Rapid averaging? Not so fast!

Brittany F. Whiting · Chris Oriet

Abstract Previous research suggests that sets of similar items are represented using a rapid averaging mechanism that automatically extracts statistical properties within 50 ms. However, typically in these studies, displays are not masked, so it is possible that the sets are available for longer than this duration. In the present study, using masked displays, we (a) tested a newly proposed strategy for extracting the mean size of a set of circles, and (b) more precisely evaluated the time course of rapid averaging. The results indicate that when viewing conditions are poor, performance can be explained by assuming that observers rely on information from previous trials. In this study, observers required at least a 200-ms exposure time in order to derive the average size of a set of circles without relying on information from previously-viewed sets, suggesting that rapid averaging is not as fast as previously assumed and, therefore, that it may not be an automatic process.

When shown a set of similar items, people can rapidly summarize the set according to statistical properties, such as the mean size. Ariely (2001) found that observers were able to determine the average size of a set of circles, but were unable to identify individual members of the set. Ariely interpreted this as evidence that the visual system can derive a statistical representation of the set without retaining specific information about the items within the set. Researchers have proposed that this is accomplished using a specialized averaging mechanism that evaluates all of the items in the set in parallel. Consistent with this proposal, Chong and Treisman (2005a) showed that averaging performance was better when attention was broadly distributed across a display than when attention was narrowly focused, suggesting that the specialized averaging mechanism operates preattentively, outside the focus of attention. Additional evidence for the automaticity of this process has been demonstrated through cuing and dual-task manipulations. Chong and Treisman (2005b) precued or postcued the relevant subset of a set of circles to be averaged. No benefit of precueing the relevant subset was observed, suggesting that observers were able to compute the mean size of two subsets of circles as easily as one. Additionally, conditions that normally limit the ability to encode information, such as performing a concurrent task (Chong & Treisman, 2005a; Joo, Shin, Chong, & Blake, 2009), or extracting the mean size of circles presented in rapid serial visual presentation sequences (Corbett & Oriet, 2010; Joo et al., 2009), have little effect on performance. Taken together, the available evidence suggests that the averaging mechanism operates preattentively, that there is no cost to computing two means concurrently, and that limiting central attention does not influence averaging performance. These findings strongly suggest that deriving the average size of a set of items is an automatic process.

Myczek and Simons (2008; see also de Fockert & Marchant, 2008) caution that before accepting the existence of a new automatic averaging mechanism, researchers must first show that performance on mean judgment tasks cannot be accounted for by focused attention strategies. Specifically, their results suggest that a number of strategies would

B. F. Whiting · C. Oriet
University of Regina,
Regina, SK, Canada

C. Oriet (✉)
Department of Psychology, University of Regina,
3737 Wascana Parkway,
Regina, SK, Canada S4S 0A2
e-mail: chris.oriet@uregina.ca

allow observers to guess the average size of the set from a small sample and to achieve high accuracy on the task. Given that these strategies do not exceed the limits of focused attention and are capable of fitting the data, Myczek and Simons suggest that it is unnecessary to posit a new averaging mechanism to explain the results obtained from previous experiments. However, Chong, Joo, Emmanouil, and Treisman (2008) argued that it is unlikely that observers would be able to first identify the appropriate strategy and then implement it before the display was removed, especially when the displays are presented briefly (e.g., 200 ms in this study).

In fact, according to Chong and Treisman (2003), observers can carry out this operation with displays as brief as 50 ms. In Chong and Treisman's (2003) study, observers compared side-by-side displays of heterogeneous or homogeneous sets of circles, or of single circles. Only the heterogeneous sets required a calculation of the mean. The results indicate that observers were able to average and compare two sets of circles as easily as they were able to compare two single circles, suggesting that the mean was determined automatically. If this process is indeed automatic, it should be carried out very rapidly. To test this hypothesis, Chong and Treisman (2003) varied the exposure duration of the display from 50 to 1,000 ms. Consistent with an automatic process, performance was not strongly affected by exposure duration. Given that visual search studies often report increases on the order of 20 to 30 ms per item as the number of items in the display increases (Wolfe, 1998), a 50-ms exposure duration does not provide adequate time to evaluate each individual item in the set. Performance under these conditions supports the claim that the set was evaluated in parallel because the display is presented too briefly to use a slow serial process of inspecting individual items.

