation of a larger mean size is a "default mode". On the other hand, probe detection in the smaller condition was much slower, suggesting that the selection of a set with a smaller mean size is not as intuitive as selecting a set with a larger mean size.

We also found a significant interaction between probe location and task setting ($F(6, 255) = 2.846$, $p < .05$, $\eta^2 = 0.09$), which gave us a better understanding of the effect of probe location on each of the task settings. In the larger condition, the RTs became significantly faster when the probe was presented in the centroid of the set with larger mean size ($F(3, 87) = 24.32$, $p < .01$, $\eta^2 = 0.46$). In other words, attention was allocated to the centroid of the set with the largest mean size, facilitating detection of the probe in this location. In the smaller and mixed conditions, however, we did not find a significant difference in RT for different probe locations (smaller: $F(3, 84) = 1.712$, $p < .171$; mixed: $F(3, 84) = 0.040$, $p = .989$). In other words, whereas participants generally responded to the probe faster in both the larger and mixed conditions, probe detection was only significantly facilitated by the probe appearing at the centroid of the largest set in the larger condition. One possible reason for this is that our probe detection task was secondary rather than primary, making the possible attentional modulation less effective. We also speculate that the nature of this difference in RT is rather additive and facilitatory, occurring only when the task instruction matches the bias inherent in the stimulus. In the smaller and mixed conditions, the interplay of different sources of attentional modulation may have weakened the stimulus-driven bias.

We found that participants were more accurate at comparing the groups with the two largest mean sizes than those with the smaller mean sizes. This result suggests that they prioritised the group with the largest mean size for further processing. The prioritisation of the largest group is not likely to have been due to perceptual discriminability, as we used Teghtsoonian's psychological scale to generate the mean sizes in order to equate the perceptual discriminability across the groups.

Although the Teghtsoonian scale, in principle, allowed us to use mean sizes with the same level of perceptual discriminability, we ran a control experiment in which we presented 15 new participants with visual arrays that contained only two groups of objects. The mean sizes of the two groups were randomly chosen from the four mean sizes used in Experiment 1, and the other aspects of the experimental procedure were identical to those in Experiment 1. We compared the accuracy of mean size comparison for each pair of mean sizes (i.e. comparing the first largest and second largest, the second and third largest, and the third and fourth largest). As shown in Figure 4, we did not observe any systematic difference in the accuracy of mean size comparison with the pair compared ($F(2,36) = 0.29$, $p = .75$). This result suggests that when the visual system did not need to select subsets from a visual array (i.e. when there were only two sets to be processed), perceptual discriminability did not differ between sets. Therefore, our main finding that the pair of sets with the two largest mean sizes yielded better

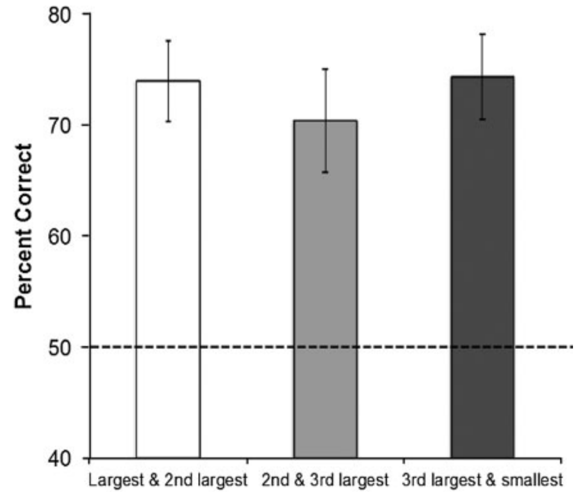

Figure 4. Results of the control experiment. The error bars indicate the standard error of the mean.

accuracy in mean size judgement reflects the prioritisation of larger sets by the attentional mechanism, rather than differences in the perceptual discriminability of the compared sets.

