Acta Psychologica 138 (2011) 289–301

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation

Jennifer E. Corbett [a,*], Chris Oriet [b]

[a] Department of Psychology, New York University, United States
[b] Department of Psychology, University of Regina, Canada

ARTICLE INFO

Article history:
Received 25 May 2011
Received in revised form 7 August 2011
Accepted 8 August 2011
Available online 7 September 2011

PsycINFO codes:
2323 Visual Perception
2320 Sensory Perception
2340 Cognitive Processes

Keywords:
Perceptual averaging
Summary representation
Mean size
Set versus individual item representation
Attentional blink

ABSTRACT

We tested Ariely's (2001) proposal that the visual system represents the overall statistical properties of sets of objects against alternative accounts of rapid averaging involving sub-sampling strategies. In four experiments, observers could rapidly extract the mean size of a set of circles presented in an RSVP sequence, but could not reliably identify individual members. Experiment 1 contrasted performance on a member identification task with performance on a mean judgment task, and showed that the tasks could be dissociated based on whether the test probe was presented before or after the sequence, suggesting that member identification and mean judgment are subserved by different mechanisms. In Experiment 2, we confirmed that when given a choice between a probe corresponding to the mean size of the set and a foil corresponding to the mean of the smallest and largest items only, the former is preferred to the latter, even when observers are explicitly instructed to average only the smallest and largest items. Experiment 3 showed that a test item corresponding to the mean size of the set could be reliably discriminated from a foil but the largest item in the set, differing by an equivalent amount, could not. In Experiment 4, observers rejected test items dissimilar to the mean size of the set in a member identification task, favoring test items that corresponded to the mean of the set over items that were actually shown. These findings suggest that mean representation is accomplished without explicitly encoding individual items.

1. Introduction

The visual system must maximize the limited attentional resources available for processing salient events, while maintaining a general sense, or "gist," of the surrounding environment. This balance may be accomplished, in part, by rapidly summarizing sets of objects, representing properties such as average size (Ariely, 2001; Chong & Treisman, 2003, 2005a,b), orientation (Dakin & Watt, 1997; Parkes, Lund, Angelucci, & Morgan, 2001), and direction of motion (Watamaniuk, Sekuler, & Williams, 1989). Ariely (2001) proposed these summary representations exclude precise information about individual objects. However, Myczek and Simons (2008) argue that observers may instead sample only a few of the items in each set, and then base their mean judgments on the average of these sub-samples. Although their simulations provide compelling evidence in support of this proposal, to date there has been no direct test of whether observers actually represent individual items in estimating the average size of sets of items.

In the present work, we explicitly tested the proposal that sets are represented in a computationally different manner than individual objects. To do so, we examined whether the average size of a set of circles was rapidly extracted under conditions that impaired or eliminated the ability to recognize individual items. Beyond previous studies, we not only compared our subjects' performance in a mean judgment task against the simulated outcome of a simplified sub-sampling strategy, but also against their actual performance when explicitly instructed to use such a sub-sampling strategy, and against their performance in an identification task requiring the explicit representation of individual set members. Furthermore, our results provide direct confirmation that items in an RSVP sequence that observers could not identify were nonetheless included in their calculations of the set's average size on the same trial.

In Experiment 1, we compared observers' performance in a mean judgment task and a member identification task involving RSVP sequences of circles to determine whether the mean size could be recalled when the size of individual circles could not. We next compared our results in the mean judgment task to simulated performance using a simple sub-sampling strategy based on explicitly encoding the largest and smallest circles in each set (M_{SL}). Experiment 2 followed with a direct comparison of observers' actual performance when specifically instructed to use each strategy, i.e., to compute the mean size of the whole set (M_W) and the mean size of only the smallest and largest circles in the set (M_{SL}). In Experiment 3, when we alternated circles and shapes on each trial, participants were still able to perform the mean judgment task, but performed poorly in a simplified member identification task. Finally,

* Corresponding author at: New York University, Department of Psychology, 6 Washington Place, New York, NY 10003, United States. Tel.: +1 646 259 2789.
E-mail address: jennifer.e.corbett@gmail.com (J.E. Corbett).

0001-6918/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2011.08.002

the results of Experiment 4 confirmed that observers' summary representations incorporated information about the sizes of individual items that were masked from awareness and could not be reliably identified. Before describing our experiments in further detail, we first outline relevant previous findings regarding the statistical representation of sets, then summarize strategies proposed to account for these findings, and finally provide a brief outline of the RSVP methodology we employed to increase the likelihood that the individual circles comprising the sets in our experiments would not be explicitly represented.

1.1. Average representation

There is much evidence to suggest that basic properties of sets are summarized in a statistical representation that does not include precise information about the individual items. For example, Torralba and Oliva (2003) reported that observers were well above chance at detecting the presence of an object from a specified category (e.g. animal, vehicle, or building) in a briefly presented image based only on categorically averaged second-order image statistics, such as luminance and roughness. Similarly, although observers could not discern the orientations of individual elements, Parkes et al. (2001) reported that they could perceive the average orientation of a whole set of Gabors. In addition, when viewing a set of moving dots, observers generally could not describe the motion of individual dots, but were quite accurate at estimating the average direction of motion of the entire set (Watamaniuk et al., 1989). Importantly, in each of these examples, the average property of the set could be computed by pooling inputs across individual local receptors without the need to explicitly encode the individual elements. The ability to compute average size, in contrast, is surprising because a corresponding mechanism for pooling the output of individual object size computations has yet to be identified (Myczek & Simons, 2008). Nevertheless, individual retinal neurons have been found that adapt rapidly to the spatial scale of images (e.g. Smirnakis, Berry, Warland, Bialek, & Meister, 1997), yielding a plausible candidate for such a mechanism. Average size calculation could be mediated by mechanisms beyond the retina as well. For example, perceived size appears to be represented in early visual cortex (Murray, Boyaci, & Kersten, 2006), and population coding of object size could also give rise to ensemble coding of average size. Irrespective of the mechanism involved, a large body of evidence suggests that, like orientation and direction of motion, statistical representations of average size can also be rapidly extracted from a set of items.

Ariely (2001) first reported that observers could not reliably determine whether a circle was a member of a set of circles presented immediately beforehand, but could construct a representation of the set that captured the mean size of the circles, regardless of whether the task oriented them to extract this property of the set. From these results, Ariely suggested that the visual system creates a statistical representation of the set that excludes precise information about the individual items. Similarly, Chong and Treisman (2003) reported that participants were able to compare which of two heterogeneous side-by-side displays of circles had the larger mean size as quickly and accurately as they could determine which of two displays of homogeneous circles had the larger mean size, or which of two single circles was larger. Neither the number nor the density of the circles affected the accuracy of mean size judgments, suggesting that computing mean size was at least as fast and accurate as comparing individual items. In addition, the mean size of a set of circles was encoded sufficiently to prime the perceived contrast of a subsequent test circle when its size corresponded to the mean size of the set, but no such priming was observed when the test circle was a member of a preceding set of circles (Marchant & de Fockert, 2009). Collectively, these findings are intriguing because they suggest that in forming visual representations, sets of objects are compressed into a summary representation that captures overall statistical properties of the entire set without retaining precise information about individual items.

1.2. Can the mean be represented when individual items cannot?

The above results suggest that different mechanisms are involved in representing sets versus individual objects. However, Myczek and Simons (2008) recently questioned whether mean size perception could be more parsimoniously explained by sub-sampling strategies assuming only a few of the items present on a given trial are represented in detail. Their simulations of Ariely's (2001) and Chong and Treisman's (2003, 2005a) experiments sampled only a few of the items on a given trial, yet yielded comparably accurate estimates of the set mean. Rosenholtz and Alvarez (2007) similarly argue that subjects' performance is qualitatively different than expected if the visual system computes the average of all display items, and is instead well modeled by a simple averaging strategy of sampling only a sub-set of the display. In summary, although these proposals do not rule out the possibility that the mean is extracted using a specialized averaging process, they argue that the existing data can also be explained by simpler focused attention mechanisms encoding the identities of only a few individual elements in each set.

Along these lines, it has been repeatedly suggested that equivalent performance could have been attained from simple sub-sampling strategies of comparing only a few circles closest to fixation (Myczek & Simons, 2008), comparing only the largest circle in each of two side-by-side displays (de Fockert & Marchant, 2008; Myczek & Simons, 2008), averaging only the smallest and largest circles present on each trial (Myczek & Simons, 2008), or switching strategies depending on the nature of individual displays (Myczek & Simons, 2008). Such conscious and deliberate strategies would be indicative of a process quite different from the rapid computational process previous research suggests is at work in set representation (e.g., Chong & Treisman, 2003, 2005a,b; Corbett, Oriet, & Rensink, 2006).

To our knowledge, Chong, Joo, Emmanouil, and Treisman (2008) conducted the only study to test directly whether the sub-sampling strategies proposed by Myczek and Simons (2008) can account for observers' performance when judging which of two side-by-side displays of circles had the larger mean size. In their Experiment 1, Chong et al. measured performance on trials for which accurate responses could be determined by: a) finding the largest circle on each side, b) averaging the smallest and largest circles on each side, and c) sampling only a few circles on each side. They then compared performance when these trial types were mixed within blocks versus when they were presented in separate blocks. If participants used a sub-sampling strategy of sampling only a few circles from each display, then mixed blocks would have required them to switch between these different strategies, leading to poorer performance for trials randomized within blocks than for trials blocked by display type. Instead, the results showed no difference in accuracy between mixed and blocked trials, supporting the proposal that participants used a consistent method of discriminating which side had the larger mean size regardless of the structure of the displays.

