Title: Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking.

Authors and Affiliations

Oscar Jacoby[1]

Marc R. Kamke[1]

Jason B. Mattingley[1,2]

[1]Queensland Brain Institute, The University of Queensland, St Lucia 4072, Australia.

[2]School of Psychology, The University of Queensland, St Lucia 4072, Australia.

Running Head: FEATURE AVERAGING & OSM

Correspondence concerning this article should be addressed to:

Oscar Jacoby
The University of Queensland
Queensland Brain Institute
St Lucia, QLD 4072, Australia
E-mail: o.jacoby@uq.edu.au
Phone: +617 3346 6326
Fax: +617 3346 6301

Abstract

We have a remarkable ability to accurately estimate average featural information across groups of objects, such as their average size or orientation. It has been suggested that, unlike individual object processing, this process of *feature averaging* occurs automatically and relatively early in the course of perceptual processing, without the need for objects to be processed to the same extent as is required for individual object identification. Here, we probed the processing stages involved in feature averaging by examining whether feature averaging is resistant to object substitution masking (OSM). Participants estimated the average size (Experiment 1) or average orientation (Experiment 2) of groups of briefly presented objects. Masking a subset of the objects using OSM reduced the extent to which these objects contributed to estimates of both average size and average orientation. Contrary to previous findings, these results suggest that feature averaging benefits from late stages of processing, subsequent to the initial registration of featural information.

Keywords: statistical processing, mean size judgment, ensemble coding, object substitution masking, recurrent processing

At any given moment, our visual environment contains more information than can be consciously perceived. The number of individual objects or locations we can accurately attend to and identify is severely limited (Alvarez & Franconeri, 2007; Franconeri, Alvarez, & Enns, 2007; Palmer, 1990). Despite these limitations, we can quickly and accurately extract average featural information (hereafter referred to as *feature averaging*) about relatively large groups of objects, such as their average size (Ariely, 2001; Brady & Alvarez, 2011; Chong & Treisman, 2003; Corbett & Oriet, 2011; Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008), orientation (Dakin & Watt, 1997; Miller & Sheldon, 1969; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), location (Alvarez & Oliva, 2008) or velocity (Atchley & Andersen, 1995; Rosenholtz, 1999; Watamaniuk & Duchon, 1992). In the present study, we set out to explore possible limits to feature averaging abilities.

Several characteristics of feature averaging suggest that it may be functionally distinct from individual object processing, and can proceed without complete processing of each object in the group (Corbett & Oriet, 2011). First, feature averaging appears to act across multiple objects in parallel rather than serially on each object. Chong and Treisman (2005a) found that estimates of average size were more accurate when objects were presented simultaneously rather than successively, even when total exposure time was matched, suggesting that parallel processing facilitates feature averaging. This suggestion is further supported by their finding that average size estimates for groups of simultaneously presented circles were more accurate when observers performed concurrent tasks requiring distributed spatial attention (e.g., attending to a stimulus spanning the entire visual display) rather than focused attention (e.g., attending to a stimulus restricted to a small part of the visual display). In a separate study, Chong and Treisman (2005b; Experiment 3) had participants estimate the average size of one of two simultaneously presented groups of objects defined on the basis of color. Accuracy did not differ as a function of whether the target group was cued before or after stimulus presentation, suggesting that the average object size of both groups was computed in parallel (but see Emmanouil & Treisman, 2008). The suggestion that feature averaging occurs in parallel across multiple objects is further supported by findings that the accuracy of average feature estimates is quite stable as the number of objects in the group increases (Ariely, 2001; Chong & Treisman, 2005a; Haberman, Harp, & Whitney, 2009) and is not substantially affected by exposure duration (Chong & Treisman, 2003; but see Whiting & Oriet, 2011).

A second noteworthy characteristic of feature averaging is that it appears to be more resistant to the withdrawal of attentional resources than individual object processing. Alvarez and Oliva (2008) presented two differently colored groups of four moving dots, and had participants track one of the groups. At a random point in time, all dots disappeared and participants were cued to estimate either the location of a single dot or the centroid (average location) of all four dots from one of the groups. Participants' accuracy in locating individual dots was dramatically impaired for the untracked group, relative to the tracked group. In contrast, there was only a small accuracy cost in locating the centroid of the untracked group of dots relative to the tracked group. Subsequently, Alvarez and Oliva (2009) found that observers were much more likely to detect orientation changes in groups of unattended gratings that altered their overall pattern than changes – equivalent in magnitude at an individual grating level – that did not alter their overall pattern. Thus, withdrawing attention from the gratings impaired the orientation discrimination of individual gratings more than it impaired discrimination of orientation patterns averaged across the whole group of gratings. Further support for the claim that feature averaging is resistant to the withdrawal of high-level processing resources was provided by Joo, Shin, Chong and Blake (2009), who found that the accuracy of average size estimates is not compromised during the attentional blink

(Shapiro, Raymond, & Arnell, 1997), a brief temporal window following object identification during which identification of subsequently presented *individual* objects is compromised.

A third distinction between feature averaging and individual object processing is that feature averaging is resistant to crowding, whereby identification of individual parafoveal objects is dramatically impaired when featurally similar objects are present nearby (Levi, 2008). Parkes, Lund, Angelucci, Solomon and Morgan (2001) presented a peripheral tilted grating, surrounded by eight other gratings oriented horizontally. Although orientation discrimination for the central grating was markedly impaired (relative to when that grating was presented in isolation), estimates of the average orientation of all nine gratings were influenced by the central grating as much as they were by any of the others (see also Balas, Nakano, & Rosenholtz, 2009). A subsequent study by Bulakowski, Post and Whitney (2011) further strengthened the argument that feature averaging is not contingent on the integrity of individual object representations. They demonstrated that although crowding of a target bar's orientation was stronger in the upper- relative to lower-visual field (a common finding in crowding experiments; He, Cavanagh, & Intriligator, 1996), the influence of the target bar on average orientation estimates for the group as a whole did not vary between the upper and lower visual fields.

Taken together, the studies reviewed here suggest fundamental differences between feature averaging and individual object processing. These differences have been taken by some to imply the existence of specialized feature averaging mechanisms distinct from those involved in object recognition (Alvarez, 2011). It has been further speculated that feature averaging may occur very early in perceptual processing, prior to the stage at which individual objects can be identified (Chong & Treisman, 2003, 2005a, 2005b; Choo & Franconeri, 2010). Such a possibility fits well with feature integration theory (Bouvier & Treisman, 2010; Treisman, 1996; Treisman & Gelade, 1980), which argues that individual object recognition requires two separate stages of processing. In the first stage, an initial feedforward sweep of activity through the visual system separately registers the features (e.g., color, shape, location) of all objects in the observer's field of view. According to feature integration theory, these features are not bound together to form complete representations of each object, available to conscious awareness, until a second stage involving reentrant feedback between higher and lower visual areas (also referred to as recurrent processing; see also Di Lollo, Enns, & Rensink, 2000; Lamme, 2000, 2010; Lamme & Roelfsema, 2000; Spratling & Johnson, 2004). If feature averaging only requires unbound featural information available via an initial feedforward sweep, then feature averaging may occur without recurrent processing, unlike individual object recognition.

One paradigm that has been used to investigate the locus at which feature integration occurs for individual objects is object substitution masking (OSM; Di Lollo, et al., 2000). In a typical OSM experiment, a group of objects is presented briefly, one of which is located within (but not obscured by) four dots, and participants are required to identify some feature of the object within the dots. This task is easy if the four dots offset at the same time as the object. In contrast, if the four dots remain visible longer than the object, the task becomes much more difficult (e.g., Di Lollo, et al., 2000; Dux, Visser, Goodhew, & Lipp, 2010; Enns, 2004; Koivisto & Silvanto, 2011). Di Lollo and colleagues (2000; see also Enns, 2004) put forward a prominent theoretical account of OSM. According to this perspective, the initial feedforward sweep leads to the generation of a hypothesis of what is being viewed, and this hypothesis is then tested against the information present in early visual areas by means of reentrant feedback. If the object disappears prior to the completion of this reentrant feedback but the four dots remain visible, there will be a mismatch between the hypothesis and the featural information in early visual areas, and a new hypothesis about what is being viewed will substitute the original one. While other theoretical accounts of OSM exist (e.g., Francis

& Cho, 2007; Francis & Hermens, 2002), it is generally accepted that this form of masking acts at a stage subsequent to the initial registration of featural information (Chakravarthi & Cavanagh, 2009; Dux, et al., 2010).