It might appear that attention is allocated to the largest member of an array rather than to the set with the largest mean size because a larger item always attracts more attention (Treisman & Gormican, 1988). However, this does not explain our results as the probe was always presented at the centroid of the set, not in any one of the elements' locations. Moreover, we ensured in our experimental design that the set with the largest mean size would not always contain the largest circle so that the centroid of the largest set and the location of the largest circle varied independently. Our data suggest that the set with the largest mean size was still prioritised even when the largest individual item belonged to a different set. When we separately analysed the data from the trials in which the largest circle did not belong to the set with the largest mean size, we found essentially the same result. There was a significant interaction between the task setting and the pairs compared in the mean size judgement performance ($F(4, 170) = 5.692$, $p < .01$). Specifically, participants were more accurate in comparing the sets with the largest mean sizes in the larger ($F(2, 58) = 4.827$, $p < .05$) and the mixed ($F(2, 56) = 5.566$, $p < .01$) conditions. However, the pattern was reversed in the smaller condition ($F(2,56) = 3.81$, $p < .05$), where participants were more accurate in comparing sets with the smallest mean sizes. In addition, participants more rapidly detected a probe presented at the centroid of the largest set ($F(3, 87) = 3.510$,

$p < .05$), and generally detected the probe faster in the larger and mixed task settings than in the smaller task setting ($F(2, 85) = 18.031$, $p < .01$). These results suggest that the attentional modulation observed in this study is more likely to have operated on ensembles, rather than on the largest individual member in the visual array. Previous studies have raised the concern that ensemble processing could have been replaced by an object-based sampling mechanism in which one or two elements in each set are selected and serve as representatives of their set (e.g. Myczek & Simons, 2008). Thus, we ran Experiment 2 to further examine attentional allocation in ensembles.

EXPERIMENT 2

In Experiment 2, we empirically tested whether attention was allocated to an ensemble as a unit, instead of to any individual member of an ensemble. Following the results of Experiment 1, we hypothesised that when ensembles need to be parsed and extracted, attention would be allocated towards the centroid of each ensemble. To test this hypothesis, we presented two sets of circles in a visual array and asked participants to detect and discriminate a probe (either a letter "T" or "L"). Presenting two sets of circles in a visual array provided many more individual items for selection and processing at any given time, encouraging participants to extract summary representations instead of processing individual items separately. Additionally, we used different mean sizes and different colours for each ensemble of circles in order to facilitate the parsing and segmenting of ensembles. In order to examine whether attention was allocated to the centroid of an ensemble instead of to individual items when there are multiple sets of objects to be represented, we directly compared the accuracy of probe discrimination when the probe appeared at the centroid of an ensemble with accuracy when it appeared at the location of the largest single element of the set.

Methods

Participants. Forty undergraduate students from Yonsei University participated in this study, either for course credit or for monetary compensation. None of these students had participated in Experiment 1.

Apparatus and stimuli. The apparatus and stimuli were identical to those in Experiment 1, except for the following changes. To facilitate the parsing of the two sets, in Experiment 2 we only used the sets from Experiment 1 with the largest and the smallest mean size, and we also used a different colour for each set. Examples of displays are shown in Figure 5. For a probe, we used one of the letters "T" and "L", instead of the black square of Experiment 1. The letter probe stimuli had a visual angle of $0.72° \times 0.72°$. The probe location was determined to be either at the centroid of the set or at the location of the largest individual member of one set. In order to minimise the perceptual differences between the probe locations, the locations of the circles were generated in a pseudorandom manner, with the constraint that the eccentricity of the probe should be identical at the centroid of an ensemble and at the location of the largest individual element.