Although the results of Chong et al.'s (2008) Experiment 1 provide convincing evidence that participants were likely not switching between strategies from trial-to-trial, the conclusions of their remaining experiments are less clear. Rather than eliminating the possibility of using a sampling strategy, the experimenters attempted to show that a small sample of the circles presented near the fovea yielded no better performance than a sample of circles from random locations (Experiment 2), and that observers were not sampling one circle (the largest) from each side when determining which of two side-by-side sets had the larger mean size. As Simons and Myczek (2008) correctly point out, the fact that observers are not using *these* strategies does not preclude the possibility that they are using some other sampling strategy.

It would seem that rather than trying to identify and systematically exclude every possible sampling strategy that could be used to determine the correct response without averaging the whole set, a new approach is needed. Here we examined how mean size is represented when observers are explicitly instructed to use a sub-sampling strategy, and whether the mean of a property of a set can be represented when sub-sampling strategies are not possible. To accomplish this, instead of presenting displays of circles all at once, side-by-side, we presented RSVP sequences of individual circles under conditions known to impair or prevent the consolidation of individual items in visual short-term memory. In addition, to examine whether different processes are involved in representing the mean of a set of items and representing individual items, we compared performance in a mean judgment task to performance in a member identification task (see also, Ariely, 2001, 2008).

1.3. RSVP methodology

Potter and Levy (1969) first reported that observers were poor at recognizing individual pictures presented for around 100 ms in RSVP sequences. However, either previewing the target picture, or providing its category name in advance of the stream was sufficient to allow them to report its presence about 65–70% of the time (Potter, 1976). In a related line of research, when Weichselgartner and Sperling (1987) presented subjects with RSVP streams of digits, each for 100 ms, and asked them to identify a given target digit plus the three digits that immediately followed it, subjects typically could name only the target and perhaps the next digit, but rarely reported items appearing in the interval between 100 and 300 ms after the target. Similarly, Dell' Acqua, Jolicoeur, Luria, and Pluchino (2009) reported that when one, two, or three digit targets were embedded among letter distractors in RSVP sequences with items presented at a rate of 84 ms each with 0 ms ISIs, targets that followed the first target were rarely correctly identified. These findings suggest that little can be recalled about a single item from an RSVP sequence if the target is not specified in advance. Taken together, these results demonstrate a transient impairment in encoding masked items, such that recognizing the first target impairs the recognition of items that follow it in close temporal succession.

There been a handful of previous attempts to demonstrate extraction of the mean of a property of a set of items presented in RSVP sequences. However, these studies have either: a) used longer stimulus presentation times (Chong & Treisman, 2005a; Weiss & Anderson, 1969), allowing for sufficient time to overcome backward masking effects (see DiLollo, Hanson, & McIntyre, 1983), b) rely on temporal order judgments (Haberman & Whitney, 2009), which usually cannot be retained from RSVP sequences (e.g., Hommel & Akyürek, 2005), or c) treat stimuli as a single object morphing over time (Albrecht & Scholl, 2010). For example, Chong and Treisman (2005a) presented a sequence of circles, each for 250 ms, at either a single, central location, or at random locations across the display. Regardless of where circles were presented, subjects needed a difference of about 22% to discriminate which of two probes corresponded to the mean size of the circles, but needed a difference of approximately 45% to determine the probe that corresponded to a randomly post-cued member of the set. Yet, because individual items appeared for 250 ms, it is unlikely that each item would be masked by its successor in this experiment (DiLollo et al., 1983). Weiss and Anderson (1969) obtained similar results, with subjects correctly determining the average length of line segments presented for up to 4 s each, separated by 2 s ISIs. As such, in these studies, it is unclear whether subjects used a specialized averaging mechanism, or simply encoded a small sample of individual circles and guessed the average size based on this sample.

In addition to the masking that occurs due to the processing of a first target item in an RSVP stream, Choo and Franconeri (2010) took advantage of object substitution masking to eliminate two circles from awareness by surrounding them with four red dots. Although it is clear that these masked circles contributed to judgments of mean size, it is possible that the red dots surrounding these circles increased their salience, and thus biased an estimate of mean size. In summary, although all of these studies suggest that observers are fairly accurate in their estimates of the mean size, it is not possible to conclude whether results were obtained using a sub-sampling strategy that differs from a specialized averaging mechanism we argue is at work in summary set representations. Therefore, in the present study, we presented observers with items at considerably faster rates (50 ms, separated by 50 ms ISIs) to limit the ability to consolidate representations of individual items, and to prevent the fusion of stimuli into one continuous object perceived to undergo change.

2. Experiment 1: mean judgment versus member identification

We first compared how well observers could identify the mean diameter of a set of circles briefly presented in an RSVP stream, versus how well they could recall the identities of the individual circles. Specifically, observers viewed RSVP streams of circles with distinctly different sizes, and then performed either a mean judgment task, in which they determined whether a test circle represented the mean size of the set, or a member identification task, in which they determined whether the test circle was present in the preceding set. If the mean size of the set is extracted over time, but information about the size of individual circles in the set is not included in this representation, then performance should be more accurate on the mean judgment versus the member identification task.

To tease apart any differences in the strategies participants used to perform each task, we also varied whether the test circle was presented before or after the RSVP stream. Given the difficulty of remembering an entire sequence of items, we anticipated that performance would be better in the member identification task when the test circle was presented first and subjects could search for it in the upcoming sequence, versus when the test followed the sequence and they had to compare it against what they remembered (e.g., Potter, 1976). On the other hand, if the mean representation is established by averaging all of the items in the sequence, it should matter little when the test is presented. If anything, performance should be superior when the test is presented after the sequence and there are no added demands to hold it in working memory. Importantly, if observers take the same approach to both tasks (e.g., if, despite instructions, they base their responses in the member identification task on the test circle's similarity to the mean, or base their responses in the mean judgment task on the test circle's similarity to a single, randomly-selected member of the sequence), the effect of varying whether the test precedes or follows the sequence of circles should be the same. Thus, this test circle order manipulation allowed us an additional opportunity to examine whether observers were approaching the two tasks differently.

2.1. Methods

2.1.1. Participants

Eleven undergraduates from New York University (6 female) between the ages of 18 and 26 (M = 22) volunteered in exchange for partial course credit. All had normal or corrected-to-normal vision.

2.1.2. Stimuli and apparatus

Stimuli consisted of rapid serial visual presentations (RSVP) of outline circles presented in the center of the display. In all four experiments, a 486 PC displayed the black circles against a white background on a color monitor with a 75 Hz vertical refresh rate, and recorded responses made using a computer keyboard. Circles in Experiment 1 varied in diameter in steps of 4 pixels, from 4 pixels to 96 pixels, for a total set of twenty-four circles. Viewed from a distance of 57 cm, at a resolution of 640×480 pixels (in all experiments), the circles in Experiment 1 subtended between 0.4° and 7.8° of visual angle. MEL software (Schneider, 1988) was used to control all display, timing, and response functions.

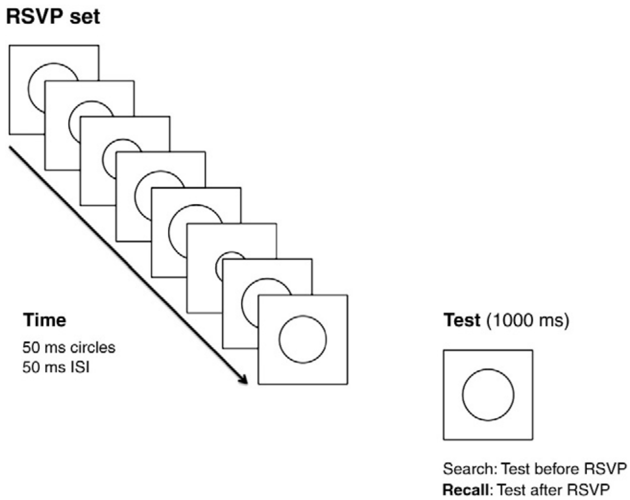

Fig. 1. Experiment 1, trial sequence: observers viewed RSVP sequences of five to eleven circles, each for 50 ms and followed by a 50 ms ISI. In the *Mean Judgment Task*, they reported whether a Test circle represented the mean size (diameter) of the preceding RSVP set; in the *Member Identification Task*, they indicated whether the Test was a member of the set. In phases when the Test was presented *before* the RSVP set, observers searched the upcoming sequence to determine whether it was the mean or a member of the set (depending on specified task); in phases when the Test appeared *after* the set, they recalled whether it was the mean or a member of the preceding set.

2.1.3. Procedure

A sample trial sequence is illustrated in Fig. 1. Each trial began with a small fixation cross, presented at the center of the display for 1000 ms. 500 ms after it was erased, a set of between five and eleven circles was chosen at random, without replacement, from the set of twenty-four circles and displayed as an RSVP sequence. Each circle was centered at fixation and shown for 50 ms (3 frames), separated by 50 ms ISIs. A test circle was also shown on each trial. Observers completed four phases, each consisting of six blocks of twenty trials, and each preceded by a single block of ten practice trials. In two of these phases, the test circle appeared for 1000 ms, and was followed by a blank screen for an additional 1000 ms before the sequence was initiated. In the other two phases, the test circle appeared 1000 ms after the sequence ended, and remained in view until a response was made. We crossed task type (mean judgment or member identification) with test circle order (test circle before or after the sequence), and counterbalanced the order of the resultant four phases across participants. At the start of each phase, observers were instructed to compare the test circle to the sequence and perform one of the two tasks: 1) in the mean judgment task, they decided whether the test circle corresponded to the (arithmetic) mean diameter of the set, and 2) in the member identification task, they decided whether the test circle was one of the circles shown in the set.

To indicate that the test circle corresponded to the mean of the set in the mean judgment task, or that it was present in the set in the member identification task, observers pressed the “s” key on the keyboard; to indicate that the test circle did not correspond to the mean of the set, or that it was absent in the set, they pressed the “x” key. Responses were not speeded, and feedback was provided immediately after each response to encourage high accuracy. Error feedback was provided by deleting the vertical segment of the fixation cross, presented at the beginning of the trial following the error, leaving a “–” sign; no change to the cross was made following correct responses. On each trial, the computer program calculated the mean and standard deviation of the diameters of the circles included in the set.