Choo and Franconeri (2010) used OSM to examine the importance of relatively late object representations in size averaging by testing whether masking individual circles using OSM reduces their contribution to estimates of the average size of a group of circles. They reasoned that if size averaging only requires the information available from relatively early object representations, then OSM should have no effect on a circle's contribution to average size estimates. In contrast, if size averaging involves later processing stages, then OSM should interfere with a circle's influence on average size estimates. Choo and Franconeri found that masking two circles using OSM did not reduce the extent to which these circles contributed to estimates of average size.

Although the results of Choo and Franconeri provide some support for the resistance of feature averaging to OSM, caution must always be exercised when interpreting null results, particularly when there are trends in the direction predicted by the alternative hypothesis: In both experiments conducted by Choo and Franconeri (2010), the average influence of the two circles masked by OSM was smaller than the average influence of the same two circles when they were not masked by OSM, as would be predicted if OSM interfered with size averaging. In addition, in their study, the two masked circles could occupy any of the eight possible locations on any given trial. Thus, on some trials the two masks would have been adjacent to each other, in which case participants' attention might have been captured to a small region of the visual display, whereas on other trials the two masks would have been on opposite sides of the visual field, which may have facilitated a more distributed mode of attention. Given that the accuracy of average size estimates is influenced by the distribution of attention (Chong & Treisman, 2005a), this aspect of the paradigm employed by Choo and Franconeri was perhaps not optimal for examining the influence of OSM on size averaging, as the magnitude of any effect could have varied as a function of mask locations.

Additionally, even if OSM does not affect size averaging, it remains possible that estimating other features could be influenced by OSM. According to Myczek and Simons (2008; see also Simons & Myczek, 2008) estimating average size may draw upon different processing mechanisms than estimating other features, such as orientation. Although estimates of average orientation could plausibly be achieved by pooling across orientation-selective neurons in early visual cortex, no analogous mechanism exists for size averaging, as there are no regions of early visual cortex sensitive to specific object sizes. For this reason, it is possible that the influence of OSM on feature averaging varies with the particular feature being averaged.

In the present study, we revisited the hypothesis that feature averaging is resistant to OSM using a protocol designed to avoid local attentional capture by the masks. We also investigated the generality of any effects by examining orientation averaging as well as size averaging. In Experiment 1, we examined whether masking circles using OSM reduced their contribution to average size estimates, in a paradigm similar to that used by Choo and Franconeri (2010). Importantly, we introduced the constraint that the two masks would always appear at opposite locations, thus avoiding local capture and encouraging a more diffuse distribution of attention across trials. In Experiment 2, we used a similar paradigm to examine whether OSM influences estimates of the average orientation of a group of Gabor patches. This allowed us to test whether any effect of OSM on feature averaging is consistent between features that are selectively processed in early visual cortex (i.e., orientation) and features that are not (i.e., size). To anticipate, contrary to the study of Choo and Franconeri, our findings suggest that OSM interferes with estimates of both average size and average orientation.

Experiment 1

Method

Participants. Eighteen undergraduates at The University of Queensland participated in Experiment 1 for course credit. All participants reported normal or corrected-to-normal vision. All procedures were conducted in accordance with the principles expressed in the Declaration of Helsinki, and were approved by The University of Queensland Ethics Committee.

Stimuli and Apparatus. Visual stimuli were presented on a 21-inch CRT monitor (NEC, Accusync 120) at a screen resolution of 1024 × 768 pixels and a refresh rate of 100 Hz. All visual stimuli were superimposed on a mid-gray background (RGB coordinates 127, 127, 127). Participants were seated at a viewing distance of approximately 60 cm from the monitor. Stimulus presentation and response recording were controlled using Cogent software (Cogent 2000 toolbox: FIL, ICN, and Wellcome Department of Imaging Neuroscience) in Matlab version 7.8 (www.mathworks.com), running on a desktop computer.

Each participant completed two tasks (see Figure 1), identical to those used by Choo and Franconeri (2010; Experiment 2). One required participants to judge the size of an individual circle, and the other required them to judge the average size of a group of circles. Trials in both tasks involved the presentation of dark gray (RGB coordinates 64, 64, 64) hollow circles (line thickness 0.1°) with center points at one of eight locations equally spaced on an imaginary circle 10° in diameter. A dark gray fixation cross 0.5° wide and high was presented at the center of the imaginary circle. Masks consisted of four red (RGB coordinates 255, 0, 0) circular dots 0.4° in diameter, located at the corners of an imaginary square 2.9° wide, concentric with the hollow circle being masked.

Procedure. Figure 1A illustrates the procedure for the object size judgment task in Experiment 1. Each trial began with a fixation cross for a random time interval of 1,500 – 2,500 ms. In the object size judgment task, eight circles were then presented for 30 ms. Participants were required to identify the size of the circle located within the single four-dot mask present on each trial. The mask either offset simultaneously with the circles (simultaneous offset mask) or remained on screen for an additional 320 ms (delayed offset mask). The circle within the mask was either 0.9° or 2.4° in diameter (small or large circle, respectively). The remaining seven circles present on each trial all had a diameter of 1.8°. Participants reported whether the target circle was large or small by pressing the up or down arrow key, respectively, on a standard keyboard. After completing 6 practice trials, participants completed 112 trials for each of the four conditions created by the crossed factors of target circle size (small, large) and mask offset (simultaneous, delayed). The target circle appeared at each of the eight possible locations an equal number of times for each condition. Participants were encouraged to take rest breaks every 64 trials to avoid fatigue, and were not given feedback about their performance (Bauer, 2009).

In the average size judgment task (Figure 1B), each set of circles was accompanied by two four-dot masks. The sizes of the six unmasked circles were randomly chosen with replacement from six possible sizes (0.9°, 1.2°, 1.5°, 1.8°, 2.1°, and 2.4°) in each trial. Within the masks, there could either be no circles (mask only), circles 0.9° in diameter (small circles), or circles 2.4° in diameter (large circles). Participants were required to identify the average size of all circles present in each trial. As in the object size judgment task, the masks either offset simultaneously with the circles or remained on screen for an additional 320 ms. In contrast to the procedure employed by Choo and Franconeri (2010), the two masks in the present study always occupied diagonally opposite locations on the imaginary circle. The masks appeared at each of the four possible opposite pairs of locations an equal number of times for each condition. A single probe circle was presented at fixation 320 ms after the offset of the circles, and participants had to report whether this probe circle was larger or smaller than the average of the circles they had just viewed by pressing the up or down arrow key, respectively, on a standard keyboard. After completing 12 practice trials, participants completed 120 trials for each of the six conditions created by the crossed factors of masked circle type (mask only, small circles, and large circles) and mask offset (simultaneous, delayed). Participants took rest breaks every 72 trials, and were not given feedback about their performance.

The size of the probe circle in the average size judgment task was calculated as a proportion of the average of the six unmasked circles present in each trial. This proportion was adjusted (separately for each of the six conditions) using the same staircase procedure employed by Choo and Franconeri (2010). Whenever the participant reported that the probe circle was larger than the average size of the circles they had just viewed, the scaling factor for the probe circle in the next trial for that condition was decreased by 3%. Whenever the participant reported that the probe circle was smaller than the average circle size, the scaling factor for that condition was increased by 3%. The initial value of the scaling factor for each condition was either 20% smaller or 20% larger than the average size of the six unmasked circles, and was counterbalanced (between participants) across the three levels of the masked circle type factor (AAB, ABA, ABB, BAA, BAB, BBA). The average scaling factor for the probe circle across the last 12 staircase reversal trials for each condition was defined as the point of subjective equality (PSE) for that condition, and used as a measure of the perceived average circle size – relative to the size of the six unmasked circles – in each condition. As in Choo and Franconeri (2010; Experiment 2), all participants completed the average size judgment task before completing the object size judgment task. Both tasks were completed in the same session, which lasted approximately 90 minutes.