To accept this conclusion, however, it is necessary to assume that observers are unable to continue processing the set after it has been removed from the display; given that the displays were unmasked, this seems unlikely (Enns & Di Lollo, 2000). If unmasked sets are available for processing for longer durations, then it increases the likelihood that observers could be using the focused attention strategies proposed by Myczek and Simons (2008). One solution that ensures the displays are available to be processed only for a specific, limited duration is to mask the items within the set. If observers are unable to compute the average size of the circles within this limited exposure duration, it may not, in fact, be possible to determine the mean size from the display as quickly as has been suggested by Chong and Treisman (2003).

To date, researchers have generally assumed that estimates of the mean size of a set on a given trial are guided (primarily or exclusively) by information provided on that trial only. In an attempt to discourage reliance on information from previous trials, researchers have introduced trial-by-trial variability in mean size by multiplying the circles in each set by a constant value (e.g., Ariely, 2001). Additionally, the distribution of trial means has generally been rectangular, such that each trial mean occurs with equal frequency. Introducing variability in trial means, however, leads only to a larger range of possible trial means, and even if trial means occur with equal frequency, observers may nevertheless acquire a representation of the mean of these trial means (i.e., cumulative mean; Crawford, Huttenlocher, & Engebretson, 2000), which could be used to aid judgments of the trial mean. This will obviously be of benefit if trial means are determined randomly and are normally distributed (Choo & Franconeri, 2010; Corbett & Oriet, 2010) because many trial means will be similar to the cumulative mean. Even if trial means are chosen with equal frequency and follow a rectangular distribution (Chong & Treisman, 2005a, b; de Fockert & Marchant, 2008), some trial means, by definition, will be similar to the cumulative mean. If observers are given a choice between a probe that corresponds to the trial mean and one that differs by a fixed amount, this will usually mean that the option closer to the cumulative mean is as likely to be correct as it is to be incorrect. However, the more closely the trial mean approximates the cumulative mean, the greater the probability that the target will be closer to the cumulative mean than will the distractor. Specifically, if the difference between the size of the target and the cumulative mean is smaller than the difference between the distractor and the cumulative mean, then choosing the probe closer to the cumulative mean leads to choosing the target.

Consider, for example, an experiment that uses three trial means that have been determined by multiplying the circles in one set by a constant value (e.g., .9, 1.0, and 1.1; de Fockert & Marchant, 2008). If each trial mean occurs with equal frequency, then on one-third of the trials, observers will be given a choice between the cumulative mean of the set and a distractor that is a fixed percentage (e.g., 13.8%) smaller or larger than both the trial mean and the cumulative mean. Observers choosing the circle closest to the cumulative mean could choose the target on 100% of these trials without needing to compute the trial mean at all. Although this strategy would lead to chance performance on the other two-thirds of trials, performance could nevertheless be very good across the whole experiment (e.g., .5X + 1.0X + .5X = 2X; if X = 100 trials of each type, then observers would choose the correct response on two-thirds of trials). Even if the trial means do not exactly correspond to the cumulative mean but are similar, this strategy could be used successfully as long as the distractor is further from the cumulative mean than is the target (Fig. 1). If observers do use such strategies, they would be
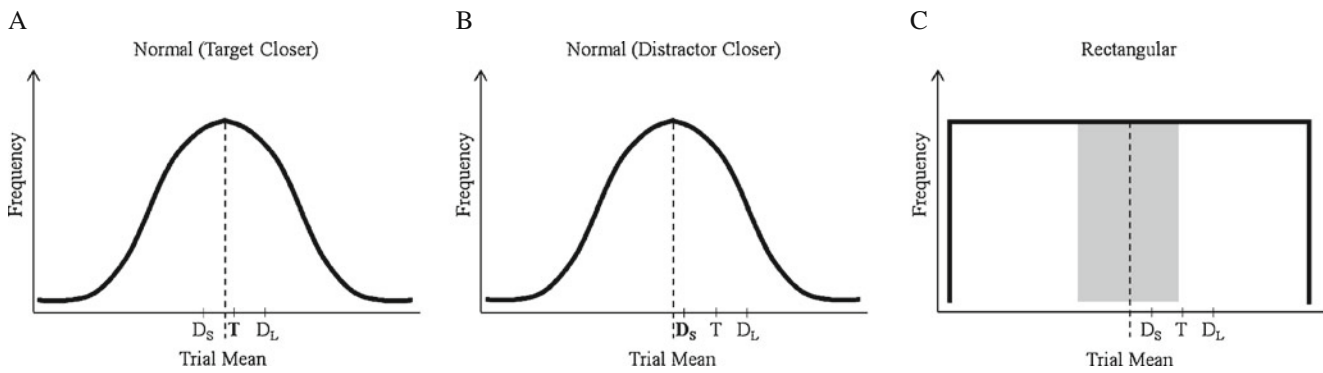
Fig. 1 *Dashed line* indicates cumulative mean of each distribution. **a** Normal distribution of trial means. Most frequently occurring trial means are similar to cumulative mean, thus choosing the probe closest to the cumulative mean favors selection of target (T) over smaller (D_S) or larger (D_L) distractors. In the example shown, irrespective of whether T is paired with D_S or D_L, T will be chosen. **b** Normal distribution of trial means. Less frequently occurring trial means are dissimilar from cumulative mean; thus, choosing the probe closest to the cumulative mean leads to selection of the target or distractor with equal frequency. In the example shown, if T is paired with D_S, D_S will be chosen, but if T is paired with D_L, T will be chosen. **c** Rectangular distribution of trial means. In this distribution, the target (T) cannot be similar to the cumulative mean (excluded trial means indicated by shaded region), and each trial mean occurs with equal frequency. If T is paired with D_S, D_S will be chosen, but if T is paired with D_L, T will be chosen, so target and distractor should be chosen with equal frequency

