Statistical summary representations of bound features

Aysecan Boduroglu¹ · Irem Yildirim²

Published online: 2 January 2020
© The Psychonomic Society, Inc. 2020

Abstract

The visual system can efficiently summarize various lower-level and higher-level features of ensembles. However, no research to date has directly investigated how different features interact with each other within a single summary and whether people can efficiently report an integrated summary of two feature dimensions. In the first two experiments, we specifically investigated whether individuals can integrate spatial and size information to report a bound spatial summary, the *center of mass* (CoM), as efficiently as the *centroid*, which is devoid of size information. Both experiments revealed that viewers were equally accurate in extracting the centroid and the CoM, with the size distribution inadvertently affecting the centroid estimates. In the final experiment, we investigated whether encouraging observers to attend to individual item size would cause the centroid estimates to be biased toward the CoM. When item size was task-irrelevant, as in the centroid task, observers were able to selectively focus on spatial location, eliminating any impact from the size distribution. Taken together, these findings demonstrate that viewers are capable of extracting integrated summaries, possibly through a mechanism that allows them to represent the spatial distribution of sizes. We discuss possible mechanisms that may support the extraction of integrated summaries, and highlight the need to consider multilevel mechanisms extending beyond simple feature- and object-based mechanisms.

Keywords Visual perception · Ensemble perception · Binding · Statistical summary representations

We experience the world in a continuous and holistic fashion, despite the severe limitations of our attentional and working memory systems (Noë, Pessoa, & Thompson, 2000). Anne Treisman's work has contributed to our understanding of the capacities and limitations of our visual system. Specifically, her feature integration theory proposed that the visual system bypasses the limitations of the feature-processing modules by utilizing spatially selective attentional mechanisms to bind features into more durable object-file representations (Kahneman, Treisman, & Gibbs, 1992; Treisman, 1988, 1996; Treisman & Gelade, 1980). Much early research in visual processing was devoted to understanding how different features are integrated to form object representations (for reviews, see Luck, 2008; Scholl, 2001). However, our increased understanding of how feature binding works in objects has not directly addressed the issue of our phenomenological experience of a rich and detailed visual environment. During the last two

decades, researchers have been drawn to the notion of ensemble perception, arguing that summary representations of encountered scenes, along with high-resolution representations of a few objects, contribute to such rich phenomenological experiences (Cohen, Dennett, & Kanwisher, 2016). Treisman's work on ensemble perception and on the role of attention in summarizing has been critical in shaping the key questions in this literature (e.g., Chong & Treisman, 2003, 2005a, 2005b; Emmanouil & Treisman, 2008; Treisman, 2006).

Research has documented how the visual system can efficiently summarize various lower-level and higher-level features of ensembles (for a review, see Whitney & Yamanashi Leib, 2018), with intraindividual differences revealing separate mechanisms operating at each of these levels (Haberma, Brady, & Alvarez, 2015), and possibly for lower-level dimensions such as size and orientation (e.g., Yörük & Boduroglu, 2019, in this issue). These rapid and highly accurate statistical summary representations are functional, in that they have been shown to contribute to the formation of individual object representations (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Mutlutürk & Boduroglu, 2014).

One major question for ensemble perception research has been identifying the role of attention in summarization. Whereas some have shown that selective attention is *not*