2.1.4. Mean judgment

In the mean judgment task, if the test circle corresponded to the mean of the set (50% of trials), the diameter equaled the mean diameter of the circles shown. If the test circle was different from the mean (the remaining 50%), it was chosen by adding or subtracting pixels from the mean diameter, determined as follows. The test circle was randomly selected to be larger or smaller than the mean of the set of circles shown on the trial, and the standard deviation of the set was calculated. To manipulate how discriminable the test was from the actual mean of the set, the diameter of the test circle was then chosen at random to be one of four sizes larger or smaller than the mean of the set, and thus increased or decreased by .25, .53, .84, or 1.28 standard deviations. Treating the set as a continuous normal distribution of circle diameters with the mean representing the median circle in the set, these values ensured that the test circle was larger than 10%, 20%, 30%, 40%, 60%, 70%, 80%, or 90% of the circles (i.e., sample means) that could theoretically be drawn from the set.[1] In practice, with sets of between five and eleven items, a test circle computed in this way sometimes fell out of range (i.e., was smaller than the smallest circle shown, or was larger than the largest circle shown). To correct for this, the procedure described above for the mean judgment task was carried out, and these trials were excluded from analysis.

2.1.5. Member identification

In the member identification task, if the test circle was present in the set (50% of trials), one of the circles shown was chosen at random and displayed as the test circle. If it was absent (the remaining 50%), one of the circles that was not shown was chosen, with the restriction that its diameter had to be larger than that of the smallest circle shown and smaller than the diameter of the largest circle shown. If either of these conditions was violated, another circle was chosen. If the conditions could not be met with the circles remaining in the set, the test circle was equal to either the smallest of the circles shown plus 2 pixels or the largest circle shown minus 2 pixels, determined randomly on each trial. Because this adjustment was smaller than the smallest difference in diameter among the items in the set, this procedure ensured that the test circle on target absent trials could never accidentally coincide with the diameter of an item actually shown. Data from these trials were excluded from analysis.

2.2. Results

We calculated the percentage of correct responses for each cell of the design separately for each observer, and then averaged these subject means to produce a group mean for each cell. We analyzed the resultant means using a 2 (Task: mean judgment vs. member identification) × 2 (Test Circle Order: before vs. after RSVP sequence) repeated-measures analysis of variance (ANOVA). Results are

[1] The test circle for the member identification and mean judgment tasks was necessarily selected using different criteria (i.e., on half of the trials, the test in the member identification task was an item actually shown; in contrast, the test's size was determined as a function of the mean size of the set in the mean judgment task). To determine whether this led to systematic differences in the discriminability of the test from the other circles in the set, and from the mean of the set, we computed the average difference between the size of each circle displayed and the size of the test circle, as well as the average difference between the size of each circle displayed and the mean of the set. The average difference across trials between each circle in the set and the test circle ranged from 27.5 to 30.2 pixels over set sizes five to eleven for the member identification task, and from 24.9 to 27.2 pixels for the mean judgment task. The average difference between the mean circle size and the size of the test actually shown across trials (excluding those trials on which the test corresponded to the mean) ranged from 21.1 to 23.1 pixels over set sizes five to eleven for the member identification task and 20.5 to 21.5 pixels in the mean judgment task. As a result, despite the difference in test selection method, the discriminability of the test circle from the other members of the set and from the mean of the set did not differ significantly across the two tasks (with discriminability slightly favoring the member identification task). Moreover, when considering the type of discriminability relevant for the member identification task (the average difference between the circles in set and the test circle) versus the mean judgment task (the difference between the test and the mean of set), the magnitude of the former difference (about 26 pixels) is larger than the magnitude of the latter difference (about 20 pixels). Therefore, differences in performance across these two tasks do not appear to be attributable to differences in test selection.

displayed in Fig. 2a. When the test circle was displayed prior to the RSVP sequence, performance on the mean judgment and member identification tasks was very similar, and rather poor (approximately 58%). When the target circle followed the RSVP sequence, performance fell to chance on the member identification task, but improved to just under 70% on the mean judgment task. This pattern yielded a significant Task × Test Circle Order interaction, $F(1, 10) = 50.54$, $MSE = 20.3$, $p < .001$.

To verify that a calculation of the mean was involved in making the mean judgment, we conducted a second repeated-measures ANOVA on the mean judgment task data only, with Test Circle Order (before vs. after RSVP sequence) and Difficulty (0, .25, .53, .84, or 1.28 standard deviations from the mean) as factors. Here we analyzed subjects' average percentage of "target present" responses (i.e., the proportion of trials on which subjects indicated that the test circle corresponded to the average size of the set) for each level of difficulty, to better illustrate how performance varied as a function of the difference in size between the test circle and the mean of the set. This analysis revealed a main effect of Difficulty, $F(4,40) = 28.16$, $MSE = 330$, $p < .001$, such that when the test circle's diameter was close in size to the mean of the set (.25 standard deviations), observers made more false alarms, tending to incorrectly identify the test as the mean circle (Fig. 2b). As the difference between the mean diameter and the diameter of the test circle increased, accuracy in judging the mean increased linearly, as indicated by a significant linear trend in the effect of Difficulty, $F(1,10) = 45.50$, $MSE = 792$, $p < .001$. Using linear interpolation, we estimated that subjects needed a difference greater than .42 standard deviations between the size of the test and the mean size to exceed 50% correct rejections. There was also a significant interaction between Test Circle Order and Difficulty, indicating that computing the mean facilitated comparisons more when the test circle followed versus preceded the sequence, $F(4, 40) = 6.41$, $MSE = 159$, $p < .001$. In other words, when the test circle followed: a) observers were more likely to correctly identify the test as the mean (a "hit"), but b) when the test was similar to the mean, they were more likely to incorrectly identify it as the mean (a "false alarm").

Finally, we also computed mean accuracy (percentage correct) as a function of Set Size in the mean judgment and member identification tasks, averaging over Test Circle Order. Consistent with Ariely's (2001) findings, accuracy was relatively constant in the mean judgment task as the number of circles in the set increased from five to eleven (62% to 59%), but decreased considerably over the same range for the member identification task (59% to 47%; Fig. 2c). We fit the difference in performance in each task at each Set Size to a linear function, which captured 69.5% of the variability in the difference scores. In addition, this linear trend in the difference scores was significant when tested as a within-subjects contrast for the Task × Set Size interaction, $F(1,10) = 5.24$, $MSE = 104$, $p < .045$, indicating that the difference in performance between the mean judgment and member identification tasks increased with set size.

2.3. Discussion

The results of Experiment 1 showed that observers could determine whether the size of a test circle presented after a stream of circles corresponded to the average size of the set, but performed no better than expected by chance when determining whether a single circle was a member of a series of circles presented in succession immediately beforehand. These results suggest that the mean size of the set was encoded over the course of the RSVP stream, but the identities of individual circles were not. Furthermore, false alarms on the mean judgment task exceeded correct rejections for test circles within about .5 standard deviation of the mean, but the ratio of correct rejections to false alarms increased monotonically as the difference between the size of the test circle and mean size increased. Our

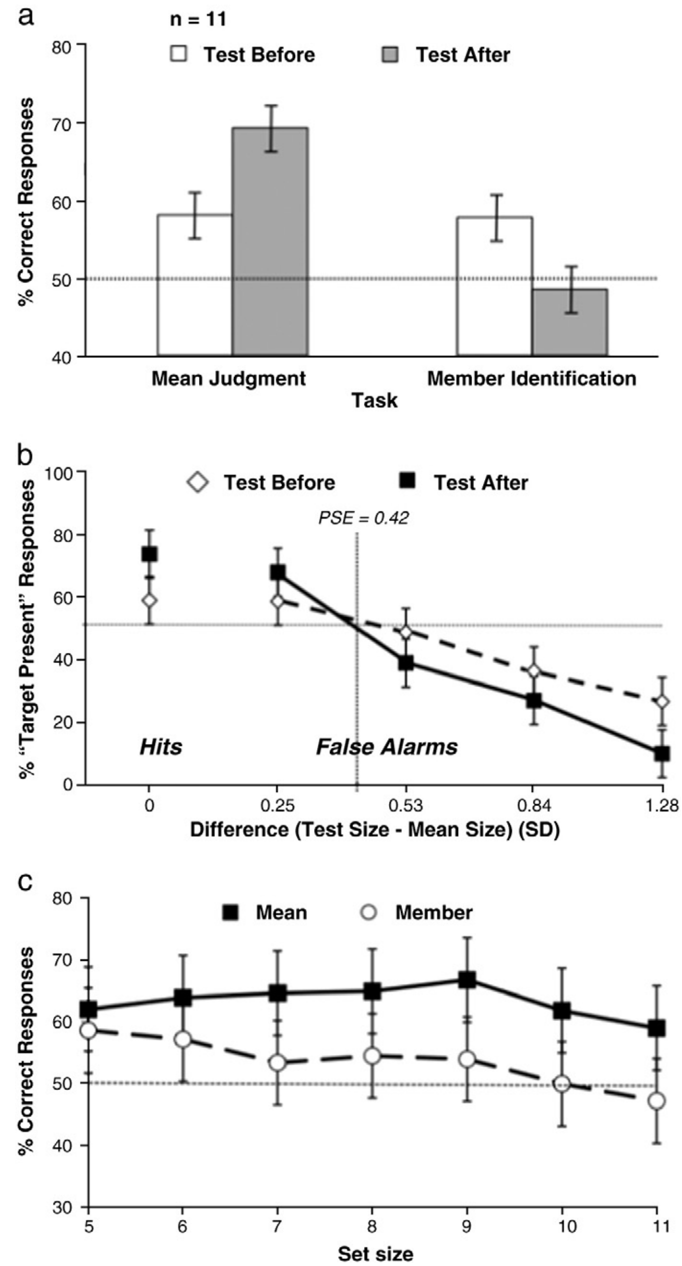

Fig. 2. Experiment 1, results: a) Although search for a target circle was similar in the mean judgment and member identification tasks, observers were better at determining whether a test circle presented after the set represented the mean size, but at chance to identify it as a member of the set, b) Observers correctly identified the test circle as the mean ("hits") more often when it was presented after versus before the set. Regardless of when the test was presented, observers' false alarms (i.e., when they incorrectly identified the test as the mean of the set) decreased linearly as the difference in standard deviation units between the test and the mean of the set increased, c) As set size increased from five to eleven, accuracy was relatively constant for the mean judgment, but decreased for the member identification task. (Error bars = 95% within-subject confidence intervals, dashed horizontal lines = chance performance).

findings parallel those of Ariely (2001), demonstrating good performance in a mean judgment task and chance performance in a member identification task for a test circle presented immediately after a display of circles.