Results and Discussion

Data from one participant were excluded from all statistical analyses as she did not follow task instructions. All statistical tests were conducted with a two-tailed alpha level of .05. Mauchly's Test of Sphericity was applied to all within-subjects *F* tests. Greenhouse-Geisser epsilon adjustments were made to degrees of freedom for these *F* tests wherever the assumption of sphericity was untenable. Unadjusted degrees of freedom are reported for all *F* tests.

Object size judgment task. Mean accuracy scores for each condition in the object size judgment task are displayed in Figure 2. Accuracy was roughly equivalent between target circle sizes, and impaired in the delayed- relative to simultaneous-mask offset conditions. To test this statistically, accuracy scores were subjected to a 2 × 2 within-subjects ANOVA with factors of target circle size (small, large) and mask offset (simultaneous, delayed). A significant main effect of mask offset, $F(1,16) = 48.56$, $p < .001$, $\eta_p^2 = .752$, indicated object size identification was impaired in the delayed ($M = 81.75\%$, $SE = 2.83\%$) relative to simultaneous mask condition ($M = 93.72\%$, $SE = 1.52\%$). There was no main effect of target circle size and no interaction between the factors (*F*s < 1).

Accuracies in the object size judgement task were also converted to d' values for each mask offset by defining correct responses to the large target circles as hits, and incorrect responses to the small target circles as false alarms. Accuracies of 0 or 100% were adjusted to 0.833% (1/*N trials*) or 99.167% ([*N trials*-1]/*N trials*), respectively, to avoid infinite d' values. A within-subjects *t* test, $t(16) = 9.92$, $p < .001$, revealed that sensitivity was significantly reduced in the delayed mask condition ($M = 2.05$, $SE = 0.21$) relative to the simultaneous mask condition ($M = 3.40$, $SE = 0.21$). Despite the significant performance impairment in the delayed mask condition, sensitivity was significantly greater than zero in both the simultaneous and delayed mask conditions [simultaneous, $t(16) = 16.11$, $p < .001$; delayed, $t(16) = 9.58$, $p < .001$].

Average size judgment task. Mean PSEs for each condition in the average size judgment task are displayed in Figure 3A. In the simultaneous mask offset conditions (white bars in Figure 3A; when no OSM should have occurred), perceived average size was roughly equivalent to the average size of the six unmasked circles when accompanied by the masks alone, smaller when accompanied by two additional small circles, and larger when accompanied by two additional large circles. This pattern suggests that, as expected, perceived average size was biased toward the size of the additional circles. To test whether OSM affected the influence of the additional circles on perceived average size, PSEs were subjected to a 3 × 2 within-subjects ANOVA with factors of masked circle type (mask only, small circles, and large circles) and mask offset (simultaneous, delayed). Significant main effects of masked circle type, $F(2,32) = 28.39$, $\varepsilon = .70$, $p < .001$, $\eta_p^2 = .640$, and mask offset, $F(1,16) = 6.13$, $p = .025$, $\eta_p^2 = .277$, were qualified by a significant interaction between the factors, $F(2,32) = 8.66$, $p < .001$, $\eta_p^2 = .351$. The interaction was followed up with within-subjects *t*-tests between simultaneous and delayed offset masks for each masked circle type. There was a significant effect of mask offset when the masked circles were small, $t(16) = 3.68$, $p = .002$, such that PSEs were smaller in the simultaneous mask condition ($M = -3.90\%$, $SE = 4.28\%$) than in the delayed mask condition ($M = 3.29\%$, $SE = 4.71\%$). When no masked circles were present, there was a trend for PSEs to be larger in the delayed mask condition ($M = 5.85\%$, $SE = 4.47\%$) relative to the simultaneous mask condition ($M = 1.46\%$, $SE = 4.44\%$, $t(16) = 2.05$, $p = .057$). PSEs did not differ as a function of mask offset when the masked circles were large, $t(16) = 0.51$, $p = .616$.

On first inspection, these results seem to indicate that OSM reduced the biasing of perceived average circle size caused by the small masked circles, but had no effect on the bias caused by the large masked circles. This stands in contrast to the finding that accuracy in the object size judgment task was impaired to a similar extent for large and small circles. However, an additional consideration in the case of the average size judgment task is that the delayed masks themselves might have biased perceived average circle size to be larger. This possibility is supported by the observation that PSEs were larger in the delayed- relative to simultaneous-offset mask conditions, even when no additional circles were present. This effect may have worked against a reduction in the effect of the additional large circles on PSEs in the delayed offset/ large circle condition, leaving the overall PSE seemingly unchanged (relative to the simultaneous offset/ large circle condition). This issue was also raised by Choo and Franconeri (2010), who suggested that their participants might have incorporated the size of the imaginary squares created by the four-dot masks into their average size judgments.

To allow an analysis of the PSE data uncontaminated by any influence of the masks themselves, we followed Choo and Franconeri (2010) and created PSE difference scores by subtracting each participant’s PSE in the mask only condition from their PSEs in the large and small masked circle conditions, separately for the two mask offsets (see Figure 3B). Inspection of Figure 3B suggests that PSE difference scores were more biased towards the size of the masked circles in the delayed offset conditions relative to simultaneous offset conditions for both large and small masked circles. To confirm this statistically, PSE difference scores were subjected to a 2 × 2 within-subjects ANOVA with factors of masked circle type (small, large) and mask offset (simultaneous, delayed). A significant main effect of masked circle type, $F(1,16) = 32.54$, $p < .001$, $\eta_p^2 = .670$, was qualified by a significant interaction between the two factors, $F(1,16) = 13.27$, $p = .002$, $\eta_p^2 = .453$. The interaction was followed up with within-subjects *t*-tests between simultaneous and delayed offset masks, separately for the large and small masked circles. The simple effect of mask offset was significant for large masked circles, *t*(16) = 2.51, *p* = .023, such that PSE difference scores were less positively biased in the delayed offset mask condition (*M* = 7.68%, *SE* = 2.06%) than in the simultaneous offset mask condition (*M* = 12.75%, *SE* = 2.01%). Despite this reduction in bias, PSE difference scores were still significantly greater than zero in the both the simultaneous and delayed offset/ large circle conditions, [simultaneous, *t*(16) = 6.33, *p* < .001; delayed, *t*(16) = 3.73, *p* = .002]. The simple effect of mask offset for small masked circles approached significance, *t*(16) = 1.86, *p* = .082, but in the opposite direction, such that PSE difference scores were less *negatively* biased in the delayed offset mask condition (*M* = -2.56%, *SE* = 1.70%) than in the simultaneous offset mask condition (*M* = -5.35%, *SE* = 1.50%). Additionally, PSE difference scores were significantly smaller than zero in the simultaneous offset/ small circle condition, *t*(16) = 3.57, *p* = .003, but not in the delayed offset/ small circle condition, *t*(16) = 1.50, *p* = .152. Although the influence of the masked circles was reduced in the delayed- relative to simultaneous-mask offset conditions, their influence was not removed altogether: There was still a significant difference between the large and small masked circle conditions when mask offset was delayed, *t*(16) = 7.12, *p* < .001. This is perhaps not surprising, given that the masks used here were only partially effective in obscuring individual object sizes. (Recall that sensitivity in the object size judgment task was significantly greater than zero even in the delayed mask offset condition.)[1]

In contrast to the study by Choo and Franconeri (2010), our findings suggest that OSM compromises not only observers' ability to identify the size of individual objects, but also their ability to use this information to estimate the average size of a group of objects. It is

---

[1] Attempting to adjust the present paradigm to reduce individual object sensitivity down to zero in the delayed mask offset conditions may have yielded an even stronger effect of OSM on feature averaging. However, pilot testing indicated that completely masking an individual object's size using OSM is very difficult to achieve, if possible at all. Thus, to maintain continuity with the study by Choo and Franconeri (2010), we decided to use the same stimulus parameters as they did.

interesting to note that the same general pattern of results was observed by Choo and Franconeri (2010; see Figure 5B), but in their study the trends were not statistically significant. As mentioned in the General Introduction, the discrepancy between the present findings and those of Choo and Franconeri could potentially be related to the single methodological difference between the two studies. Unlike Choo and Franconeri, we constrained the two masks in the average judgment task to always appear at opposite locations. In Choo and Franconeri's study, participants' attention may have been captured to a small region of the visual display on some trials but not others. Given evidence that distribution of attention influences size averaging (Chong & Treisman, 2005a), this variability could have added noise to the average size estimates made in Choo and Franconeri's study, which in turn could explain their failure to find a significant effect of OSM on size averaging.