Procedure. All aspects of procedure for Experiment 2 were identical to those of Experiment 1, other than those specified here. The probe was

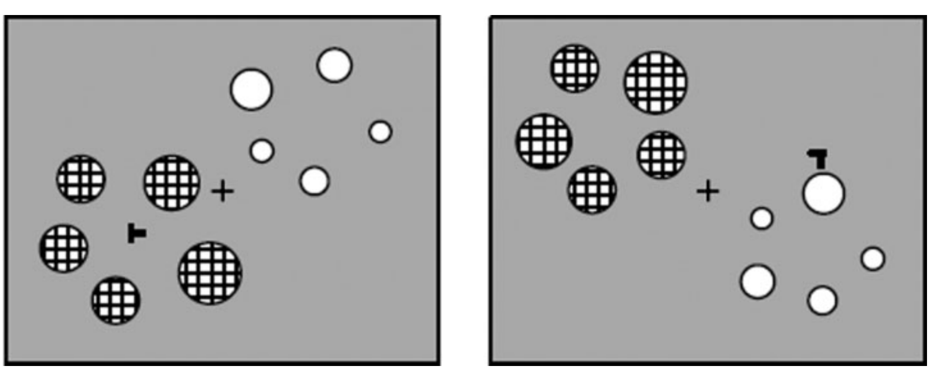

Figure 5. Sample displays for Experiment 2. A probe appeared either at the centroid of an ensemble of circles (left column) or adjacent to the largest individual circle within an ensemble (right column).

presented in every trial either at the centroid of an ensemble of circles or adjacent to the largest individual circle in a set. The task was to report whether the target probe was “T” or “L”. Note that there was no mean size judgement task; thus, the circles of the visual array were not at all relevant to the task.

Results

Figure 6 summarises the results of Experiment 2. We found that even when an ensemble of multiple circles in the visual array was not relevant to the task, the participants’ performance in identifying the probe target was significantly more accurate ($F(1, 39) = 7.93$, $p < .01$, $\eta^2 = 0.17$) and faster ($F(1, 39) = 7.41$, $p < .05$, $\eta^2 = 0.16$) when the probe was presented at the centroid of the set. Our finding of better performance in probe discrimination when the probe appeared at the centroid of the ensemble compared to when it appeared at the location of the largest circle in the ensemble was not due to perceptual difference such as the probe’s eccentricity, as we ensured that the eccentricity of the probe location was identical across the conditions. Therefore, the result of Experiment 2 further supports our claim that the centroid of an ensemble can function as the locus of selective attention operating on ensembles.

DISCUSSION

Ensemble statistics can be used to form an economical representation of a complex scene (e.g. Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003). Attention also plays a role in the efficient processing of a visual scene by selecting the most important information and prioritising it in order to maximise the outcome. The importance of information can be determined either by the bottom-up signal strength (i.e. attentional capture by onset or singularity; Folk et al., 1992; Yantis & Jonides, 1984) or by top-down relevance to the task goal (Green & Anderson, 1956). Here, we investigated how attentional selection mechanisms operate on ensemble representations. Our findings show that ensembles can be selected by attentional mechanisms in a similar manner to individual objects; the set with the largest mean size is selected preferentially, but this preference can be weakened or reversed by the top-down attentional mechanism. Our findings also reveal an intriguing aspect of ensemble-based attention: the centroid (i.e. average location) of an ensemble may serve as the focus of selective attention operating on the ensemble.

The set with the largest mean size attracted selective attention in the same way as a largest individual object (Proulx, 2010). Bottom-up attention can therefore operate not only on the individual size of an object but also on the mean size of a group of multiple objects, and in the same manner. This finding is consistent with previous studies on visual search, which have shown that searching for a large target among small distractors is easier and faster than vice versa (i.e. there is search asymmetry for a larger target; Treisman & Gormican, 1988). Our finding, however, is distinct from previous findings on visual search for one critical reason: our ensembles were different from large individual object (Treisman & Gormican, 1988) because the largest object did not necessarily belong to the largest ensemble, and the locus of attention was the centroid of an ensemble where no physical stimulus was actually presented.