✉ Aysecan Boduroglu
aysecan.boduroglu@boun.edu.tr

¹ Bogazici University, Istanbul, Turkey

² University of Delaware, Newark, DE, USA

necessary for viewers to summarize information (Alvarez & Oliva, 2008, 2009; Bronfman, Brezis, Jacobson, & Usher, 2014; Chong & Treisman, 2003, 2005a; Oriet & Brand, 2013; see Jou et al., 2004; Wilken & Ma, 2004, for summarization in the memory domain), others have argued that attention facilitates ensemble summaries (Chong & Treisman, 2005b; de Fockert & Marchant, 2008; Jackson-Nielsen, Cohen, & Pitts, 2017). Although these general conclusions hold, more recent research has further established that the role of attention in summarization increases as the structure of displays becomes more complex, with multiple perceptual groups and specifically when viewers are asked to report the perceptual group as opposed to entire-set summaries (for a review, see Whitney & Yamanishi Leib, 2018; see, e.g., Yildirim, Ögreden, & Boduroglu, 2018). There have also been claims that when viewers have to summarize more than one feature, additional attentional demands may be present (Emmanouil & Treisman, 2008; but see also Attarha & Moore, 2015). However, no research to date has directly investigated how different features interact with each other within a single summary and whether people can efficiently report an integrated summary of two feature dimensions. In a series of experiments, we specifically investigated whether individuals can integrate spatial and size information to report a bound spatial summary (Exps. 1 and 2), the center of mass, as efficiently as the centroid, which is devoid of size information (see Fig. 1), and whether encouraging observers to attend to the size distribution leads to a bias in the centroid responses toward the center of mass (Exp. 3).

Considerable research has investigated how viewers can simultaneously extract multiple summaries. Although early studies argued that when attention is divided across sets, no costs are associated with summarizing two sets along the same dimension (e.g., size; Chong & Treisman, 2005b), these findings were explained by methodological confounds; once these confounds were eliminated, costs were reported for concurrent averaging (Brand, Oriet, & Tottenham, 2012). Furthermore, Attarha and colleagues also reported that there were costs to averaging size, with larger costs being associated with four

Fig. 1 Locations of the center of mass and the centroid in a uniquely sized display. The *center of mass* is the average of the size-weighted central coordinates of all circles. The *centroid*, in contrast, is the average of the equally weighted central coordinates of all circles. The figure illustrates how the center of mass is pulled toward the larger items

simultaneously presented perceptual groups than with two pairs of sequentially presented ones (Attarha, Moore, & Vecera, 2014), highlighting that capacity limits are driven by how attention is visually divided. Research from our own lab, using a method of adjustment instead of forced choice procedures used in these earlier studies, also demonstrated that when viewers were presented with spatially segregated perceptual groups, precues improved perceptual group summaries of size, but summaries of the entire sets were not impacted by explicit manipulations of attention (Yildirim et al., 2018). We further found that when greater uncertainty was associated with a perceptual group summary (e.g., in a postcue as opposed to a precue condition), there was a greater bias toward the entire-set mean, suggesting that entire-set summaries may be readily available to the visual system. These findings as a whole suggest that the presence of perceptual groups within sets does not interfere with the averaging of entire sets (Brady & Alvarez, 2011); nevertheless, summarizing separate perceptual groups on the same featural dimension may be associated with attentional costs.

A separate set of studies investigated whether the attentional costs associated with concurrent averaging were partly due to the fact that each perceptual group was summarized on the same dimension. When viewers were asked to summarize two spatially segregated yet otherwise similar sets on one of two separate dimensions (either size and speed or size and orientation), significant precue advantages emerged, suggesting that attentional costs were associated with concurrently summarizing two perceptual groups on two separate dimensions (Emmanouil & Treisman, 2008). However, Attarha and Moore (2015) pointed out that in the postcue condition of Emmanouil and Treisman's study, viewers had to extract two within-feature summaries for two spatially segregated sets, consequently increasing the reported precue advantage. In fact, they demonstrated that, where the extraction of within-feature summaries was capacity-dependent (see also Poltoratski & Xu, 2013), the extraction of between-feature summaries (i.e., two summaries of independent features) was not limited by capacity.

In all these studies, the focus was on how dividing attention to regions and/or features impacted the summarization process. However, none of the studies investigated the role of attention in the formation of summaries integrating the properties of two dimensions of a single set of objects. By utilizing the difference between two different types of spatial summaries, centroid and center of mass, we were able to investigate whether viewers could form summaries of integrated feature dimensions as accurate as the summaries formed for single dimensions. The centroid of a display has been operationalized as the average of the central coordinates of a set of objects or the midpoint of a configuration (e.g., Alvarez & Oliva, 2008; Boduroglu & Shah, 2014). Unlike the centroid, for which the central coordinates of each item are weighted

equally, the center of mass has generally been operationalized as a weighted summary in which the centroid is pulled toward larger and farther away from smaller objects (see Fan, Turk-Browne, & Taylor, 2016, for a similar distinction between centroid vs. center of mass during perceptual learning). To our knowledge, only in one other, very recently published study were participants asked to generate the center of mass; they were instructed to report the size-weighted centroid on those trials (Rodriguez-Cintrón, Wright, Chubb, & Sperling, 2019). In our study, we chose not to specify to participants how the center of mass should be determined; they could possibly have used either the circumference or the area of each circle as the weight for each central coordinate. Instead, we were interested in participants' intuitive estimations of the center of mass.