When the test circle was presented first, observers could search the upcoming RSVP stream for the mean or the identity of an individual circle and perform the comparison with the test circle more accurately than expected by chance. However, performance on the mean judgment task was much better when the test circle followed the sequence (69% correct) than when it preceded the sequence (58% correct), consistent with the result expected if observers consider all (or most) of the sequence of circles in extracting a representation of mean size. If individual members of the set were encoded in a durable form (i.e., sufficiently consolidated to prevent masking), performance on the member identification task should have exceeded chance. Our findings of similar performance for the two tasks when the test circle was presented before the RSVP sequence, and superior performance in the mean judgment task but chance performance in the member identification task when the test circle followed the sequence support the proposal that participants used different strategies to accomplish the two tasks. Moreover, as the number of circles in the set increased, the accuracy of member identification decreased, but the accuracy of the mean judgment remained relatively constant. Such a result would be expected if subjects approached the two tasks differently; increasing the number of items in the set increases the number of individual items that must be remembered to perform the member identification task, but should have little effect on the ability to represent a single mean value.

If observers base their judgments on a small subset of items, they would likely have to attend to, and encode several circles from the sequence. Yet, if encoding a single item in an RSVP sequence impairs observers' ability to process subsequent circles, as reported for the recognition of pictures (e.g., Potter & Levy, 1969) and single letters (Dell' Acqua et al., 2009) in an RSVP stream, encoding even one circle would be expected to interfere with encoding a second circle for at least 300 ms. If the duration of this impairment does not vary as a function of the number of items already encoded, then observers could attend to, at most, two or three of the 5 to eleven circles we presented at a rate of 50 ms (with 50 ms ISIs) per item. Similarly, increasing the number of individual items to be represented should make it harder for observers to identify one of these items, whereas increasing set size should have little effect on the representation of a single summary statistic (i.e., mean size). Consistent with this interpretation, accuracy in the member identification task in Experiment 1 decreased as set size increased, but the mean judgment was unaffected. Given these differences between encoding a single versus an entire series of circles, it is unlikely that the results of Experiment 1 are due to observers sampling only a subset of the circles as a basis for their mean judgments on each trial.

2.3.1. Simulation: whole set averaging versus sub-sampling

Although our results show the pattern expected if participants were extracting the mean of the RSVP sequence of circles without representing individual circles, the possibility remained that they may have used one, or a combination of the sampling strategies outlined in the Introduction. If observers did sample only a few circles on each trial, a particularly effective strategy might be to compute the mean of the smallest and largest circles (M_{SL}) as a proxy for estimating the whole set mean (M_W). If the smallest and largest circles can be focally attended and encoded well enough to be remembered after the sequence ends, subjects would only have to identify the midpoint of a point on the circumference of the smallest and largest circles in the sequence and judge whether this point corresponds to a point on the upcoming test circle. However, such a focused attention strategy seems unlikely; it implies some foreknowledge that the particular circle being sampled is the smallest or largest, which would not be possible to know until all the circles in the sequence have been presented. Nevertheless, we carried out an ideal observer simulation to evaluate the extent to which this sampling strategy would be effective in our Experiment 1 mean judgment task when the test circle followed the RSVP sequence.

We modeled performance assuming that subjects were sampling only the smallest and largest circles from the sequence shown, and then compared the mean diameter of this subset (rather than the mean of the whole sequence) to the diameter of the test circle. In order to establish a decision rule for determining whether or not the test circle corresponded to the mean of the set (see Ariely, 2008), we interpolated the point of subjective equality (PSE) from our data in the mean judgment task in Experiment 1. Doing so, we determined that a test circle with a diameter .42 standard deviations smaller or larger than the mean would be identified as corresponding to the mean at chance (50%) levels of performance (see Fig. 2b). Thus, if the test circle fell within .42 standard deviations of M_{SL}, we assumed that an ideal observer (like subjects in Experiment 1) would classify the test as corresponding to the mean, and otherwise, the test would be rejected as not corresponding to the mean. The simulation was programmed in Matlab to run 10,000 times for each set size (5, 6, 7, 8, 9, 10, or 11) to produce 10,000*7 = 70,000 trials of data.

The results of the simulation are displayed in Table 1. Observed data from the corresponding condition in Experiment 1 are also reproduced for comparison. The outcome of the simulation (top section of Table 1) suggests that observers would correctly classify the test circle as the mean of the set (i.e., hits; 0 SD condition) based on computing M_{SL} as a proxy for M_W with considerable variability (between 40% and 100%), with the number of hits increasing as the number of circles presented during the trial (set size) increases. The false alarm rate would also vary systematically with increases in set size when the test item was .25 standard deviations smaller or larger than M_W. In the observed data, however, there is no systematic effect of set size on either hits or false alarms at these levels of difficulty (bottom section of Table 1).

Given the ideal observer categorizes perfectly relative to the PSE criterion, the simulated performance represents a theoretical maximum proportion of "target present" responses that could be obtained using this criterion. As such, the better performance (increased proportion of hits, reduced proportion of false alarms) for sets of 7 or more items than what is actually observed could be explained by making comparisons based on M_{SL} rather than on M_W. However, the simulation suggests that the levels of accuracy actually observed for set sizes of five and six could never be achieved by computing M_{SL} as a proxy for M_W. This poses considerable difficulty for the M_{SL} strategy, because it should be easier to explicitly encode the identities of the smallest and largest items in the sequence as the total number of items decreases. Arguably, this is perhaps more difficult than maintaining a single, continuously updated representation of the mean, and also requires the additional step of computing an average after identifying the circles to be considered. For these reasons, the results

Table 1
Experiment 1, simulation: Top: results of an ideal observer simulation assuming subjects compute M_{SL}. Bottom: actual performance (n = 11) in Experiment 1, when observers were instructed to compute M_W. Values represent percentage of target present responses as a function of the difference (in standard deviation units) between the test item and the mean of the set of circles shown. In the 0 condition, the values shown correspond to the percentage of "hits" expected and observed. For the remaining 4 conditions, values correspond to the percentage of "false alarms" expected and observed.

Set size	0 (hits)	.25 (f/a)	.53 (fla)	.84 (f/a)	1.28 (fla)
Simulated (M_{SL})					
5	38	37	38	18	13
6	59	53	43	19	7
7	73	64	38	13	1
8	84	70	33	8	2
9	89	78	39	6	1
10	93	80	30	6	1
11	98	81	23	3	0
Observed (M_W)					
5	73	73	56	29	11
6	78	58	32	35	7
7	69	59	25	29	20
8	81	63	39	36	7
9	75	82	24	26	8
10	69	57	47	17	9
11	69	72	60	14	0

of this simple simulation provide no support for the sampling account in which only the smallest and largest items are included in the calculation of the mean diameter of the trial set. Nevertheless, in Experiment 2, we explicitly instructed subjects to compute either M_{SL} or M_W to test empirically between these two averaging strategies. If M_{SL} is the preferred strategy in this task, subjects should favor a test circle that is closer to M_{SL} both when instructed to compute M_{SL} and when instructed to compute M_W.

3. Experiment 2: whole set mean (M_W) versus Small/Large mean (M_{SL})

The results of Experiment 1 suggest that subjects are able to extract the mean diameter of a rapid sequence of circles without separately encoding their individual sizes, as previously demonstrated for slower sequences of circles (Chong & Treisman, 2005a), and for arrays of circles in static displays (Ariely, 2001). Comparing the simulation to the results obtained for observers in Experiment 1 in the mean judgment task when the test followed the RSVP sequence shows that performance is not well accounted for by the simplified sub-sampling strategy of calculating the mean of only the smallest and largest circles in each sequence (M_{SL}). Although this strategy generally yields accurate estimates of the set mean with larger set sizes, predicted and observed performance based on this strategy differs for small sets and for smaller differences between the set mean and the test circle. Moreover, because the smallest and largest circles cannot be determined until the sequence ends, to use this strategy carries the additional requirements of maintaining two separate, continuously updated representations (one for the smallest circle and one for the largest). To better compare observers' actual performance based on these two strategies, in Experiment 2 we explicitly instructed participants to either compute the mean of the entire sequence (M_W; as in Experiment 1), or to compute the mean of only the smallest and largest circles in the set (M_{SL}).

A difficulty in comparing these two strategies is that, under most circumstances, they are both effective. This difficulty was exacerbated in Experiment 1 by the random selection of items, which yielded little variability in the smallest and largest items in the set. Therefore, even if participants could compute the mean of all circles in the set, it might not have been necessary to do so to perform well on most trials in Experiment 1, leading them to adopt this alternative strategy. However, the outcome of our simulation suggests that when the number of circles is small and the difference between the mean and the test is small (<.53 standard deviations), a calculation including only the smallest and largest circles in the sequence tends to underestimate observed performance. In Experiment 1, the PSE corresponded to a test circle between .25 and .53 standard deviations from the mean. The average difference between the mean and a test .25 standard deviations from the mean was 7.56 pixels, and the average difference between the mean and a test .53 standard deviations from the mean was 15.7 pixels. Thus, in Experiment 2 we constructed sequences that yielded a difference between the mean of the whole trial set (M_W) and the mean of only the smallest and largest circles (M_{SL}) approximately midway between these two values (12 pixels on average). Regardless of whether subjects were instructed to compute M_W or M_{SL}, test circles either corresponded to M_W or to M_{SL}. If subjects have no preference for computing M_W or M_{SL}, these alternatives should be equally likely to be judged as corresponding to the mean (i.e., performance should be at chance in each task). As well, all sets included only six circles to ensure that observers could easily identify the smallest and largest in the sequence.