It is possible – both in the present study and that of Choo and Franconeri – that the mere presence of the masks encouraged the use of different mean estimation strategies than would normally be used, had the masks not been present. At the extreme, capture of attention by the masks may have caused participants to estimate the average circle size based solely on the size of one of the two masked circles, rather than using the information from all circles present (for discussion, see Chong, Joo, Emmanouil, & Treisman, 2008; de Fockert & Marchant, 2008; Myczek & Simons, 2008; Simons & Myczek, 2008). If our participants had been estimating average size solely on the basis of the size of one of the masked circles, our observed influence of OSM on these estimates could essentially be explained as an effect of OSM on individual object judgments, and would provide no information about the influence of OSM on size averaging. To test whether such a strategy might have been adopted by our participants, we performed two additional analyses on the response data from the average size judgment task.

For the first additional analysis, we considered only trials in which an estimate based solely on *masked circle size* should have led to a different response than an estimate based on the *average of all circles* (i.e., trials in which the probe circle was smaller than the two masked circles, but larger than the average of all eight circles, or vice-versa). (Note that we excluded trials in which no masked circles were present in both this and the next additional analysis, as it would obviously not have been possible to base average size estimates on masked circle size in these trials.) For this subset of trials, we compared the number of times participants responded in the direction predicted if their judgments had been based solely on masked circle size to the number of times they responded in the direction predicted if their judgments had been based on the average of all eight circles. A within-subjects *t* test, $t(16) = 2.59$, $p = .020$, revealed that participants responded in accordance with an estimate based on *all circles* significantly more often ($M = 89.30$, $SE = 6.48$) than they did in accordance with an estimate based *solely on masked circle size* ($M = 73.00$, $SE = 6.72$).[2]

To further test the possibility that participants were estimating average size based solely on the size of one of the masked circles, we performed a series of hierarchical logistic regressions on each participant's response data. The binary outcome variable was whether the participant reported that the probe was larger (1) or smaller (0) than the average circle size on each trial. The two continuous predictors were: (1) the size difference between the probe circle and the two masked circles (masked-circle size deviation), and (2) the size difference between the probe circle and the average of the six unmasked circles (unmasked-circle size deviation). In Block 1 of the regression, we entered masked-circle size deviation. In Block 2, we entered unmasked-circle size deviation. If participants had estimated average size based

---

[2] Note that this comparison was conducted on the actual numbers of trials, rather than on percentages, to account for between-subjects variability in the number of trials in the analysed subset. The comparison remains statistically significant in the same direction if percentages are compared instead.

solely on masked circle size, the predictive power of the model should not improve when unmasked circle size deviation is entered into the model. Against this possibility, Block 2 χ^2 was statistically significant (*p*s $\leq .007$) in all except one ($p = .138$) of the 17 participants. It is worth pointing out that Block 1 χ^2 for this participant was also not statistically significant ($p = .365$), which is not consistent with the assumption that this participant estimated average size based solely on masked circle size. Additionally, removing this participant from the main analyses had no impact on the overall pattern or statistical significance of the results, suggesting that this participant alone did not drive the observed effect of OSM on size averaging. This finding further supports our argument that participants incorporated the size of all eight circles into their average size estimates, rather than estimating average size solely on the basis of the size of one of the masked circles.

Experiment 2

In Experiment 2, we sought to further test the hypothesis that average feature estimation is resistant to OSM, now using the feature dimension of orientation. This also allowed us to test the generalizability of the findings in Experiment 1. Current evidence suggests that visual cortical neurons show preferences for orientation, color and motion, but not for size (Myczek & Simons, 2008). Therefore, size averaging likely recruits different neural circuits than those for averaging other features registered by feature-selective neurons in early visual cortex (such as orientation). Consequently, the influence of OSM on feature averaging may depend on the particular visual feature in question. A second reason for examining the influence of OSM on average orientation estimates was that average orientation estimates should not be biased one way or another by the delayed offset masks, unlike estimates of average size.

Method

Participants. Eighteen undergraduates at The University of Queensland participated in Experiment 2 for course credit. No participant took part in both experiments. All participants reported normal or corrected-to-normal vision. All experimental procedures were conducted in accordance with the principles expressed in the Declaration of Helsinki, and were approved by The University of Queensland Ethics Committee.

Stimuli and Apparatus. Stimulus delivery and apparatus were identical to Experiment 1, except that the hollow circles were replaced by Gabor patches (diameter = 2.4°, spatial frequency = 0.83 cycles per degree, peak contrast = 60%). The mean luminance of the Gabors was set to dark gray (RGB coordinates 78, 78, 78) on the basis of pilot testing indicating that the orientation of Gabors with a mean luminance equal to that of the background was not effectively masked by the four dots, even at very low peak contrasts.

Procedure. The procedure for Experiment 2 (illustrated in Figure 4) was similar to that of Experiment 1, except that rather than making judgments on groups of differently sized circles, participants judged groups of differently oriented Gabors. In the object orientation judgment task (Figure 4A), participants were required to identify whether the Gabor located within the four-dot mask was tilted (20°) counter-clockwise (left) or clockwise (right) of vertical by pressing the left or right arrow key, respectively. The remaining seven Gabors were all oriented vertically. As in Experiment 1, masks either offset simultaneously with the Gabors (simultaneous mask) or remained on screen for an additional 320 ms (delayed mask). After completing six practice trials, participants completed 96 trials for each of the four conditions created by the crossed factors of target orientation (right, left) and mask offset (simultaneous, delayed). The target Gabor appeared at each of the eight possible locations an equal number of times per condition. Participants rested every 64 trials, and were not given performance feedback.

In the average orientation judgment task (Figure 4B), each set of eight Gabors was accompanied by two four-dot masks. The orientations of the six unmasked Gabors were chosen randomly with replacement from six possible orientations (-10°, -6°, -2°, 2°, 6°, or 10° relative to vertical). The two Gabors within the masks were rotated 20° either rightward (clockwise) or leftward (counter-clockwise) relative to the average orientation of the six unmasked Gabors on each trial. Participants were required to identify the average orientation of all eight Gabors on each trial. The two masks always occupied diagonally opposite locations, and either offset simultaneously with the Gabors (simultaneous offset mask) or remained on screen for an additional 320 ms (delayed offset mask). A probe Gabor was presented at fixation 320 ms after the offset of the set of eight Gabors, and participants reported whether this probe Gabor was oriented more leftward or rightward than the average orientation of the Gabors they had just viewed using the left and right arrow keys. After completing 12 practice trials, participants completed 120 trials for each of the four conditions created by the crossed factors of masked orientation (rotated right, rotated left) and mask offset (simultaneous, delayed). Because we thought it unlikely that the delayed offset mask would have any systematic effect on average orientation estimates (as opposed to the size judgments of Experiment 1), we chose not to include a “mask only” condition in Experiment 2. The masks appeared at each of the four possible diagonally opposite pairs of locations an equal number of times for each condition. Participants rested every 48 trials, and were not given performance feedback.