If viewers could report the center of mass as well as they could report the centroid of uniquely sized circles, this would suggest that they are capable not only of processing one-dimensional summaries of objects, but of summarizing independent features in an integrative fashion. In the first experiment, we asked participants to encode briefly presented sets that consisted of either homogeneously or uniquely sized circles. Viewers reported either the centroid or the center of mass of these sets. We compared the errors across these conditions and tested whether there was a bias toward the alternative spatial summary, to determine whether one type of spatial summary was more readily available and was pulling the summary response toward itself. In Experiment 2, we investigated whether there were temporal differences in the extraction of these two types of spatial summaries. These two studies demonstrated that center-of-mass information was represented as efficiently as centroid information, even when sets were viewed for only 50 ms with backward masking. In the final experiment, we compared centroid estimations across two conditions to determine whether task conditions requiring selective attention to circle size would increase the pull exerted by the size distribution, and so bias the centroid estimates toward the center of mass. However, we did not find such evidence; instead, we found that the influence of the center of mass on the centroid responses was the same, regardless of the incentives to focus on the size distribution, suggesting possible parallel summarizing of both spatial summaries.

Experiment 1

Method

Participants Twenty-one undergraduates from Bogazici University (15 female, six male; $M_{\text{age}} = 22.29$ years, age range: 19–32 years) took part in the first experiment in exchange for course credit. All participants reported normal or corrected-to-normal vision.

Materials and stimuli The experiment was programmed in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) and presented on 17-in. monitors with their resolution set to 640×480 pixels. Participants were seated 57 cm away from screens, so that 1° of visual angle corresponded to approximately 20 pixels.

The experiment consisted of centroid and center-of-mass (CoM) estimation tasks. These two tasks were blocked and counterbalanced across participants. In each block there were 96 uniquely sized and 48 homogeneously sized sets, resulting in 144 trials in total. In both conditions, the correct response was in the same quadrant in no more than three consecutive trials. At the beginning of each block, participants were informed about the particular task. The centroid was defined as the mean center of all the circles' centers. The CoM was defined as the *balanced* center of all circles' centers. For both tasks, participants completed an instruction-and-familiarization phase before starting the experimental trials. They observed some examples for which they did not generate any responses. These examples consisted of homogeneously sized circle sets, and the sets' spatial summary was the same point for the centroid and the CoM. We showed sets of two to eight circles in the first eight examples, and sets of nine circles in the following eight examples. The displays were presented for 1 s, and a red asterisk indicating the spatial summary appeared together with the circles for another 1 s. For the CoM estimation task, we additionally instructed observers that when objects are uniquely sized, the bigger objects pull the CoM toward themselves more than smaller objects do. Following the instructions, four additional examples were presented. The first example started with two same-sized circles, and the CoM was indicated with a red asterisk. In the subsequent examples, one of the circles got larger and a red arrow appeared showing the direction and approximate distance that the CoM should move. We specifically did not train participants with uniquely sized sets, to capture participants' natural tendency to deal with variation in size in determining the CoM.

Each trial began with a fixation mark, a plus sign, that was presented at the screen center for 500 ms. After the fixation, a circle set appeared for 200 ms. A blank response screen immediately followed the set and stayed until a response had been given. During this response screen, the mouse cursor was visible and always positioned at the center. Participants responded by moving the cursor to the sought-out location, and once they were certain, they made a right click. Participants immediately saw a red asterisk appear at the location they had specified. Following the response, participants were asked whether their report was a guess or a "remember" response. They were informed beforehand that the "guess" option should be used only when their report was totally random because they had missed the trial, and all other reports should be "remember." They pressed D to indicate "remember" and K to indicate "guess" responses, with no restriction

on which hand to use. A blank screen was presented during the 1,000-ms intertrial interval. No feedback was presented during the experimental trials.