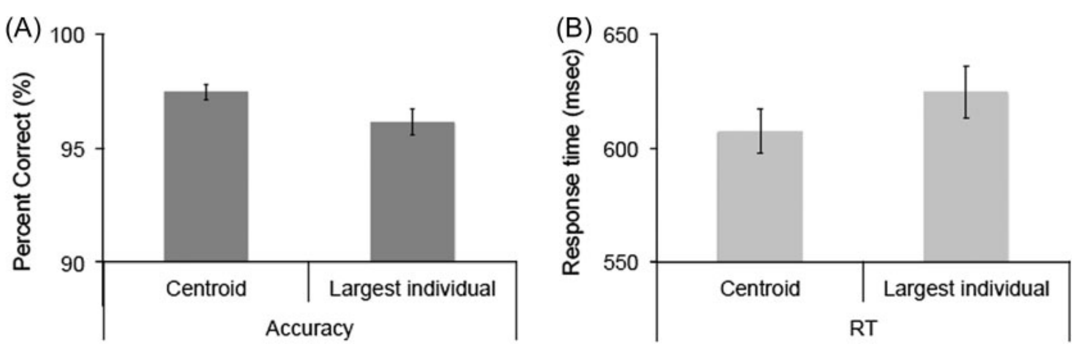

Figure 6. Results of probe discrimination from Experiment 2: (A) Mean accuracy of probe discrimination when the probe appeared at the centroid of an ensemble and at the location of the largest individual circle in an ensemble. (B) Mean RT for the probe discrimination task when the probe appeared at the centroid of the ensemble and at the location of the largest individual circle in an ensemble.

Why is attention attracted to a large object or a larger mean-sized group of objects? Although the underlying mechanism is unknown, a previous study demonstrated that this perceptive preference for larger objects is observed from early infancy (Newman, Atkinson, & Braddick, 2001). From a developmental perspective, it has been speculated that visual aspects of magnitude such as size, height or quantity may be internally represented as falling on a mental scale with "taller, larger or more" at the positive end and "shorter, smaller or less" at the negative end, and that children prefer and master the "positive" terms (i.e. taller, larger or more) earlier than the "negative" terms (i.e. shorter, smaller or less; Barner & Snedeker, 2008; Donaldson & Balfour, 1968; Klatzky, Clark, & Macken, 1973). "Positive" terms are used to denote "the excess … within a dimension, with respect to some 'standard'" (Vendler, 1968, p. 95). On the other hand, "negative" terms are used to denote a defect or deficiency with respect to a standard. This conceptual preference for positive attributes such as "larger" may also mediate the perceptual preference for larger objects and ensembles than smaller ones, as observed in this study.

Researchers have suggested that visual attention can be guided by various properties in the environment. For instance, scene information can direct our attention to the objects to be encoded (Hollingworth & Henderson, 2003). Also, repeated object locations and shapes form a context that can guide our attention to find a target (Chun & Jiang, 1998, 1999). The extraction of this scene information can be used to facilitate the perception of both subsequent scenes and objects within the scene (Sanocki & Epstein, 1997). Here, we have shown that ensemble representation, such as mean size of a set of items, can modulate the deployment of attention. This attentional guidance towards to sets with larger mean size may contribute to the visual search for objects in complex scenes via a non-selective pathway through which the basic statistical information of an ensemble, rather than of individual items, is extracted (Wolfe, Võ, Evans, & Greene, 2011).

Our results show that the task set can also modulate the attentional selection mechanisms operating on the mean sizes of ensembles. Different task instructions affected the manner in which participants' attention was deployed. This finding is consistent with those of previous studies on top-down attentional sets (Most et al., 2005; Kiss & Eimer, 2011). It is worth noting that unlike most previous studies on top-down attentional sets (Most et al., 2005; Kiss & Eimer, 2011), we changed the task demand within the single feature domain of mean size, rather than across different feature domains. Although most previous studies have asked participants to shift their attention across different feature sets, a task shift may also include a change of attentional deployment within a feature set, if the goal achievement depends on different aspects of the stimuli (Ravizza & Carter, 2008). The definition of "attentional set", however, remains to be clarified by further studies.