3.1. Methods

3.1.1. Participants

Twenty undergraduate students (19 female) between the ages of 18 and 58 (M = 24.3) at the University of Regina participated in exchange for partial course credit. All had normal or corrected-to-normal vision.

3.1.2. Stimuli and apparatus

Stimuli were drawn from a set of circles ranging in diameter from 4 pixels to 120 pixels, in steps of 4 pixels. Because the sets had to be constructed such that they yielded a difference in M_W and M_{SL} of approximately 12 pixels, the computer program did not generate the sets randomly as in Experiment 1. Instead, a set of six items was randomly drawn from a sub-set of the circles ranging in size from 4 pixels to 96 pixels. Restricting sets to six items ensured that it was possible to create multiple sets of circles that met the criterion that M_W and M_{SL} differed by at least 12 pixels. M_W and M_{SL}, and the absolute value of the difference between them, $|M_W - M_{SL}|$, was computed. The set was retained if $|M_W - M_{SL}|$ was approximately 12 pixels, and discarded otherwise. This process was then repeated nine times to yield ten different sets of six circles. Sixty additional sets of six circles were created by adding constants of 4, 8, 12, 16, 20, and 24 to each circle in the ten randomly generated sets. This was done to ensure that the smallest and largest circles in the set varied from trial-to-trial without increasing the range of the set, which in turn ensured variability in both M_W and M_{SL} across trials (the former ranged from 20.0 to 60.7 pixels; the latter varied from 32.0 pixels to 74.0 pixels). The apparatus used was identical to that of Experiment 1.

3.1.3. Procedure

The procedure was as in Experiment 1 (Fig. 1), with the following exceptions. The member identification task was omitted, and instead subjects performed one phase of six blocks of thirty trials judging either M_W or M_{SL}, then a second phase performing the other task. The order of tasks was counterbalanced across subjects. Each phase began with a block of ten practice trials to familiarize subjects with the task. In each task, on each trial, the test circle was randomly selected to be either M_W or M_{SL}, and was presented after the RSVP sequence, as in Experiment 1. Subjects were instructed to decide whether the test circle shown corresponded to the relevant mean, but were not told anything about how test circles were chosen. In contrast to Experiment 1, no feedback was provided to ensure that response choices were not influenced by task demands (e.g., Bauer, 2009).

3.2. Results

We computed the mean percentage of correct responses for each subject in each task, and then averaged them together to yield group means (Fig. 3). To determine whether the order in which observers performed the M_W and M_{SL} tasks influenced their strategies, we then analyzed group means in a 2 × 2 mixed-model ANOVA with Task Order as a between-subjects factor, and Task (M_W or M_{SL}) as a within-subjects factor. There was no main effect of Order, and no Order × Task interaction, $F(1, 18) = 2.45$, $MSE = 232$, $p > .13$, but there was a significant main effect of Task, $F(1, 18) = 4.64$, $MSE = 232$, $p < .05$. Accuracy was poor overall, suggesting that subjects had difficulty differentiating the two types of means. Nevertheless, their performance was reliably better than chance when they were instructed to judge M_W ($M = 56\%$; $t(19) = 2.14$, $p < .05$), and tended to be worse than chance when they were instructed to judge M_{SL} ($M = 44\%$; $t(19) = -1.77$, $p < .10$). Because the test circle on target absent trials in each case was always the other mean regardless of whether observers were performing the M_W or M_{SL} task, the results suggest participants were slightly more likely to categorize a test circle corresponding to M_W versus M_{SL} as the mean of the set. In other words, irrespective of which task they were performing, on over half of the trials in Experiment 2, participants judged M_W to correspond to the mean of the set.

3.3. Discussion

Although computing M_{SL} is generally an accurate proxy for computing M_W, the results of Experiment 2 suggest that this strategy is

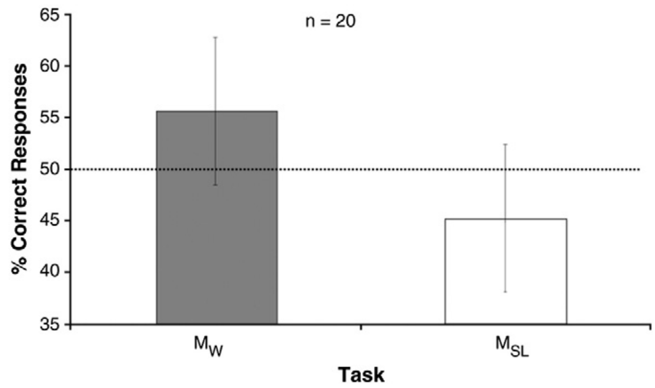

Fig. 3. Experiment 2, results: although performance should have been at chance for each task if subjects have no preference for computing M_W or M_{SL}, they were reliably better than chance at determining whether a test circle corresponded to M_W, but tended to be worse than chance at determining whether the test corresponded to M_{SL}. As in each task, M_W or M_{SL} was randomly chosen as the test on every trial, results suggest participants were more likely to categorize a test corresponding to M_W as the mean of the set than they were to categorize a test corresponding to M_{SL} as the set mean (Error bars = 95% within-subject confidence intervals, dashed horizontal line = chance performance).

not used, at least not when the smallest and largest circles in the set (and M_{SL}) vary non-systematically across trials. Despite the similarity between the two possible test circles, subjects reliably chose M_W more often than expected if they had no preference for either strategy, even when they were instructed to compute M_{SL}. Thus, when instructed to compare M_{SL} to a test circle, test circles that corresponded to M_W were judged to be more similar to M_{SL} than were test circles that actually corresponded to M_{SL}. More importantly, in the absence of feedback that could have potentially influenced responses in Experiment 1, no evidence emerged to suggest that subjects compute M_{SL} when instructed to compute M_W.

4. Experiment 3: mean judgment versus largest circle identification

Accuracy in judging M_W in Experiment 2 was near (although reliably above) chance, as expected given that the two test circles corresponding to M_W and M_{SL} differed by just 12 pixels, a difference that is close to the .42 standard deviation difference that yielded 50% "target present" responses in Experiment 1. We conducted Experiment 3 to provide converging evidence that subjects compute M_W, versus computing the mean of a sub-sample of the circles in each set. To do so, we compared performance when subjects were instructed to compute the mean of the entire sequence of circles, as in Experiments 1 and 2, against performance in a new member identification task, in which they were asked to identify only one, very conspicuous member of the preceding set, the largest circle. Towards these ends, the purpose of Experiment 3 was fourfold. First, we sought additional evidence against the possibility that subjects base their responses on a calculation of the mean that includes only a subset of circles (the smallest and largest). Second, we wished to confirm that encoding a single circle from a sequence of circles for later report is more difficult than extracting the average diameter of the sequence. Third, we hoped to verify that the apparent advantage of extracting the mean diameter of the sequence of circles over reporting the identity of a given individual circle observed in Experiment 1 was not an artifact of differences in memory demands or differences in how the test circle was determined between the mean judgment and member identification tasks (see Footnote 1). Finally, we wished to ensure that subjects could extract M_W of a rapid sequence of individual circles, even when the sequence could not be perceived as a single morphing disk changing rapidly in size over time (cf. Albrecht & Scholl, 2010). Therefore, in Experiment 3, we asked participants to perform the mean judgment task as in Experiments 1 and 2, and compared their performance in this task to their performance in determining whether a test circle corresponded to the largest circle in the preceding sequence.

4.1. Methods

4.1.1. Participants

Sixteen undergraduate students from the University of Regina (12 female) between the ages of 18 and 52 (M = 22.3) participated in exchange for partial course credit. All had normal or corrected-to-normal vision.

4.1.2. Stimuli and apparatus

Stimuli included circles and random shapes, all drawn in MEL. Non-circle shapes were selected from a set containing twelve shapes: triangle, square, trapezoid, pentagon, hexagon, octagon, parallelogram, diamond, crescent, cross, five-pointed star, and six-pointed star. Shapes were chosen from this set at random on each trial, and circles were selected at random on each trial from the same set as in Experiment 1. Although the circles varied in size from 4 to 96 pixels in diameter (subtending approximately 0.4 to 7.8° of visual angle), the shapes were of approximately uniform size (about 30 pixels in width and height), and subtended 2.5° of visual angle. The apparatus was identical to that used in Experiments 1 and 2.

4.1.3. Procedure

Sequences of five to eleven circles were interleaved with sequences of an equal number of shapes (i.e., circles were alternated with other random shapes within each sequence). Each sequence began with a randomly selected circle and ended with a randomly selected shape. As in Experiment 2, a single test circle was displayed at fixation 1000 ms after the last item in the sequence, and remained in view until the subject responded. Subjects completed two counterbalanced phases, performing a member identification task in one phase and a mean judgment task in the other. In the member identification task, they were instructed to judge whether the test circle corresponded to the largest circle in the sequence. On half of member identification trials, the largest circle in the set was selected as the test circle (target present trials); on the other half, the second-largest circle in the set was selected as the test circle (target absent trials).