In each study display there were nine black-filled circles on a gray background. The edges of the display screen were always left empty (2° on the right and left, and 1° on the top and bottom). The circles were pseudorandomly positioned on the rest of the screen, with the following constraints. First, we kept at least 1° distance between the circumferences of any two circles. Second, a distance of at least 1.75° separated the centroid, the CoM calculated by circumference, the CoM calculated by area, and the screen center (note that, for homogeneously sized sets, the former three corresponded to the same location; see the Results and Discussion section for the formulas). Third, across trials, the probed spatial summary corresponded equally to each quadrant of the screen. The diameters of circles were chosen randomly to be between 1.3° and 4.3° . Within this size range, at least a 0.2° size difference is needed between circles so that people can perceive the circles as being differently sized. We calculated this just noticeable difference value on the basis of a function between perceived and physical sizes with an exponent of 0.76 (Teghtsoonian, 1965). Therefore, for uniquely sized sets, each circle diameter was selected from the range above, with 0.2° increments. The sample variance of a uniquely sized set ranged from 6° to 14.2° . For homogeneously sized sets, the diameters of all circles in a set were equal and were selected from the same range, between 1.3° and 4.3° with 0.2° increments. The possible circle sizes in homogeneously sized sets were presented equally frequently within blocks.

Results and discussion

In the centroid block, for each trial, we computed the distance between the actual centroid location (as determined by Eq. 1 below) and the response. For the CoM blocks, for each display we used both formulas (Eqs. 2 and 3 below: weighted by circumference vs. weighted by area) to compute two potential CoM coordinates: (x_{com}, y_{com}) and (x_{acom}, y_{acom}) , respectively. We determined the error in each trial by calculating the distance between the responses and both sets of CoM coordinates. A paired sample t test revealed that errors were similar, regardless of which CoM coordinates were taken as the correct location, $p = .7$. We also checked whether responses were biased toward either of the CoM locations. To do this, for each trial we computed the proportion of the error to the CoM weighted by circumference to the error to the CoM weighted by area. If participants did not have any particular bias to either location, then we would expect this proportion to be around 1; values greater than 1 would indicate a bias toward the CoM weighted by circumference, and values less than 1 would indicate a bias toward the CoM weighted by area. A one-sample t test with a criterion set to 1 showed that

participants' estimates were significantly closer to the CoM weighted by circumference ($M = 1.54$, $SD = 0.5$), $t(20) = 4.95$, $p < .001$, $d = 1.08$, suggesting that participants were weighing the center coordinates by the circumferences of circles, rather than by their areas, when estimating the CoM. This finding is in fact in line with earlier research. Solomon, Morgan, and Chubb (2011) reported that when viewers were computing mean size, they relied on size as opposed to area (luminance). Similarly, Allik, Toom, Raidvee, Averin, and Kreegipuu (2014) demonstrated that the reported mean size of a set of circles approached the mean circumference and not the area. The very recent work by Rodriguez-Cintron, Wright, Chubb, and Sperling (2019) also demonstrated that when viewers were determining weighted centroid locations, they relied on the sizes of squares rather than their areas. In light of these results, in the remainder of the analyses, for the CoM calculations we relied on Eq. 2:

$$x_c = \frac{\sum x_i}{n} \quad y_c = \frac{\sum y_i}{n} \quad (1)$$

$$x_{com} = \frac{\sum r_i x_i}{\sum r_i} \quad y_{com} = \frac{\sum r_i y_i}{\sum r_i} \quad (2)$$

$$x_{acom} = \frac{\sum r_i^2 x_i}{\sum r_i^2} \quad y_{acom} = \frac{\sum r_i^2 y_i}{\sum r_i^2} \quad (3)$$

To measure the accuracy of the spatial summary estimations, for each trial we calculated the absolute error by finding the distance between the reported and the correct summary location, in visual degrees, and determined the average error per condition. We excluded participants' trials that were above three standard deviations from the error distribution for that particular condition. "Remember" responses consisted of 95%, 97%, and 97% of the responses for the CoM, centroid, and homogeneous conditions, respectively, and these values did not differ, $F = 1.02$, $p = .37$. Here we report the results obtained from "remember" trials only.