Most theories of attentional guidance, such as Guided Search (Wolfe, 1994) and Contingent Capture (Folk et al., 1992), rely on the notion that observers can be actively prepared to identify a target by selectively enhancing relevant features defining the target (i.e. feature-based attention). We therefore concluded that our manipulation of the task demand could encourage participants to pay more top-down attention to one of the two feature values in the domain of the mean size, in the same way as feature-based attention can be made to shift between red and blue stimuli in the domain of colour.

We also found intriguing evidence that the centroid of a set of multiple items can be the focus of guided attention for the ensemble statistics of the set. The centroid is an ensemble statistic that represents the average location of a set. That attention is guided towards the centroid of a set itself interesting given that there is possibly no physical stimulus at this location, even though it describes the average location of a set of items as a whole. Centroid information remains available to observers even when they cannot effectively localise individual members of the set (Alvarez & Oliva, 2008). Observers also position their eyes on the centroid of multiple items when they attentively track multiple items (Fehd & Seiffert, 2008). The centroid of a set, therefore, may serve as the locus of ensemble-based representation for selection, just as the location of an individual item can be the target of selective attention for further processing. To our knowledge, this is the first empirical evidence that the centroid can serve as the focus of ensemble-based attention.

Finally, the results of Experiment 2 suggest that the centroid of an ensemble attracted attention even when participants did not need to extract any kind of ensemble representation. In Experiment 2, the task was to report whether the target probe was the letter "T" or "L", and the circles were not at all relevant to the task. Although the ensembles of circles were not relevant to the task, we found that attention was allocated towards the centroid of an ensemble, rather than towards an individual member of the ensemble. When there were many items to be grouped into ensembles, it seems that the selection mechanism operated at ensembles level by allocating attention to the centroid of a set of multiple items. Therefore, the results of Experiment 2 not only verify that the centroid of a set of multiple items can serve as the focus of ensemble-based attention, but they also suggest that the centroids of ensembles can be extracted in a compulsive, obligatory manner to direct the deployment of attention, as are other ensemble feature of visual textures (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014; Oriet & Brand, 2013).

Together, our results suggest that the mean size of a set of multiple items can be used as a higher-order unit for attentional selection. Our results support the framework of a hierarchical visual representation in which different levels of representations can be extracted from objects and groups of objects in a flexible manner. The visual system can extract ensemble representations from a visual array and can then select some of the ensembles for further processing via attentional mechanisms, just as it does with individual objects.

In real-life situations, attentional selection based on ensembles instead of on individual objects may be sufficient for the purposes of understanding and interacting with complex visual scenes. When many similar items are present in a visual scene, it is highly inefficient to index and select each individual object separately. Instead, one may attend to those items through a single reference to the entire group. Representing an ensemble as a single unit of similar items increases the efficiency of the selection mechanism by providing more information in a compressed form about a larger number of objects than the visual system can process individually. The selected ensembles can then be manipulated and further stored together in the same way as individual objects (Im & Chong, 2014). Selection based on ensembles thus enhances visual processing (Alvarez, 2011; Brady & Tenenbaum, 2010, 2013) by increasing the diversity and flexibility of the visual system, and allowing more efficient use of its limited capacity to process items simultaneously.

Original manuscript received February 2014
Revised manuscript received October 2014
Revised manuscript accepted October 2014
First published online December 2014

REFERENCES

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*, 427–432. doi:10.1111/1467-9280.01458

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, *101*, 34–40.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131. doi:10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398. doi:10.1111/j.1467-9280.2008.02098.x

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. doi:10.1073/pnas.0808981106

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162. doi:10.1111/1467-9280.00327

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*, 371–394. doi:10.1068/p010371

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child Development*, *79*, 594–608. doi:10.1111/j.1467-8624.2008.01145.x

Brady, T. F., & Tenenbaum, J. B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 411–416). Austin, TX: Cognitive Science.