The orientation of the probe Gabor was calculated as a variable number of degrees rotated from the average orientation of the six unmasked Gabors on every trial. This number was adjusted (separately for each of the four conditions) using a staircase procedure similar to that employed in Experiment 1. Whenever the participant reported that the probe Gabor was oriented more leftward than the average of the Gabors they had just viewed, the probe in the next trial for that condition was rotated 1° further to the right of the average of the six unmasked Gabors. Conversely, whenever the participant reported that the probe Gabor was oriented more rightward than the average, the probe in the next trial for that condition was rotated 1° further to the left. The initial value for the probe rotations was 20° leftward or rightward of the average of the six unmasked Gabors, counterbalanced between participants across the two levels of the masked orientation factor (AA, BB, AB, BA). The average probe rotation across the last 12 staircase reversals for each condition was defined as the point of subjective equality (PSE) for that condition, and used as a measure of the perceived average orientation (relative to the six unmasked Gabors) in each condition. As in Experiment 1, all participants completed the average judgment task before the object judgment task. Both tasks were completed in the same session, which lasted approximately 60 minutes.

Results

Data from one participant were excluded from all statistical analyses as she did not follow task instructions. Statistical tests were conducted using the same guidelines as employed in Experiment 1.

Object orientation judgment task. Mean accuracy scores for each condition in the object orientation judgment task are displayed in **Error! Reference source not found.**. Similar to Experiment 1, accuracy was roughly equivalent between target orientations, and impaired in the delayed- relative to simultaneous-mask offset conditions. To test this statistically, accuracy scores were subjected to a 2 × 2 within-subjects ANOVA with factors of target orientation (right, left) and mask offset (simultaneous, delayed). A significant main effect of mask offset, $F(1,16) = 45.05$, $p < .001$, $\eta_p^2 = .738$, indicated object orientation identification was impaired in the delayed ($M = 74.63\%$, $SE = 3.83\%$) relative to the simultaneous mask condition ($M = 84.62\%$, $SE = 3.63\%$). There was no main effect of target orientation and no interaction between the factors (Fs < 1).

Accuracy scores in the object orientation judgement task were also converted to *d′* values for each mask offset by defining correct responses to the left-oriented targets as hits, and incorrect responses to the right-oriented targets as false alarms. Accuracies of 0 or 100% were adjusted to 1.042% (1/*N trials*) or 98.958% ([*N trials*-1]/*N trials*), respectively, to avoid infinite *d′* values. A within-subjects *t* test, $t(16) = 6.53$, $p < .001$, revealed that sensitivity was significantly impaired in the delayed mask condition ($M = 1.60$, $SE = 0.30$) relative to the simultaneous mask condition ($M = 2.61$, $SE = 0.37$). Despite the significant performance impairment in the delayed mask condition, sensitivity was significantly greater than zero in both the simultaneous and delayed mask conditions [simultaneous, $t(16) = 7.01$, $p < .001$; delayed, $t(16) = 5.34$, $p < .001$].

Average orientation judgment task. Mean PSEs for each condition in the average orientation judgment task are displayed in **Error! Reference source not found.**. In the simultaneous offset mask conditions (white bars in **Error! Reference source not found.**; when no OSM should have occurred), perceived average orientation tended to be positive (i.e., rightward of the average orientation of the six unmasked Gabors) when the masked Gabors were oriented rightward of the average, and negative (i.e., leftward of the average orientation of the six unmasked Gabors) when the masked Gabors were oriented leftward of the average. This pattern suggests that, as expected, perceived average orientation was biased toward the orientation of the masked Gabors. Comparing these results to those in the delayed offset mask conditions (black bars in **Error! Reference source not found.**; when OSM should have occurred), the biasing influence of the masked Gabors is clearly less pronounced. To confirm this statistically, PSEs were subjected to a 2 × 2 within-subjects ANOVA with factors of masked orientation (right, left) and mask offset (simultaneous, delayed). A significant main effect of masked orientation, $F(1,16) = 15.31$, $p = .001$, $\eta_p^2 = .489$, was qualified by a significant interaction between the factors, $F(1,16) = 11.09$, $p = .004$, $\eta_p^2 =$

.409. The interaction was followed up with within-subjects *t*-tests between simultaneous and delayed offset masks for each masked orientation. The simple effect of mask offset was significant for left masked Gabors, *t*(16) = 3.11, *p* = .007, such that PSEs were less leftward of the average unmasked orientation in the delayed offset mask condition (*M* = -6.18°, *SE* = 1.80°) than in the simultaneous offset mask condition (*M* = -9.71°, *SE* = 2.13°). Despite this reduction in bias, PSEs were still significantly greater than zero in the both the simultaneous and delayed offset/ left Gabor conditions, [simultaneous, *t*(16) = 4.55, *p* < .001; delayed, *t*(16) = 3.43, *p* = .003]. The simple effect of mask offset for right masked Gabors approached significance, *t*(16) = 1.93, *p* = .072, but in the opposite direction, such that PSEs were less *rightward* of the average unmasked orientation in the delayed offset mask condition (*M* = 3.32°, *SE* = 2.41°) than in the simultaneous offset mask condition (*M* = 5.63°, *SE* = 2.06°). Additionally, PSEs were significantly rightward from zero in the simultaneous offset/ right Gabor condition, *t*(16) = 2.73, *p* = .015, but not in the delayed offset/ right Gabor condition, *t*(16) = 1.38, *p* = .187. As with Experiment 1, although the influence of the masked Gabors was reduced in the delayed- relative to simultaneous-mask offset conditions, their influence was not removed altogether. There was still a significant difference between the right and left masked Gabor conditions when mask offset was delayed, *t*(16) = 2.78, *p* = .013.

As with Experiment 1, we conducted additional analyses to test the possibility that participants had estimated average Gabor orientation based solely on the orientation of one of the masked Gabors. First, we examined only trials in which an estimate based solely on the orientation of the *masked* Gabors should have led to a different response than an estimate based on the *average of all* Gabors. On these trials, once again, participants responded in accordance with an estimate based on all Gabors significantly more often (*M* = 99.29, *SE* = 11.90) than they did in accordance with an estimate based solely on masked Gabor size (*M* = 76.65, *SE* = 11.16), *t*(16) = 2.77, *p* = .014. This finding supports our argument that participants incorporated the orientation of all eight Gabors into their average orientation estimates, rather than estimating average orientation based solely on the orientation of one of the masked Gabors.

We then performed a series of hierarchical binary logistic regressions on whether the participant reported that the probe Gabor was oriented more leftward or rightward than the average Gabor orientation on each trial. The two continuous predictors were the orientation difference between the probe Gabor and the two masked Gabors (masked Gabor deviation), and the orientation difference between the probe Gabor and the average of the six unmasked Gabors (unmasked Gabor deviation). We entered masked Gabor deviation in Block 1 of the regression, and unmasked Gabor deviation in Block 2. If participants had estimated average orientation based solely on masked Gabor orientation, the predictive power of the model should not have improved when unmasked Gabor deviation was entered in Block 2. Against this possibility, Block 2 χ^2 was statistically significant (ps $\leq .018$) in all except 3 of the 17 participants. Removing these participants from the main analyses had no impact on the overall pattern or statistical significance of the main results, suggesting that these participants alone did not drive the observed effect of OSM on orientation averaging.

General Discussion

In two experiments, we tested the hypothesis that feature averaging is resistant to masking by object substitution (Choo & Franconeri, 2010). Participants estimated the average size (Experiment 1) or orientation (Experiment 2) of a group of briefly presented objects, two of which were surrounded by masks. The masks could either offset simultaneously with the objects (no OSM) or after the objects (OSM). The dependent measure was the extent to which the two masked objects influenced average feature estimates. In both experiments, masked objects exerted significantly less influence on average feature estimates than objects that were not masked. Separate tasks in which participants judged individual masked objects confirmed that OSM also reduced accuracy in identifying the size or orientation of individual objects.

It is possible that we found a significant effect of OSM on average size estimates where Choo and Franconeri (2010) did not due to a methodological difference between the studies. Unlike the previous study, we chose to constrain the two masks to always appear at opposite locations, ensuring that participants' distribution of attention was uniform across trials. Choo and Franconeri's approach of allowing the two masks to vary randomly in their positions across trials may have led to variability in the manner in which participants' attention was distributed across the visual field. As attention is known to influence the accuracy of average size estimates (Chong & Treisman, 2005a), this variability might therefore have reduced the likelihood of them observing a reliable masking effect in their study.