First we analyzed the homogeneously sized sets across different blocks that required different spatial centers. We wish to highlight two things about the homogeneously sized sets again: (1) There is only one possible spatial center for homogeneous sets, and thus the centroid and the CoM corresponded to the same spatial center, and (2) the numbers of homogeneous sets were equal across two blocks (i.e., 48 trials in each), and their total number was equal to the number of uniquely sized sets in each block (i.e., 96 trials in each). We compared the absolute errors in visual degrees for homogeneously sized sets between two blocks. People made similar amounts of error in finding the spatial centers of homogeneous sets when the center asked for was the centroid ($M = 1.68$, $SD = 0.46$) or the CoM ($M = 1.59$, $SD = 0.31$), $t(20) = 1.04$, $p = .31$, $d = 0.23$. Also, we computed a Bayes factor (BF_{10}) to contrast the likelihood of our data fitting the alternative hypothesis (i.e., significant difference) with the likelihood of our data fitting the

null hypothesis (i.e., no difference). The BF_{10} was 0.37,¹ suggesting that the null hypothesis was 2.71 times more probable than the alternative hypothesis, given our data. These analyses indicate that for homogeneously sized sets, asking for either the centroid or the CoM in an intermixed fashion did not impact accuracy. Thus, we collapsed these two blocks with homogeneously sized sets and used them as a control condition for the main analyses.

To test our main question of interest, we compared the errors for the CoM, centroid, and homogeneous/control conditions via a one-way repeated measures analysis of variance (ANOVA) (Fig. 2a). We found a significant effect of condition, $F(2, 40) = 20.67$, $MSE = .13$, $p < .001$, $\eta_p^2 = .51$. Pairwise comparisons² revealed that for uniquely sized sets, viewers made similar levels of errors (in visual degrees) for the centroid ($M = 2.32$, $SD = 0.49$) and CoM ($M = 2.15$, $SD = 0.49$) conditions, $p = .6$, $d = 0.29$; in both these conditions, the error was significantly higher than in the homogeneously sized ($M = 1.63$, $SD = 0.34$) sets, both $ps < .001$, both $ds > 1.28$. When block order was added as a between-subjects variable, we found no main effect nor an interaction involving block order, $p > .6$. A Bayesian ANOVA was also in line with these results, $BF_{10} = 56,879.24$ for the main effect model. Pairwise tests showed that the posterior odds favored the null hypothesis for the difference between centroid and CoM, $BF_{10} = 0.29$. However, the odds were higher for the alternative hypothesis for the comparisons with the homogeneously sized sets ($BF_{10} = 1,368.47$ for CoM vs. homogeneous, $BF_{10} = 1,974.67$ for centroid vs. homogeneous), providing strong support and replicating earlier findings that the distributional properties of irrelevant features impact the accuracy of summary statistics (e.g., Chong & Treisman, 2003).

To determine whether one spatial summary was more readily available, and thus more likely to create a pull toward itself, we also calculated a bias score for uniquely sized set trials. For each trial, the distance of the reported location from the other (nonprobed) spatial center was divided by the distance from the probed center. That is, for the CoM trials, the distance to the centroid was divided by the distance to the CoM, and vice versa in the centroid block. A value of 1 indicated no particular bias, a score below 1 indicated a bias toward the other (nonprobed) spatial summary, and a score above 1 indicated a bias toward the probed spatial center. The bias scores from uniquely sized trials were averaged for each block. Comparison of the bias scores against the no-bias point (1, in this case) via one-