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109. doi:10.1037/a0030779

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, & Psychophysics*, *72*, 86–99.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. doi:10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365. doi:10.1111/1467-9280.00168

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*, 3181–3192.

Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British*

Journal of Psychology, *59*, 461–471. doi:10.1111/j.2044-8295.1968.tb01163.x

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177. doi:10.1037/0096-3445.123.2.161

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, *108*(1), 201–209. doi:10.1016/j.cognition.2007.11.008

Feigenson, L. (2011). Objects, sets, and ensembles. In S. Dehaene & E. Brannon (Eds.), *Attention and performance* (Vol. XIV, pp. 13–22). Oxford: Oxford University Press.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 847–858. doi:10.1037/0096-1523.24.3.847

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1030–1044. doi:10.1037/0096-1523.18.4.1030

Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of Experimental Psychology*, *51*(1), 19–24. doi:10.1037/h0047484

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, R751–R753.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*, 572–576. doi:10.1111/j.1467-9280.2006.01746.x

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory & Cognition*, *31*, 930–940.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*, 663–676. doi:10.1068/p7719

Im, H. Y., & Halberda, J. (2012). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*, 278–286.

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 1–12. doi:10.1167/9.9.7

Kiss, M., & Eimer, M. (2011). Attentional capture by size singletons is determined by top-down search goals. *Psychophysiology*, *48*, 784–787. doi:10.1111/j.1469-8986.2010.01145.x

Klatzky, R. L., Clark, E. V., & Macken, M. (1973). Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual? *Journal of Experimental Child Psychology*, *16*(1), 32–46. doi:10.1016/0022-0965(73)90060-X

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research*, *39*, 2929–2946.

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What you see is what you set: Sustained in attentional blindness and the capture of awareness. *Psychological Review*, *112*(1), 217–242. doi:10.1037/0033-295X.112.1.217

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*, 772–788. doi:10.3758/PP.70.5.772

Newman, C., Atkinson, J., & Braddick, O. (2001). The development of reaching and looking preferences in infants to objects of different sizes. *Developmental Psychology*, *37*, 561–572. doi:10.1037/0012-1649.37.4.561

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8–16.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744. doi:10.1038/89532

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366

Proulx, M. J. (2010). Size matters: Large objects capture attention in visual search. *PLoS One*, *5*, e15293. doi:10.1371/journal.pone.0015293.t002

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97. doi:10.1016/0010-0277(89)90014-0

Ravizza, S. M., & Carter, C. S. (2008). Shifting set about task switching: Behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia*, *46*, 2924–2935. doi:10.1016/j.neuropsychologia.2008.06.006

Rensink, R. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. doi:10.1146/annurev.psych.53.100901.135125

Sagi, D., & Julesz, B. (1987). Short-range limitation on detection of feature differences. *Spatial Vision*, *2*(1), 39–49. doi:10.1163/156856887X00042

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*, 374–378. doi:10.1111/j.1467-9280.1997.tb00428.x

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290. doi:10.1006/cogp.1998.0698

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*, 329–337. doi:10.1037/a0028712

Teghtsoonian, M. (1965).The judgment of size. *American Journal of Psychology*, *78*, 392–402. doi:10.2307/1420573

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*, 599–606. doi:10.3758/BF03211656

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48. doi:10.1037/0033-295X.95.1.15

Urbach, D., & Spitzer, H. (1995). Attentional effort modulated by task difficulty. *Vision Research*, *35*, 2169–2177. doi:10.1016/0042-6989(94)00305-X

Vendler, Z. (1968). *Adjectives and nominalizations*. The Hague: Mouton.

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, *18*, 484–489. doi:10.3758/s13423-011-0071-3

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. doi:10.3758/BF03200774

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*, 77–84. doi:10.1016/j.tics.2010.12.001

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 601–621. doi:10.1037/0096-1523.10.5.601