The difference between the diameters of the largest and second-largest circles varied depending on the circles included in the set, and was not constrained in any way. In the mean judgment task, subjects were instructed to judge whether the diameter of the test circle corresponded to the mean diameter of the circles shown in the sequence. On half of mean judgment trials, the test circle corresponded to M_W (target present trials). On the other half, the test circle varied from the mean by an amount that was equal to the difference between the largest and second-largest circles in the set (target absent trials). If M_W was less than 50 pixels (the midpoint of the smallest possible and largest possible values of M_W for this set, 12 pixels and 88 pixels), the test circle was larger than M_W; if M_W was more than 50 pixels, the test circle was smaller than M_W. As such, for both tasks, the difference between test circles on target present trials and target absent trials was identical, and equal to the difference between the largest and second-largest circles in the sequence of circles shown. Target present and target absent trials were randomized within blocks. Each phase began with one block of ten practice trials to familiarize subjects with the task. Feedback was provided during these practice trials, in which a "+" or "−" was displayed at screen center for 1000 ms prior to the onset of the sequence of shapes to indicate a correct or incorrect response, respectively. However, as in Experiment 2, no feedback was provided during experimental trials, so as not to introduce bias towards a particular strategy for calculating the mean. Following the practice block, subjects completed eight blocks of twenty experimental trials in each phase.

4.2. Results

Because test circles could differ from the mean or largest circle by as few as 4 pixels, or as many as 80 pixels, we first binned accuracy into quintiles of the size of this difference. Trials with small differences occurred much more frequently than trials with large differences, so despite our attempt to equate the number of observations in each bin, the number of observations varied slightly across quintiles. The quintiles were determined by calculating the difference between the largest and second-largest circles on each trial, irrespective of the judgment to be performed. This yielded the following quintiles (the number of observations in each quintile is indicated in brackets; first for the mean judgment and then for the member identification task): 4 pixels difference (828 observations, 790 observations), 8 pixels difference (506, 539), 12 pixels difference (399, 401), 16–20 pixels difference (409, 439), and 24–68 pixels difference (319, 299).

The mean percentage of "target present" responses was computed for each subject in each task as a function of quintile, and subject means were averaged to yield group means. Results are displayed in Fig. 4. Because half of trials within each task were target absent trials, the design of the experiment afforded calculation of both sensitivity (d′) and response bias (β) parameters; these measures are reported in Table 2. Consistent with Experiment 1, accuracy was higher in the mean judgment task (M = 57%) than in the member identification task (M = 48%), leading to a significant main effect of Task, $F(1,15)=18.6$, $MSE=300$, $p<.001$. Performance in the mean judgment task reliably exceeded chance, $t(15)=5.92$, $p<.001$, but performance in the member identification task did not reliably differ from chance, $t(15)=0.90$, $p>.38$. Performance in the mean judgment task improved from 50% to 73% as the difference between the largest and second-largest circles increased from 4 pixels to 24+ pixels, but no such improvement was observed in the member identification task, in which performance never exceeded chance over the same range of differences (51% to 55%, with the latter value not reliably greater than chance, $t(15)=.78$, $p>.44$). This led to a significant interaction between Task and Bin, $F(4, 60)=3.08$, $MSE=167$, $p<.03$. Simple effects tests confirmed that the effect of Bin was significant for the mean judgment task, $F(4, 60)=11.6$, $MSE=.109$, $p<.001$, but not for the member identification task, $F(4, 60)=2.94$, $MSE=198$, $p>.08$. As indicated in Table 2, d′ exceeded zero (with α set to .005 using the Bonferroni correction for multiple comparisons) only at the larger probe differences, and only in the mean judgment task. No significant response biases were observed.

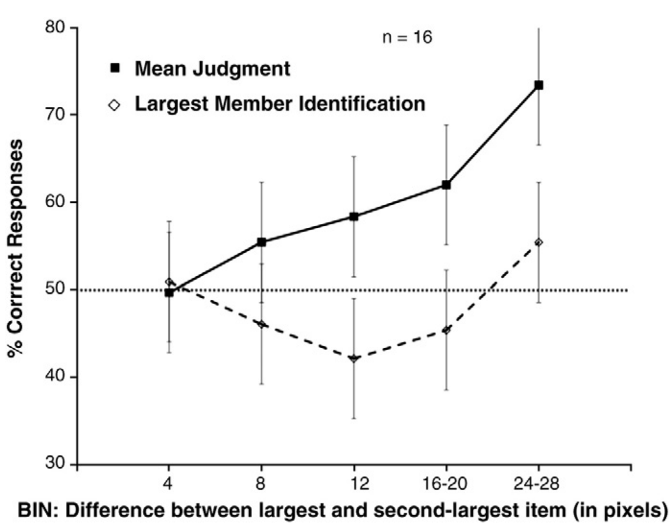

Fig. 4. Experiment 3, results: accuracy was higher in the mean judgment task than in the largest member identification task. Whereas performance was not reliably better than chance in discriminating whether a test circle corresponded to the largest (versus second-largest) circle in each set, observers were consistently better than chance when discriminating whether a test circle was the mean (versus a circle that was smaller or larger than the mean by an amount corresponding to the difference in size between the largest and second largest circle in the set). (Error bars = 95% within-subject confidence intervals, dashed horizontal line = chance performance).

Table 2
Experiment 3, results: d′ and β are displayed for each task (Mean judgment and Largest member identification) as a function of the difference in the number of pixels between the largest and second largest items appearing in the sequence (Bin).

Bin	4	8	12	16–20	24+
Mean judgment					
d′	−.14	.21	.64*	.81*	1.45*
β	.19	−.05	−.55	−.41	.38
Largest member identification					
d′	−.04	−.34	−.38	−.33	.39
β	.01	.11	.07	.21	−.44

* *p<.005.*

4.3. Discussion

As in Experiment 1, subjects were able to judge the mean size of the sequence of circles, but were not able to determine whether a single item specified in advance (in Experiment 3, the largest circle) was presented in the sequence. Given that subjects could not encode just the largest circle in the set well enough to discriminate it from the second-largest circle, even when it was more than 24 pixels smaller than the largest circle, it is unlikely that the sub-sampling strategy of encoding and averaging both largest and smallest circles in the set as a proxy for computing the mean diameter of the entire sequence is a viable alternative. Yet, this same difference in size[2] between M_W and a circle that was smaller or larger led to considerably better performance in the mean judgment task (73% correct). This discrepancy suggested that, under precisely the same conditions, a summary representation of the sequence was consolidated when individual circle representations were not. Under conditions that limit the ability to encode individual circles for later report, identification of single circles – even the largest, most conspicuous circle – is impaired relative to extracting a summary representation of the set. Furthermore, superior performance in the mean judgment versus largest member identification task cannot be attributed to greater memory demands in the latter than in the former. In Experiment 3, observers were only required to remember a single, conspicuous circle. Arguably, this places less of a demand on working memory than performing a calculation that requires encoding multiple circles from the RSVP sequence, even if it is assumed that subjects only need to encode two circles to arrive at a reasonable guess about the mean of the whole set (Myczek & Simons, 2008). Finally, Experiment 3 replicates the key findings of Experiment 1, but with a sequence in which target items (circles) are interleaved with distractors (random shapes), ensuring that items are clearly seen as

[2] We note that, proportionally, a difference of N pixels between M_W and a test circle is greater than a difference of N pixels between the largest and second-largest circles in the set. For example, a difference of 24 pixels between a circle of 72 pixels and a circle of 96 pixels represents a 33% difference in size. For a trial on which M_W was 50 pixels, a test circle differing by 74 pixels would correspond to a 48% difference in size. We do not believe, however, that this difference is sufficient to account for the results obtained. Examination of Fig. 4 suggests that a difference between M_W and the test circle of 8 pixels (a difference from 50 pixels of 16% in the preceding example) yields performance similar in the mean judgment task to a difference of at least 24 pixels between the largest and second-largest circles in the set (a difference of 33% in the preceding example). At a difference of 33% (approximately 17 pixels, in the preceding example), accuracy in the mean judgment task is over 60%, which is at least as good (if not better) than the 55% correct performance observed in the member identification task with a difference of 33% between the largest and second largest items.

individual circles (rather than as a single object changing over time — see Albrecht & Scholl, 2010).

5. Experiment 4: mean size in the attentional blink

A difficulty that consistently arises in comparing performance in the mean judgment and member identification tasks in the three previous experiments is that subjects were instructed to make different judgments in each case (i.e., compute the mean vs. remember one circle). Arguably, they could have performed the mean judgment task by encoding and remembering a small number of circles and computing the mean on that small sub-sample, whereas determining whether a specific circle was present (even if it was the largest or smallest) always required a consideration of all circles shown in the set. In Experiment 4, we eliminated this difference in task demands by examining a situation in which we could directly verify that subjects could not reliably distinguish a specific individual circle (the largest or smallest) from a distractor, while confirming that the unreportable circle contributed to their estimates of the mean on the same trial. Specifically, if individual circle identities are not included in the representation of the set's mean size, but all the circles are nonetheless factored into the calculation of the mean (versus sub-sampling strategies), circles that are not able to be identified should nevertheless affect the perceived mean size of the set.

A dual-task variant of the RSVP method we employed has been used extensively in the *Attentional Blink (AB)* paradigm (e.g., Raymond, Shapiro, & Arnell, 1992), in which observers' task is to report the identities of two targets embedded in an RSVP sequence. The typical finding is that the first target item (T1) can be fully processed for identification and reported with high accuracy, but the second target (T2) is detected or identified much less frequently if it appears within about 500 ms of the first. This second-target deficit decreases as the lag between T1 and T2 increases, suggesting that it takes about half a second to fully process T1. Researchers have frequently interpreted this effect as a transient difficulty in consolidating a representation of T2 in short-term memory that results from insufficient availability of central attentional resources (e.g., Jolicœur, 1999). Although it is clear that T2 is processed to some extent (e.g., Luck, Vogel, & Shapiro, 1996), this processing is insufficient to yield a representation that is sufficiently durable to allow T2 to be discriminated from a distractor at the end of the trial. If, as we have argued, representing average size does not require encoding of individual items, it should be possible to extract average size even when the ability to consolidate representations of individual circles in the sequence is impaired by the attentional blink.