sample t tests suggested that viewers were biased toward the required center (see Fig. 2b). That is, their responses were biased toward the CoM rather than the centroid when the task was to estimate the CoM ($M = 2.11$, $SD = 0.44$), $t(20) = 11.48$, $p < .001$, $d = 2.51$, and toward the centroid when the task was to estimate the centroid ($M = 1.4$, $SD = 0.53$), $t(20) = 3.49$, $p = .004$, $d = 0.76$. There was also a significant difference between the bias scores, $t(20) = 4.61$, $p < .001$, $d = 1.01$, suggesting that the centroid reports were more likely to be influenced by the CoM than the CoM was likely to be influenced by the centroid. When added as a variable, block order (CoM first vs. centroid first) did not affect the bias scores, $p = .34$, nor did it interact with the difference between the CoM and centroid, $p = .97$. A Bayesian analysis also provided strong evidence in line with these results. The ratio of odds favored the alternative hypothesis that the bias toward the required center was different from 1 (i.e., the no-bias point) for both the CoM, $BF_{10} = 3.392 \times 10^7$, and the centroid, $BF_{10} = 17.5$, and that the bias toward the correct over the other center was different for the CoM and the centroid, $BF_{10} = 173.9$.

Overall, Experiment 1 demonstrated that observers were not able to extract the spatial summary of items in uniquely sized sets as precisely as they did for displays that consisted of homogeneously sized elements (e.g., Chong & Treisman, 2003; Fan et al., 2016). More critically, though, viewers were able to extract an integrated summary of size and spatial features within 200 ms.³ The similar levels of error in the centroid and the CoM trials suggest that no additional cost was associated with attending to size information in the CoM block as compared to the centroid block. However, the bias results implied that it may be difficult for viewers to ignore size information when asked solely to report the centroid. Although the bias scores showed that viewers were attending to the task demands (given the greater pull toward the required spatial summary), the difference in the bias scores suggested that the centroid responses were inadvertently being impacted by size information. Thus, even when the tasks were blocked, viewers were not able to completely ignore size information while estimating the centroid.

Critical to our question of interest, the CoM performance observed in Experiment 1 could not have been achieved by merely integrating separate summaries of the size and spatial distributions, because an identical set of items (with identical summaries of size) could have yielded different CoM summaries based on where items were positioned (for a similar argument with scatterplots, in which identical summaries emerge from different distributions, see Anscombe, 1973).

¹ All the Bayes factors for t tests in this article were scaled with Jeffrey–Zelner–Siow (Cauchy) prior with a .707 scaled r on effect size, to approach the most objective results (Rouder, Speckman, Sun, Morey, & Iverson, 2009). All Bayes factors for pairwise comparisons were corrected for multiple tests by fixing the prior probability to .5 for the null hypothesis being true for all comparisons (Westfall, Johnson, & Utts, 1997).

² All p values for paired-samples and one-sample t tests in this article were Bonferroni-corrected for multiple comparisons.

³ In a task focusing on the roles of different types of feedback on perceptual learning, Fan et al. (2016) also reported better-than-chance accuracies for centroids and CoM responses at similar presentation durations. We also found better-than-chance accuracies for the match between the reported and the correct quadrant of the screen.

Fig. 2 Results of Experiment 1, showing the averages of (a) errors across conditions and (b) bias in uniquely sized displays ($N = 21$). Error bars indicate 95% confidence intervals. * $p < .001$.

Even though our research was targeting whether integrated summaries could be extracted, as opposed to *how* they might be extracted, post-hoc analyses helped us eliminate at least one potential mechanism. We specifically tested whether viewers might have extracted the centroid and shifted that centroid toward the largest item presented in the display. Such a strategy would be highly feasible following a 200-ms encoding duration, given that summaries of unmasked displays can be extracted in periods as short as 50 ms (Chong & Treisman, 2003), objects can be consolidated in approximately 50 ms (Vogel, Woodman, & Luck, 2006), and three or four locations can typically be attended within that duration (e.g., Xu & Chun, 2006). To determine whether the CoM responses we recorded were actually biased toward the centroid shifted toward the largest item, we calculated a new summary point for each CoM trial by only weighing the largest item and keeping the weights of all other items at 1. Specifically, we adjusted the weight of the largest item by the ratio of its size to the mean size of the set. This allowed us to calculate a new bias score for each trial. We took the ratio of the reported location's distance from the largest-weighted center to the reported location's distance from the CoM. Thus, a ratio larger than 1 would indicate a correct bias toward the CoM rather than the largest-weighted centroid. This is exactly what we found; the bias was significantly larger than 1 ($M = 1.64$, $SD = 0.08$), $t(20) = 8.33$, $p < .001$, $d = 1.82$, indicating that participants' reports were correctly pulled toward the CoM and that participants were not merely weighting the largest item in their reports. However, the bias toward the CoM was smaller when calculated against the distance from largest-weighted centroid than against the distance from the centroid ($M = 2.11$, $SD = 0.04$), $t(20) = 6.07$, $p < .001$, $d = 1.32$. This suggests more of an influence of the largest-weighted centroid on the CoM estimations than an influence of the summary centroid. This finding is not surprising, given that the largest-weighted centroid is associated with size information from one object in the set, as opposed to no size information being associated with the summary centroid; hence, the largest-weighted centroid is closer to the CoM.