particularly advantageous when the visibility of the set is reduced. Thus, before concluding that observers can determine the mean size of a set in as little as 50 ms (Chong & Treisman, 2003), it is important to first ensure that performance does not simply reflect a strategy of relying on the cumulative mean when visibility is limited.

The present study

Following Chong and Treisman's (2003) work, we manipulated the exposure duration of sets of circles and instructed observers to compute the mean size of the sets in order to measure the speed of the averaging process. Unlike Chong and Treisman (2003), however, we interrupted processing of the displays with a trailing pattern mask to ensure that displays were available for processing for only a specific manipulated duration. To encourage observers to maintain a distributed state of attention and avoid inadvertently drawing attention to specific items (de Fockert & Marchant, 2008), all items in the display were masked. To test whether observers rely on information accrued over previous trials, we analyzed performance as a function of whether choosing the option closest to the cumulative mean would lead to selection of the correct (trial mean) response. Additionally, to manipulate whether relying on the cumulative mean was actually a useful strategy, we varied the frequency with which each trial mean occurred.[1] In the normal distribution condition, means that were similar to the cumulative mean occurred much more frequently than means that were

dissimilar to the cumulative mean, so targets were closer to the cumulative mean than were distractors on 75% of trials. In the rectangular distribution condition, each possible trial mean occurred with equal frequency but, unlike in previous experiments, no trial means corresponded to values similar to the cumulative mean, so targets were closer to the cumulative mean than were distractors on 50% of trials. This manipulation of the distribution of trial means allowed us to test (a) whether observers relied on the cumulative mean in making their judgments of trial means, (b) whether doing so had to reliably lead to selection of the correct response for this to occur, and (c) whether there was any benefit to actually showing sets whose average sizes corresponded closely with the cumulative mean. If, rather than relying on the cumulative mean, observers are simply sensitive to the overall range of possible trial means, they might refer to the midpoint of the range to make their judgment when uncertain. If determining the average size of a set of similar items is a rapid process that requires only 50 ms, then observers should be able to perform the task above chance with a 50-ms masked display, even when relying on an alternative strategy would lead to choosing the incorrect probe, and increases in exposure duration should have little effect (Chong & Treisman, 2003).

Prior to carrying out the main experiment, we conducted a pilot experiment using similar methodology. Twenty observers were shown displays of 12 circles for 0 (i.e., no circles), 50, 100, or 1,000 ms, followed by a 400-ms pattern mask. They were then instructed to choose which of two test circles corresponded to the average size of the circles shown. On circle-present trials, either a correct choice (trial mean) and an incorrect choice (distractor; randomly 20%

---

[1] We thank Heeyoung Choo and Steve Franconeri for suggesting this manipulation.

larger or smaller than the trial mean) were shown, or two distractors were shown (one 10% smaller, one 10% larger than the trial mean). On circle-absent trials, observers chose between a test circle whose size corresponded to the cumulative mean and a distractor that was randomly 20% larger or smaller. Three key findings emerged: (a) Irrespective of whether the correct choice was available, observers preferred the option closer to the cumulative mean; (b) this tendency was strongest at short stimulus onset asynchronies (SOAs; Fig. 2); (c) the probability of choosing the trial mean circle at short SOAs was comparable to the probability of choosing the cumulative mean circle on circle-absent trials, suggesting that observers are not using trial information to estimate mean size at short SOAs. The full details of this experiment are reported elsewhere (Whiting & Oriet, 2010).