The present findings suggest that OSM disrupts not only individual object processing (Di Lollo, et al., 2000; Enns & Di Lollo, 1997) but also estimates of average featural information across groups of objects. In light of evidence suggesting OSM leaves the initial registration of featural information essentially intact, and interferes primarily with later recurrent processing stages (Chakravarthi & Cavanagh, 2009; Chen & Treisman, 2009; Di Lollo, et al., 2000; Di Lollo, Enns, & Rensink, 2002; Enns, 2004; Woodman & Luck, 2003), the present results imply that feature averaging relies to some extent upon these relatively late stages of processing. We must bear in mind, however, that it has not been conclusively established that OSM *only* interferes with recurrent processing stages. As such, conclusive proof of the specific involvement of reentrant processing in feature averaging will require future research involving methods that can more directly separate feedforward from feedback stages of processing (e.g., using neurodisruption techniques such as transcranial magnetic stimulation, applied to different levels of the cortical visual system; Juan, Campana, & Walsh, 2004; Silvanto, Lavie, & Walsh, 2005).

Our finding that OSM interfered with estimates not only of average *size* but also average *orientation* rules out an important potential criticism of the average size result. According to Myczek and Simons (2008; see also Simons & Myczek, 2008), size averaging may be a fundamentally different process than averaging of other features – such as orientation, motion or spatial frequency – that are specifically encoded in early visual cortex. From this account, size averaging might rely more heavily on the late processing stages affected by OSM than orientation averaging does. Therefore, it could have been the case that OSM interfered with size averaging but not orientation averaging. Our findings contradict this possibility, and suggest that OSM interferes with feature averaging even for features encoded in early visual cortex.

Other recent research further supports the notion that feature averaging may not be as rapid and automatic as previously suggested. Previous studies used to support the automaticity of feature averaging have found that average feature estimates are accurate even at very short presentation durations (50 ms; Chong & Treisman, 2003), too short for effortful processing of each individual item (Wolfe, 1998). As pointed out by Whiting and Oriet (2011), however, these studies failed to account for the possibility that processing continued after the objects had disappeared. To address this limitation, Whiting and Oriet (2011) tested observers’ ability to estimate average object size when prolonged processing of the objects was prevented using backward masking. Here, significant performance impairments were observed when the masks replaced the objects after 100 ms or less, suggesting that feature averaging benefits from additional processing beyond this initial time window. Further evidence against the automaticity of feature averaging comes from a study by de Fockert and

Marchant (2008), who found that cueing observers to a single object in a group caused average size estimates to be biased toward the size of the cued object.

It is worth pointing out that, although the present findings suggest a role for reentrant feedback in feature averaging, this does not necessarily contradict previous findings that average feature estimates are more resistant to the withdrawal of attentional resources than individual object feature estimates (e.g., Alvarez & Oliva, 2008, 2009; Joo, et al., 2009). According to Alvarez (2011; see also Alvarez & Oliva, 2008; 2009), the apparent accuracy of average feature estimates – relative to individual object feature estimates – could simply be due to the *power of averaging:* Averaging multiple noisy measurements will yield an estimated value that is more precise than the individual measurements are because uncorrelated random errors will tend to cancel out. Thus, estimates of the average size of a group of objects should be more accurate than estimates of the size of any one individual object, even if the same information is used in each case. For the same reason, increasing the noise in each measurement (as might occur under conditions of reduced attention) should have less of an impact on average estimates than on individual estimates, even if the same information is used. This represents a plausible explanation for why average feature estimates could appear to be more resistant to the withdrawal of attentional resources than individual object processing without the requirement for assuming that separate, purely feedforward mechanisms exist for feature averaging.

To summarize, the present study has demonstrated that masking objects using OSM significantly impairs their contribution to estimates of both average size and average orientation. These findings suggest that, rather than only requiring the unbound featural information available in the initial feedforward sweep, feature averaging may benefit from prolonged recurrent processing of each individual object within the group. This research adds to the growing body of evidence suggesting that feature averaging is not as automatic as previously speculated, and may not rely on different mechanisms than individual object processing after all.

Author Note

This work was supported by Australian Postgraduate Award and UQ Advantage scholarships to OJ, and an Australian Research Council grant (DP110102925) to JBM. We thank Paul Dux and David Painter for insightful comments on previous versions of the manuscript.

References

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., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*, 1-10. doi: 10.1167/7.13.14

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 of the United States of America, 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

Atchley, P., & Andersen, G. J. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research, 35*, 3131-3144. doi: 10.1016/0042-6989(95)00057-7

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*, 1-18. doi: 10.1167/9.12.13

Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, & Psychophysics, 71*, 655-665. doi: 10.3758/app.71.3.655

Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science, 21*, 200-204. doi: 10.1177/0956797609357858

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*, 384-392. doi: 10.1177/0956797610397956

Bulakowski, P. F., Post, R. B., & Whitney, D. (2011). Reexamining the possible benefits of visual crowding: Dissociating crowding from ensemble percepts. *Attention, Perception, & Psychophysics, 73,* 1003-1009. doi: 10.3758/s13414-010-0086-2

Chakravarthi, R., & Cavanagh, P. (2009). Recovery of a crowded object by masking the flankers: Determining the locus of feature integration. *Journal of Vision, 9,* 4.1-9. doi: 10.1167/9.10.4

Chen, Z., & Treisman, A. (2009). Implicit perception and level of processing in object-substitution masking. *Psychological Science, 20,* 560-567. doi: 10.1111/j.1467-9280.2009.02328.x

Chong, S. C., Joo, S. J., Emmanouil, T. A., & 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. doi: 10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics, 67,* 1-13. doi: 10.3758/BF03195009

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. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica, 138*, 289-301. doi: 10.1016/j.actpsy.2011.08.002

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42-45.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37,* 3181-3192. doi: 10.1016/S0042-6989(97)00133-8

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

Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia, 46,* 2861-2864. doi: 10.1016/j.neuropsychologia.2008.05.014

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481-507. doi: 10.1037/0096-3445.129.4.481

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2002). Object substitution without reentry? *Journal of Experimental Psychology: General, 131*, 594-596. doi: 10.1037/0096-3445.131.4.594

Dux, P. E., Visser, T. A. W., Goodhew, S. C., & Lipp, O. V. (2010). Delayed reentrant processing impairs visual awareness: An object-substitution-masking study. *Psychological Science, 21*, 1242-1247. doi: 10.1177/0956797610379866

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, *70*, 946-954. doi: 10.3758/pp.70.6.946

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*, 1321-1331. doi: 10.1016/j.visres.2003.10.024

Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science, 8,* 135-139. doi: 10.1111/j.1467-9280.1997.tb00696.x

Francis, G., & Cho, Y. S. (2007). Testing models of object substitution with backward masking. *Perception & Psychophysics, 69*, 263-275. doi: 10.3758/BF03193748

Francis, G., & Hermens, F. (2002). Comment on "Competition for Consciousness Among Visual Events: The Psychophysics of Reentrant Visual Processes" (Di Lollo, Enns, & Rensink, 2000). *Journal of Experimental Psychology: General, 131,* 590-593. doi: 10.1037/0096-3445.131.4.590

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected at once? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1003-1012. doi: 10.1037/0096-1523.33.5.1003

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*, 1-13. doi: 10.1167/9.11.1

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334-337. doi: 10.1038/383334a0

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

Juan, C. H., Campana, G., & Walsh, V. (2004). Cortical interactions in vision and awareness: Hierarchies in reverse. *Progress in Brain Research, 144*, 117-130. doi: 10.1016/S0079-6123(03)14408-1

Koivisto, M., & Silvanto, J. (2011). Relationship between visual binding, reentry and awareness. *Consciousness and Cognition, 20,* 1293–1303. doi: 10.1016/j.concog.2011.02.008