To examine whether items that are not able to be identified due to the AB are included in the calculation of mean size, we modified our paradigm to create conditions that would vary the availability of the resources needed to consolidate representations of individual items in each sequence. In Experiment 4, subjects were instructed to identify two targets, a single shape (square or triangle; T1) presented as the last item within a sequence of random shapes, and a test circle (T2) sometimes appearing within an RSVP sequence of circles immediately following the sequence of shapes. Responses to T1 were speeded to ensure that this task consumed the attentional resources needed to process T2, presented a short time later (Jolicoeur, 1998). Each sequence contained an *Outlier (O)* that was clearly larger or smaller than the rest of the circles, and a circle whose size corresponded to the mean size of the whole set of circles (M_W). One of these two circles appeared at lag 2 (two circles later, a delay of 200 ms; when central attentional resources were likely unavailable for processing T2) and the other at lag 8 (eight circles later, a delay of 800 ms; when resources were likely available for processing T2).

On some trials, subjects were asked to judge whether the O or M_W test circles appeared in the set. If subjects' reports are influenced by a calculation of the mean, O tests should be more frequently rejected as set members than M_W tests because outliers bear less resemblance to the set as a whole than a circle with a diameter that corresponds to the mean of the set. Moreover, the probability of recalling O should be greater when it appears at lag 8 than when it appears at lag 2, because central attentional resources are more likely to be available at longer lags. If subjects actually recall the mean circle as having been present in the set (rather than determining whether it was present by comparing it to the mean size of the set), the probability of recalling the mean circle – like the outlier circle – should be influenced by the lag between T1 and T2. Finding an effect of lag on the O judgment, but not on the M_W judgment would suggest different processes are at work in making the two types of decisions. Of course, there should be no effect of lag when the test circle is a *Foil (F)*, which did not appear in the set. Such a pattern of findings would confirm that subjects' responses are not influenced by the relative position of the outlier and mean circles in the sequence per se, but rather by the similarity between the test circle and the mean size of the set.

5.1. Methods

5.1.1. Participants

Thirteen students volunteered their participation in exchange for partial course credit. Participants (9 female) ranged in age from 19 to 22 years (M = 19.9 years). All had normal or corrected-to-normal vision. Data from one participant were excluded from analysis because she apparently reversed the key mapping. Eliminating trials on which Task 1 was not made within 2500 ms excluded all trials completed by a second participant who apparently was not speeding his first response as instructed, leaving data from eleven participants in the analyses reported below.

5.1.2. Stimuli

Stimuli were drawn from a set of thirty-nine black outline circles spaced evenly on a power function (relating perceived and actual size) with the exponent .76 (Teghtsoonian, 1965). Circles ranged in diameter from 3 to 129 pixels, approximately .12° to 5.16° of visual angle, and were presented at fixation on a white background.

5.1.3. Procedure

Each trial began with the presentation of a fixation cross at screen center for 500 ms. This cross was cleared from the screen, and after a 500 ms delay, a sequence of shapes was presented. On each trial, between seven and eleven shapes (drawn from the set described in the Method of Experiment 3) were presented in a random order in an RSVP sequence, immediately followed by an RSVP sequence of eleven circles. Each item was displayed for 50 ms, followed by a blank screen for 50 ms. Nine circles were selected randomly from circles between 21 and 84 pixels in diameter. A tenth circle was an Outlier (O), chosen to be either larger than rest of the set (on a random half of trials), or smaller than the rest of the set (on the other half of trials). When larger, O was chosen from a set of circles ranging from 112 to 129 pixels in diameter; when smaller, it was chosen from a set ranging from 3 to 12 pixels. This method of selection ensured a minimum difference of five steps on the power function between the next largest (or smallest) circle in the set and O. The eleventh circle was computed as the mean of the other ten circles in the set (M_W). Trial means ranged from 36 to 67 pixels. The last of the shapes (prior to the first circle) was either a square or a triangle, determined randomly on each trial.

We instructed subjects to identify the shape (T1) by making a speeded response to indicate whether it was a square (by pressing the “z” key on the keyboard with the left ring finger), or a triangle (by pressing the “x” key with the left index finger). Instructions emphasized the importance of responding quickly. Subjects were told that either a square or a triangle would appear only once within the shapes portion of the RSVP sequence, and that this target would always be the last shape before the appearance of the first circle. The order of the presentation of circles was randomized, with the constraint that on half of the trials, the second circle (Lag 2) was O and the eighth circle (Lag 8) was M_W, and on the other half of trials, this arrangement was reversed.

Following a one second delay after the last circle was cleared, a test circle appeared and remained in view until the subject responded. Subjects were instructed to decide whether the test circle (T2) had appeared within the set of circles shown. The test item was O on 25% of trials, and M_W on 25% of trials. On the remaining 50% of trials, the test item was a Foil (F) determined by computing the midpoint between O and M_W (depending on whether O was the largest or smallest circle in the set). Thus, on half of the trials, the test circle was present in the set, and on half of the trials, it was absent but midway between two items actually shown. We instructed subjects to press the "." key with the right index finger if the test item appeared in the set shown, and the "/" with the right ring finger if it did not.

The experiment began with six blocks of ten practice trials, followed by sixteen blocks of twelve experimental trials. Subjects were given the opportunity to take a break after each block of trials. Lag (2, or 8) and Test Type (O, M_W, or F) were randomized within blocks.

5.2. Results

We computed the mean proportion of correct responses for each subject as a function of the lag between T1 and O (Lag 2 vs. 8; note that the lag between T1 and M_W is always the converse of the lag between T1 and O), and Test Type (O, M_W, or F), and averaged them together to produce group means (displayed in Fig. 5). Only Task 2 responses from trials on which the response to Task 1 was correct were analyzed, to ensure that subjects were actually devoting resources to performing Task 1 (cf. Joo, Shin, Chong, & Blake, 2009). For the same reason, only trials on which the response time in Task 1 was 2500 ms or less were considered for analysis; excluding responses that were longer than this eliminated 8.1% of trials with a correct response in Task 1. Accuracy on Task 1 ranged from 75% to 81% across conditions.

Task 2 response accuracy was analyzed in a 2 (Lag) × 3 (Test Type) repeated-measures ANOVA. The effects of both Lag $F(1, 10) = 9.44$, $MSE = 45.5$, $p < .001$, and Test Type $F(2, 20) = 9.36$, $MSE = 1243$, $p < .001$, were significant. Lag strongly affected responses to O tests but had no effect on responses to M_W or F tests, leading to a significant Lag × Test Type interaction, $F(2, 20) = 4.84$, $MSE = 245$, $p < .02$. Paired samples t-tests revealed a significant difference in accuracy for O tests at Lag 2 (.34) and Lag 8 (.56), $t(10) = 3.24$, $p < .009$, but no effect of Lag for M_W tests (.82 vs. .83; $t < 1$), or for Foil tests (.44 vs. .37; $t(10) = 1.45$, $p < .17$). Two additional ANOVAs were carried out with the same factors and levels for Task 1 accuracy and response times. No significant main effects or interactions were observed in either analysis (all $Fs < 2.18$, all $ps > .13$).

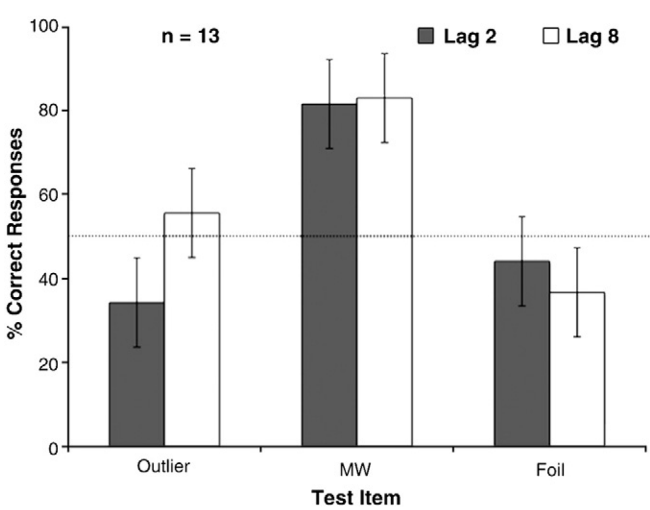

Fig. 5. Experiment 4, results: when presented with sets containing both a circle corresponding to the mean size of the set (M_W) and an outlier (O), subjects overwhelmingly endorsed the former but rejected the latter (about as frequently as a Foil, F) as present in the set. Accuracy in reporting the presence of O was strongly influenced by performing a demanding concurrent shape identification task, suggesting that resources needed for consolidating a representation of individual circles were less readily available at Lag 2 versus Lag 8. However, accuracy of reporting the presence of a circle corresponding to the average size of the set was completely unaffected by Lag, suggesting that judgments of O and M_W relied on different processes. (Error bars = 95% within-subject confidence intervals, dashed horizontal lines = chance performance).

5.3. Discussion

1In Experiment 4, subjects were not instructed to calculate the mean size of the circles shown. Nevertheless, the results strongly suggest that they did so anyway in order to perform the difficult task of judging whether a single circle had appeared within the sequence. When presented with sets containing both a circle corresponding to the mean size of the set (M_W) and an outlier (O), subjects overwhelmingly endorsed the former but rejected the latter as present in the set (about as frequently as a Foil, F). Subjects' accuracy in reporting the presence of O was strongly influenced by performing a demanding concurrent shape identification, suggesting that resources needed for consolidating a representation of individual set items were less readily available at the shorter lag than at the longer lag. However, the accuracy of reporting the presence of a circle corresponding to the average size of the circles shown was completely unaffected by its proximity to the first shape target.[3] The effect of Lag on O identification versus the lack of such an effect on M_W identification suggests that judgments of O and M_W relied on different processes: to succeed in judging an outlier as present in the set, subjects had to try to remember each individual circle shown, but to succeed in judging the mean circle as present in the set, they only had to evaluate its similarity to the representation of mean size extracted from the sequence.