Method

Participants

Forty-six observers volunteered their participation for partial course credit. Two did not complete the experiment and were replaced. All observers self-reported normal or corrected-to-normal vision and were between the ages of 18 and 43 ($M = 22.6$; $SD = 4.90$).

Stimuli and apparatus

Stimuli consisted of black outlined circles presented on a white background in 11 randomly chosen locations within
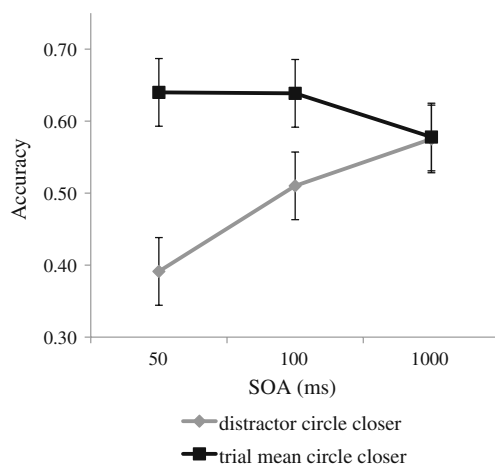
Fig. 2 Pilot experiment. Accuracy (i.e., selecting the circle corresponding to the trial mean) is displayed separately for trials on which the trial mean circle is closer in size to the cumulative mean versus trials on which the distractor circle is closer in size to the cumulative mean, as a function of stimulus onset asynchrony (SOA). *Error bars* represent 95% within-subjects confidence intervals

an imaginary 5×6 matrix encompassing the entire display. Each circle was unique in size, and diameters ranged from .12 to 4.8 degrees of visual angle in steps that were equally spaced on a power function with the exponent .76 (Teghtsoonian, 1965), resulting in eight possible trial means. The cumulative mean was calculated on a trial-by-trial basis by summing the means of the current and all previous trials and dividing by the number of trials completed. For all participants, the cumulative mean became relatively stable by the end of the first block of trials and equalled 2.0 degrees of visual angle. Some displays were masked by presenting identical masks at each of the 11 locations previously occupied by circles. Each mask consisted of a set of 10 randomly oriented 1.84° lines, determined randomly on each trial. A PC displayed stimuli on a 19-in. CRT monitor set to refresh at 60 Hz, viewed from a distance of approximately 60 cm.

Procedure

The experiment began with one block of 10 practice trials. There were two phases of experimental trials (Phase 1, unmasked displays; Phase 2, masked displays), each with four blocks of 36 trials, and all variables were manipulated within blocks. Each trial began with a fixation cross that was located in the center of the display for 1 s. The circles were displayed for a variable display-to-mask SOA of 50, 100, 200, 500, or 1,000 ms, and, if masked, they were masked for 400 ms. Following the mask, a test circle was presented on each side of fixation, and observers chose the test circle whose size corresponded to the average size of the set shown. The diameter of one circle corresponded to the trial mean, and the other was randomly 30% larger or smaller. Observers pressed the "4" or "6" key on the keyboard to choose the left or right test circle, respectively. Responses were unspeeded, and no feedback was provided.

Observers were assigned to one of two groups differing with respect to how frequently trials with each of the eight possible mean diameters (1.14, 1.37, 1.61, 1.86, 2.12, 2.40, 2.68, or 2.96 degrees of visual angle) occurred within each block of 36 trials. The eight mean sizes occurred in a 6:6:6:0:0:6:6:6 ratio for the rectangular distribution group and in a 2:4:5:7:7:5:4:2 ratio for the normal distribution group.