Lamme, V. A. F. (2000). Neural mechanisms of visual awareness: A linking proposition. *Brain and Mind, 1*, 385-406. doi: 10.1023/a:1011569019782

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience, 1*, 204-220. doi: 10.1080/17588921003731586

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*, 571-579. doi: 10.1016/S0166-2236(00)01657-X

Levi, D. M. (2008). Crowding - an essential bottleneck for object recognition: A mini-review. *Vision Research, 48*, 635-654. doi: 10.1016/j.visres.2007.12.009

Miller, A. L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology, 81*, 16-21. doi: 10.1037/h0027430

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

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 332-350. doi: 10.1037/0096-1523.16.2.332

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

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research, 39,* 3157-3163. doi: 10.1016/S0042-6989(99)00077-2

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences, 1*, 291-296. doi: 10.1016/S1364-6613(97)01094-2

Silvanto, J., Lavie, N., & Walsh, V. (2005). Double dissociation of V1 and V5/MT activity in visual awareness. *Cerebral Cortex, 15,* 1736-1741. doi: 10.1093/cercor/bhi050

Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, *70,* 1335-1336. doi: 10.3758/PP.70.7.1335

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience, 16*, 219-237. doi: 10.1162/089892904322984526

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6,* 171-178. doi: 10.1016/S0959-4388(96)80070-5

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136. doi: 10.1016/0010-0285(80)90005-5

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research, 32,* 931-941. doi: 10.1016/0042-6989(92)90036-I

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

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

Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science, 14,* 605-611. doi: 10.1046/j.0956-7976.2003.psci_1472.x

Figure Captions

Figure 1. Schematic of the procedure for Experiment 1. (A) In the object size judgment task, participants reported whether the circle located within the red four-dot mask (displayed here in gray) was large or small. The four-dot mask either offset with the circles (simultaneous mask), or remained on screen for an additional 320 ms (delayed mask). (B) In the average size judgment task, participants reported whether the probe circle presented at the end of the trial was larger or smaller than the average size of the set of circles they had just viewed. Figure adapted from Figure 1 of “Objects with reduced visibility still contribute to size averaging,” by H. Choo and S. L. Franconeri, 2010, *Attention, Perception, & Psychophysics, 72,* p. 89. Copyright 2010 by The Psychonomic Society, Inc. Adapted with kind permission from Springer Science+Business Media B.V.

Figure 2. Accuracy for the object size judgment task across the conditions of target circle size and mask offset in Experiment 1. Error bars represent within-subjects standard errors of the means (Cousineau, 2005).

Figure 3. Results for the average size judgment task in Experiment 1. (A) Point of subjective equality (PSE) across the conditions of masked circle type and mask offset. (B) Differences in PSE between the mask only conditions and the large and small circle conditions. Error bars represent within-subjects standard errors of the means (Cousineau, 2005).

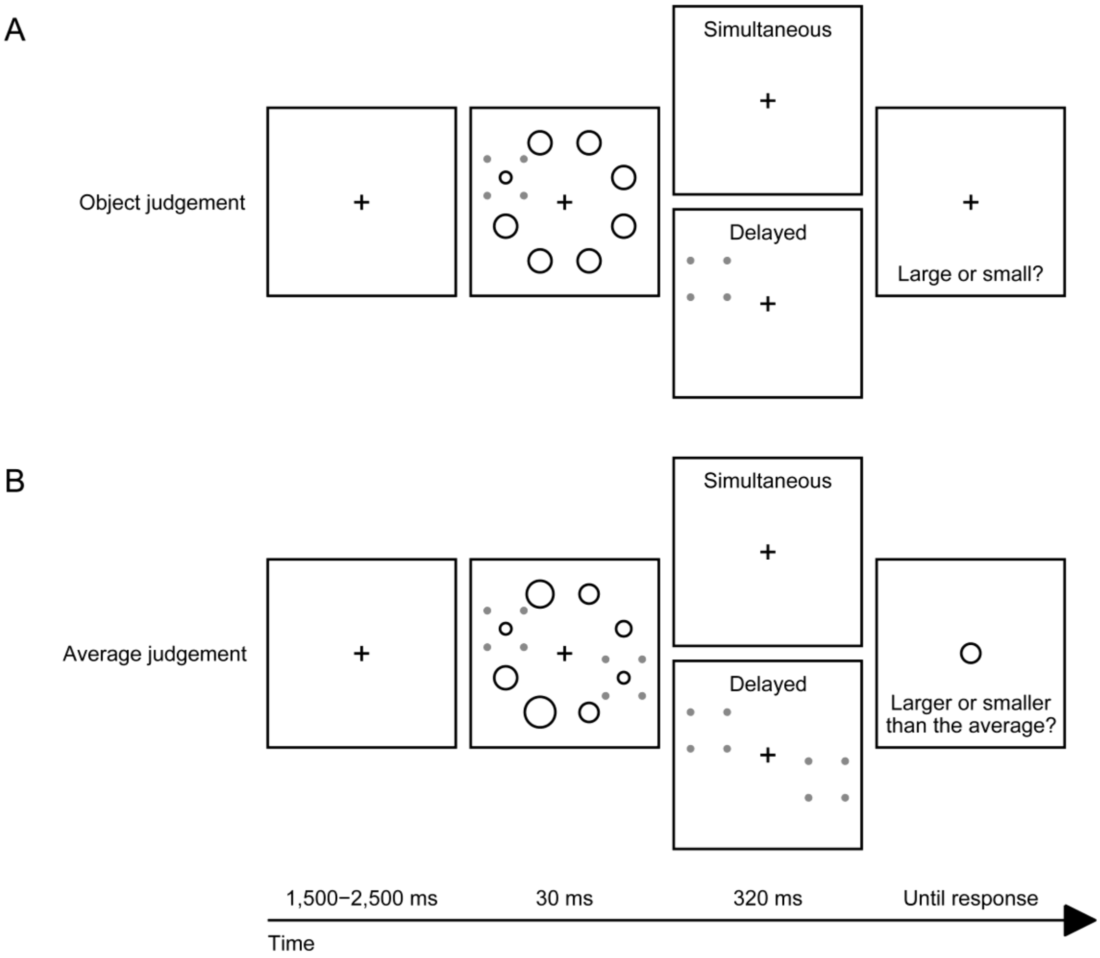
Figure 4. Schematic of the procedure for Experiment 2. (A) In the object orientation judgment task, participants reported whether the Gabor located within the red four-dot mask (displayed here in gray) was oriented to the left or right. (B) In the average orientation judgment task, participants reported whether the probe Gabor presented at the end of the trial

was oriented more leftward or more rightward than the average orientation of all the Gabors they had just viewed.

Figure 5. Accuracy for the object orientation judgment task across the conditions of target orientation and mask offset in Experiment 2. Error bars represent within-subjects standard errors of the means (Cousineau, 2005).

Figure 6. Point of subjective equality (PSE) for the average orientation judgment task across the conditions of masked orientation and mask offset in Experiment 2. Positive values are right of the average of the six unmasked Gabors; negative values are left of the average of the six unmasked Gabors. Error bars represent within-subjects standard errors of the means (Cousineau, 2005).