Importantly, this representation of mean size was extracted even though it is clear that under these conditions, the encoding of individual items was limited by the concurrent task and by limitations in the capacity of visual short-term memory. Moreover, irrespective of whether O was presented during the attentional blink (i.e., Lag 2), or not (Lag 8), M_W was equally likely to be judged as part of the set. This finding suggests that the perceived mean size of the whole set was the same whether the outlier was consolidated well (Lag 8), or not (Lag 2). Had the outlier not been factored into the calculation of the set average when it was subject to the attentional blink, participants should have been less likely to judge tests corresponding to the mean circle as part of the set when O appeared at Lag 2 than at Lag 8. This would be expected because the representation of the mean size should be skewed by the outlier, reducing the perceived similarity of M_W tests to the average size of the set. The fact that this did not happen strongly suggests that the representation of mean size incorporates items that are not consolidated as individual items sufficiently well to be reported.

We wish to note that Joo et al. (2009) recently conducted a similar study in which: a) the entire display of circles to be averaged was presented simultaneously as T2, following an RSVP stream of digits containing a single letter target (T1; Experiment 2-1), b) two single circles to be identified were presented on either side of fixation as T2, following an RSVP stream of digits and a target letter (T1; Experiment 2-1), or c) individual circles to be averaged each surrounded an item within an RSVP sequence containing letters and two digit targets (T1

[3] The finding that the relative positioning of O near the beginning or end of the sequence had no effect on estimates of average size converges with recent unpublished work by one of our Honors students (Schweitzer, 2009). Using a similar RSVP paradigm, we presented circles in ascending, descending, or random order according to size. The ordering of sizes had no effect on whether a test circle corresponding to M_W was correctly discriminated from a distractor that was 20% larger or smaller, suggesting that neither items at the beginning nor the end of the sequence are weighted differently in computing mean size.

and T2; Experiment 2-2). The authors assert that although individual circles were masked by the attentional blink such that observers could not reliably determine which of the two side-by-side circles presented as T2 was larger, there was no effect of the lag between the detection of T1 and T2 in Experiment 2-1 on observers' ability to discriminate whether a test circle presented after the RSVP sequence was the mean of the T2 display (with nonsignificant effects of Lag regardless of whether the discrimination was easy [$p = .10$], or hard [$p = .06$]). Further, they report no effect of lag in Experiment 2-2 on observers' abilities to discriminate a test circle from the mean of the RSVP sequence of circles surrounding the letter and digit stimuli, despite the effect of Lag observed for reporting the identity of the second letter target (T2) in close temporal proximity to the first letter target (T1). However, although their finding that individual circles were susceptible to the attentional blink reinforces our choice of methodology, none of their experimental tasks ensured that a given circle that could not be identified was included in the mean calculation, as in our present Experiment 4. Furthermore, it is not clear whether the results of the averaging task were conditionalized on performing the concurrent attentional blink tasks in their second experiment. Specifically, their claim that the averaging task was immune to the AB is valid only if subjects were able to report both T1 and T2; on trials on which either target was reported incorrectly, additional resources could have been available for performing the averaging task (see Dell' Acqua et al., 2009 for additional discussion on the importance of this within-trials contingency principle). Thus, it is possible that on a considerable proportion of trials in their experiments, resources may, in fact, have been available for consolidating representations of individual circles. As such, it is unclear whether averaging in Joo et al.'s experiment was accomplished by using a specialized mechanism or by relying on a sub-sampling strategy.

6. General discussion

We have demonstrated that the average size of a set of circles can be represented under conditions that impair or prevent discrimination of the identities of the individual circles making up the set. We used an RSVP method of briefly presenting a series of single items in one location to create spatiotemporal conditions well known to impair the identification of multiple targets. In Experiment 1, we showed that observers could identify the mean diameter of the series, but not the diameters of individual circles. Furthermore, the mean judgment was superior when the test circle followed the RSVP sequence and observers had to encode global properties of the entire set, versus when the test circle was specified in advance and had to be held in short-term memory for the duration of the RSVP sequence. However, increasing memory demands for the test item by presenting it after the sequence eliminated the ability to perform the member identification task altogether, while improving the accuracy of mean judgment. The differential effects of increasing memory demands on the two tasks confirm they are carried out by different processes. The results of an ideal observer simulation suggested that averaging only the smallest and largest items in the set to estimate the whole set mean was not sufficient for performing the task with accuracy as high as was actually observed for small set sizes. In Experiment 2, we explicitly instructed participants to compute either the mean of the whole set shown (M_W), or only the mean of the smallest and largest circles in the set (M_{SL}), and followed each sequence with a test circle that corresponded to one mean or the other. Our belief that identifying and averaging two pre-specified circles from the sequence would be more difficult than averaging all of them was supported by the failure to find any evidence that subjects preferred to base judgments of the test circle on the mean of the smallest and largest circles in the set (and, actually, a slight but significant bias against doing so). Experiment 3 provided additional evidence against the suggestion that subjects encode the specific identities of individual circles in evaluating mean size. When instructed to remember and report the largest circle in the sequence, subjects were unable to reliably differentiate it from the second largest circle in the sequence. However, when asked to judge the mean size of the set, they were readily able to differentiate a test circle corresponding to the mean from one that varied by an amount equivalent to the difference between the largest and second largest circles in the set. In Experiment 4, observers extracted a representation of the mean size of a set of RSVP circles after performing a shape discrimination task (T1) that incorporated the size of an Outlier circle, regardless of whether it was presented during the attentional blink, and thus not consolidated sufficiently well to be reported as T2.

6.1. Statistical averaging or sub-sampling?

A persistent criticism of arguments in favor of the existence of a rapid averaging mechanism is that comparable performance can often be achieved by averaging only a small sub-sample items, rather than by computing the average of the target characteristic for the whole set (e.g., Myczek & Simons, 2008). The results of the present investigation demonstrate that this is highly unlikely for several reasons. First, in all conditions in which subjects were required to individuate one or more items in the sequence, performance costs were observed. This was true even when subjects only had to monitor the sequence for the largest item, and knew that their task only required remembering the size of this one circle. Although it is possible that subjects indeed relied on only one or two circles in comparing test circles to the mean of the set, our results suggest that doing so perfectly would not yield performance similar to what we observed (M_{SL} simulation of Experiment 1), and that whether instructed to do so or not, subjects did not use the most effective sampling strategy available, namely averaging the smallest and largest circles in the set. Further, Experiment 3 showed that dropping the requirement to identify only one, highly conspicuous circle did not improve performance: subjects could not discriminate even the largest circle from the second-largest, but could discriminate the mean from a test circle differing in size by the same amount. Second, as shown in Experiment 4, attending to a single circle impairs subsequent report of a single target item, but processing an entire sequence of circles does not. These findings indicate that observers do not use the same approach when processing the single circle as they do when processing the entire sequence of circles. If a sampling strategy is used, it does not appear to involve consolidating stable representations of individual items in the RSVP sequence, as is required to accurately report the identity of a single circle, square, or triangle from the sequence.

Overall, our results extend Ariely's (2001) proposal by demonstrating that the visual system forms a representation of the mean size of a sequence of similar objects that does not include identifying information about individual elements. Although encoding is not sufficient to consolidate a stable representation of each item individually, enough information survives masking by successive items to allow the system to construct a statistical summary of certain characteristics of the whole set. In closing, the present results support the existence of an efficient mode of perceptual representation that does not rely on sampling and explicitly encoding individual elements, but instead captures the average properties of the entire set.

Acknowledgment

This research was supported by a Natural Sciences and Engineering Research Council of Canada grant to CO.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*, 560–567.

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

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, *70*, 1325–1326.

Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, & Psychophysics*, *71*, 655–665.

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

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

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

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

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

Corbett, J. E., Oriet, C., & Rensink, R. A. (2006). The rapid extraction of numeric meaning. *Vision Research*, *46*, 1559–1573.

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

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

Dell' Acqua, R., Jolicoeur, P., Luria, R., & Pluchino, P. (2009). Reevaluating encoding-capacity limitations as a cause of the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance*, *35*, 338–351.

DiLollo, V., Hanson, D., & McIntyre, J. S. (1983). Initial stages of visual information processing in dyslexia. *Journal of Experimental Psychology. Human Perception and Performance*, *9*, 923–935.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Summary statistical representations of faces. *Journal of Experimental Psychology. Human Perception and Performance*, *35*, 718–734.

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *Quarterly Journal of Experimental Psychology*, *58A*, 1415–1433.

Jolicoeur, P. (1998). Modulation of the attentional blink by on-line response selection: Evidence from speeded and unspeeded Task1 decisions. *Memory & Cognition*, *26*, 1014–1032.

Jolicœur, P. (1999). Concurrent response selection demands modulate the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance*, *25*, 1097–1113.

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*(7), 1–12.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings are accessed but cannot be reported during the attentional blink. *Nature*, *383*, 616–618.

Marchant, A. P., & de Fockert, J. W. (2009). Priming by the mean representation of a set. *Quarterly Journal of Experimental Psychology*, *62*, 1889–1895.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*, 429–434.

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.

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

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 509–522.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*, 10–15.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, *18*, 849–860.

Rosenholtz, R., & Alvarez, G. A. (2007). How and why we perceive sets: What does modeling tell us? *Talk presented at the annual meeting of the European Conference on Visual Perception, August 27–31*.

Schneider, W. (1988). Micro-experimental laboratory: An integrated system for IBM-PC compatibles. *Behavior Research Methods, Instruments, & Computers*, *20*, 206–217.

Schweitzer, M. J. (2009). Extracting the mean size from a series of circles. Unpublished Honours thesis, University of Regina, Regina, Canada.

Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, *70*, 1335–1336.

Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature*, *386*, 69–73.

Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, *78*, 392–402.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412.

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*, 47–59.

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science*, *238*, 778–780.

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, *82*, 52–63.