Results

Mean accuracy was computed for each observer as a function of group (rectangular vs. normal distribution), test circle (target closer vs. distractor closer to cumulative mean), SOA (50, 100, 200, 500, 1,000 ms), and display (unmasked vs. masked), and was averaged to produce

group means; results are shown in Fig. 3. A mixed-model ANOVA was used to analyze accuracy. There was a strong suggestion of an interaction between test circle, SOA, and display, $F(4, 176) = 2.33$, $MSE = .020$, $p < .06$, $\eta_p^2 = .050$, so we analyzed the Test Circle X SOA interaction separately for each display condition. This interaction was not significant for the unmasked trials ($F < 1$) but was significant for the masked trials, $F(4, 180) = 5.30$, $MSE = .021$, $p < .001$, $\eta_p^2 = .11$. Accuracy was relatively unaffected by SOA (ranging from .62 to .68) when the target circle was closer to the cumulative mean than was the distractor, but was strongly affected by SOA (ranging from .47 to .60) when the distractor circle was closer to the cumulative mean than was the target. The distribution of trial means had little effect on performance; only the Group X SOA interaction was significant, $F(4, 176) = 2.66$, $MSE = .023$, $p < .04$, $\eta_p^2 = .06$. Examination of Fig. 3 suggests that this interaction may have resulted from a difference in how SOA affected the unmasked trials when the distractor was closer to the target in the two groups (i.e., left panel; open squares vs. open triangles).

To determine whether accuracy exceeded chance at each SOA, we computed one-sample t tests at $\alpha = .05$ using Tukey's HSD posthoc correction. When the target was closer to the cumulative mean than was the distractor, performance reliably exceeded chance in all but a few conditions. When the distractor was closer to the cumulative mean than was the target, however, many comparisons, particularly those at the shortest SOAs, did not differ reliably from chance (Fig. 3).

The results suggest that when the strategy favored selection of a distractor, estimates of the trial mean were comparable to chance. In this condition, observers must rely on trial level information, and it is clear that their ability to do so was compromised at short SOAs, particularly with masked dis-

plays. Observers in the rectangular distribution condition were apparently unaffected by the absence of trials with means similar to the cumulative mean; the tendency to favor the option closer to the cumulative mean was equally strong in the two distribution conditions and influenced performance similarly in both. When the visibility of the circles was limited by a short exposure duration, observers could compensate for this limitation by choosing the option closer to the cumulative mean. When this option was, in fact, a distractor, only trial level information could be used to guide this judgment; thus, performance could not exceed chance-with limited visibility in this circumstance.

To summarize, when trial level information was limited by a brief exposure and relying on the cumulative mean strategy led to choosing the wrong probe, choosing the correct response could reliably be accomplished neither by averaging the circles shown, nor by relying on the cumulative mean strategy.

Discussion

The results suggest that when there is limited exposure to the items within a set, observers use information from previous trials to aid in their determination of the average size. Interestingly, this information was used even when no trial means were similar to the cumulative mean (i.e., rectangular distribution). This would be expected if observers were unaware that when trial means are normally distributed, those close to the cumulative mean occur more frequently than those further from the cumulative mean. Alternatively, the strategy may be used only when the trial mean is dissimilar to the cumulative mean, and when it is clear which of the two options is closer to the cumulative mean (e.g., Fig. 1b). If so, the frequency of such trials would

Fig. 3 Accuracy (i.e., selecting the circle corresponding to the trial mean) when the distractor circle is closer in size to the cumulative mean (*left panel*) versus when the trial mean circle is closer in size to the cumulative mean (*right panel*) as a function of SOA. Values above the *dashed line* are significant at $p < .05$, calculated using Tukey's HSD posthoc correction. *Error bars* represent 95% within-subjects confidence intervals
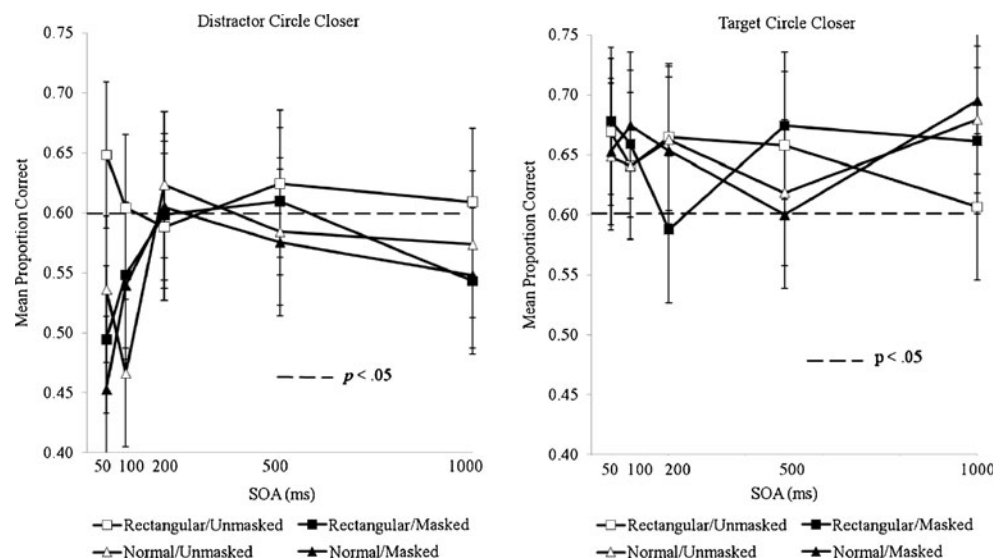
be similar across the two distributions, leading to similar performance across the two groups. In fact, if observers used this strategy only when the target and distractor each differed from the cumulative mean by more than 10 pixels (.4°), they could succeed on between 75–85% of those trials.