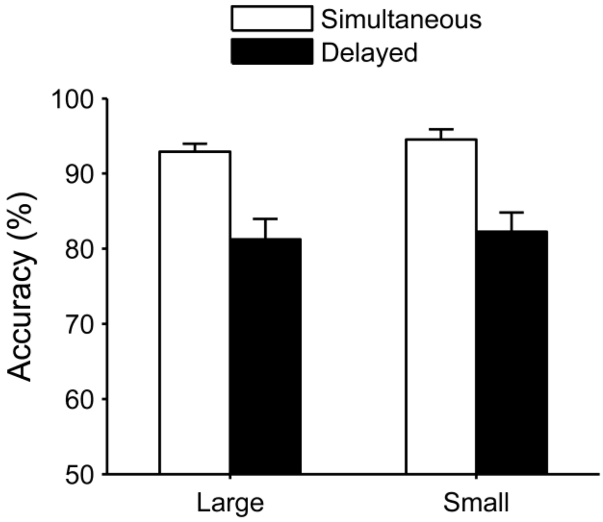
846 **Figure 1**



847

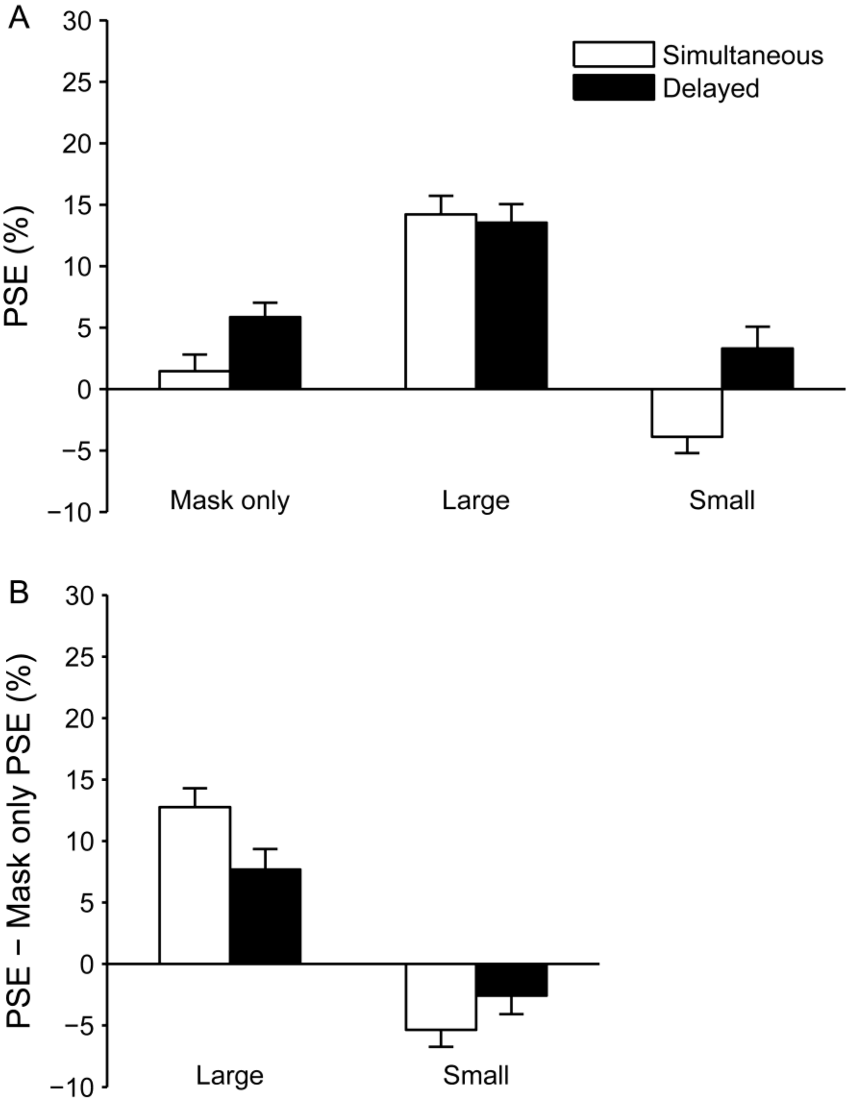
848

Figure 2


Simultaneous
Delayed
Accuracy (%)
100
90
80
70
60
50
Large
Small

852

Figure 3

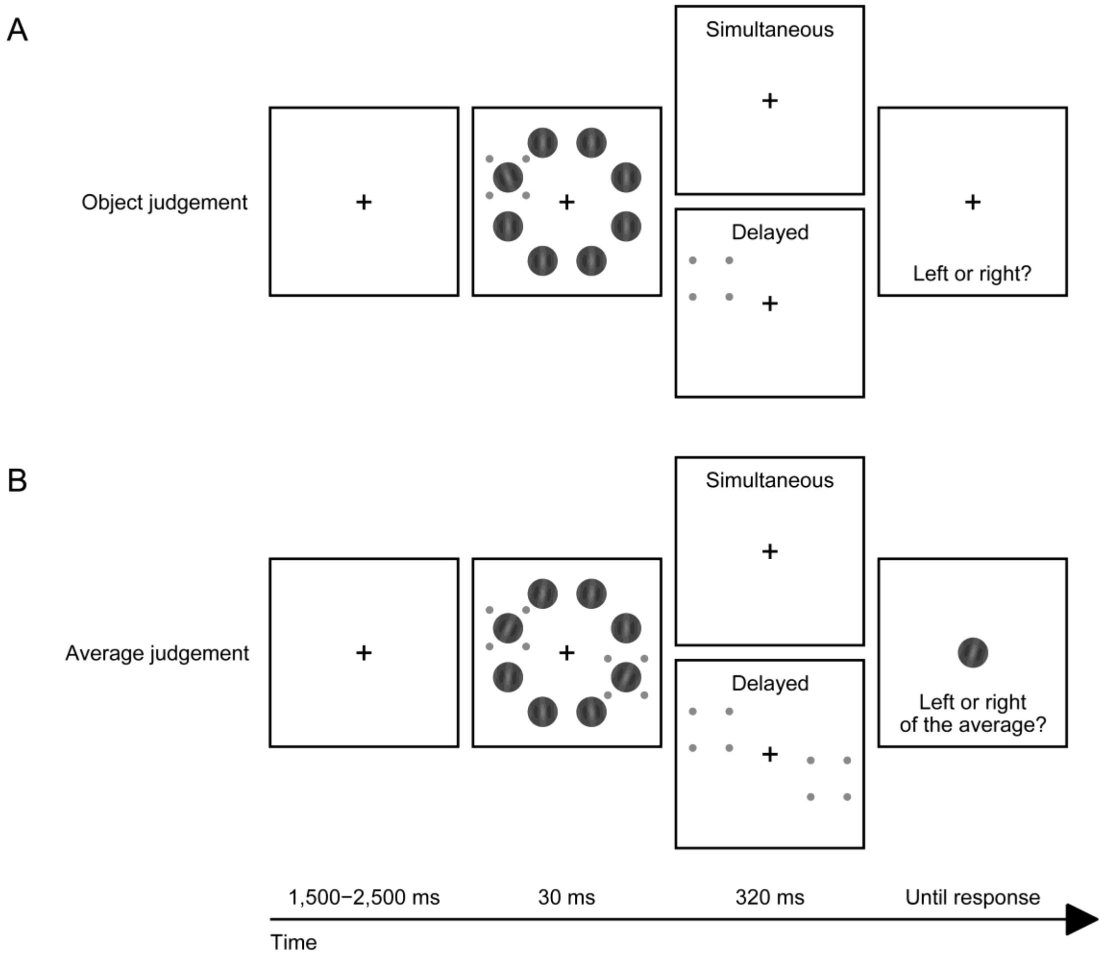


853

854

855

Figure 4



856

857

Figure 5

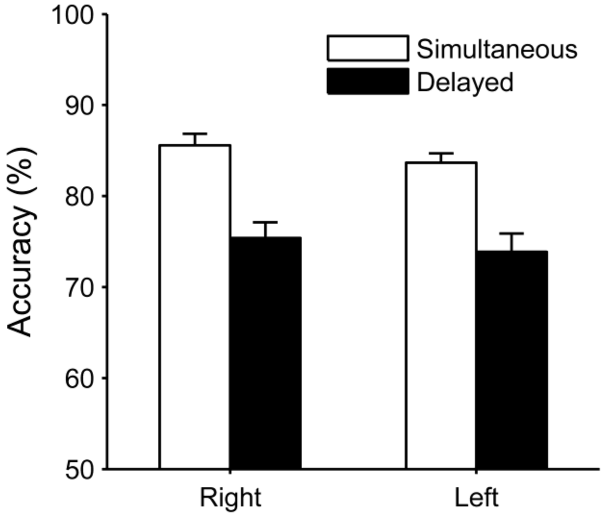


Figure 6