Although it is unclear why performance differed little across the two distribution types, it is quite clear that when given a choice between a circle corresponding to the trial mean and a distractor, observers' choices were strongly biased toward the option closer to the cumulative mean, irrespective of which was correct. This was especially true at short exposure durations with masked stimuli, and, in contrast to the results of our pilot work, this strategy appears to be used even at longer exposure durations. Thus, observers were clearly using information beyond that provided in the display to aid their judgments of mean size. The apparent insensitivity to frequency, however, hints that observers learn about the range of stimuli used over the experiment and reference the midpoint of the range rather than the cumulative mean, per se.

Researchers point to the speed with which accurate judgments of mean size can be made as evidence of the automaticity of this process (Chong & Treisman, 2003). Evidence for this claim rests on the assumption that processing of set information is discontinued after the removal of the display. Without the use of masked displays, however, this assumption is questionable. Thus, observers may be using their overall impression of the cumulative mean, or, more likely, the range of possible means in determining their responses.

The present study suggests that previous experiments overestimate the speed with which observers are able to determine average size (i.e., 50 ms), suggesting instead that performance on mean judgment tasks with limited visibility can be explained by reliance on information accumulated across trials. A strict interpretation of the results would suggest that observers cannot succeed in estimating the trial mean if the target is not reliably closer to the cumulative mean than is the distractor. This seems unlikely, given the wide range of conditions under which perceptual averaging has been demonstrated. Nevertheless, future studies could control for this strategy by ensuring a similar number of trials on which the distractor is closer to the cumulative mean than the target, and vice versa. Alternatively, assuming that observers are not using the strategies outlined by Myczek and Simons (2008), and that they can compute the mean size of two subsets of circles concurrently, using the side-by-side comparison method (Chong & Treisman, 2003) in which observers determine which of two displays of circles has the larger average size could circumvent this problem. Because observers are required to make a relative judgment, performance would not benefit from referring to information accrued across trials.

In conclusion, although the present findings do not rule out the possibility of an automatic averaging mechanism, they suggest that research designs must carefully consider the contribution of information accrued from previously-viewed trials in evaluating the speed of this mechanism. If performance is neither better than what is expected by a focused attention strategy (Myczek & Simons, 2008) nor consistent with the time course of a parallel, preattentive process, there is little to suggest that rapid averaging is automatic.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.

Chong, S. C., Joo, S. J., Emmanouil, T., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70*, 1327–1334. doi:10.3758/PP.70.7.1327

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

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67*, 1–13.

Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*, 891–900. doi:10.1016/j.visres.2004.10.004

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, & Psychophysics, 72*, 86–99. doi:10.3758/APP.72.1.86

Corbett, J. E., & Oriet, C. (2010). *Statistical representation and the temporal allocation of attention: Mean size is rapidly extracted over attentionally-demanding conditions*. Manuscript submitted for publication.

Crawford, L. E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science, 11*, 280–284.

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70*, 789–794. doi:10.3758/PP.70.5.789

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345–352.

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*, 1–12. doi:10.1167/9.9.7

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

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

Whiting, B. F., & Oriet, C. (2010, June). *How rapid is rapid averaging?* Poster presented at the meeting of the Canadian Society for Brain, Behavior and Cognitive Science, Halifax, Nova Scotia, Canada.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*, 33–39. doi:10.1111/1467-9280.00006