Experiment 2

Experiment 1 demonstrated that viewers are capable of extracting bound summaries. In Experiment 2 we wanted to determine whether both spatial summaries were extracted equally rapidly and efficiently from the visual displays. First we manipulated encoding durations, to test whether the pattern of results obtained in the first experiment would replicate with shorter durations of 50 and 100 ms, as well as with 200 ms (as in Exp. 1). Second, we used backward masking by presenting a noise mask following the circle sets, to prevent any persistence of afterimages. The bias results from Experiment 1 suggested that size information may have been hard to ignore during the centroid estimations; this may have been partly due to the available object representations and/or the rapidly established and readily available CoM summaries creating a subtle pull toward the CoM. Experiment 2 tested the latter possibility.

Method

Participants The participants were a different set of 24 Bogazici University undergraduates (14 female, 10 male; $M_{\text{age}} = 21.83$ years, age range: 19–26 years), who completed the experiment in exchange for course credit. All reported normal or corrected-to-normal vision. We excluded the responses from two participants who never said “remember” in at least one of the conditions, reducing the sample size to 22 (13 female, nine male).

Materials and stimuli All materials and stimuli were identical to those of Experiment 1, with the following exceptions: (1) We only used uniquely sized displays; (2) we introduced a white noise mask for 200 ms, presented at the offset of the study display; and (3) we manipulated the encoding duration to be either 50, 100, or 200 ms. The whole experiment consisted of four blocks, half of which were the CoM trials, and the rest of which were the centroid trials. The two blocks of the same task, CoM or centroid, always followed each

other, and the task order was counterbalanced across participants. There were 96 trials in each block, with encoding durations equally and randomly intermixed within a block, and with the constraint that no more than three consecutive trials had the same encoding duration.

Results and discussion

Even though the percentage of “remember” responses was less than in Experiment 1, participants still reported more “remember” than “guess” responses (for CoM: 79% for 50 ms, 81% for 100 ms, and 84% for 200 ms; for centroid: 83% for 50 ms, 85% for 100 ms, and 87% for 200 ms). This decrease might have been due to the increased task difficulty induced by backward masking and the shorter encoding durations. A 2 (task: CoM vs. centroid) $\times$ 3 (duration: 50 vs. 100 vs. 200 ms) ANOVA on the percentages of “remember” responses revealed a main effect of duration, $F(1.48, 34.1) = 7.08$, $MSE = .003$, $p = .006$, $\eta_p^2 = .24$. Pairwise comparisons showed that participants were significantly less confident for 50-ms trials than for 200-ms trials, $p = .018$. The 100-ms condition did not differ from the other two conditions in “remember” percentage, $ps > .06$. We observed neither a main effect of task, $F(1, 23) = 1.56$, $p = .32$, nor a significant interaction, $F < 1$, $p = .63$. As in the first experiment, the following analyses were carried out on the data from “remember” trials only. We followed the same steps as in Experiment 1, to calculate the error per condition and to clean outliers. As in Experiment 1, the errors were similar for the CoM estimations no matter how CoM was calculated, $p > .11$ for all encoding durations. But when we analyzed the proportion of errors with CoM weighted by circumference to errors with CoM weighted by area, we found that participants’ estimates were closer to the former value for all encoding durations, all $ps < .001$. Therefore, for the remainder of the analyses we focused on the error data computed by taking the CoM weighted by circumference as the correct answer.

To compare the errors (in visual degrees) between centroid and CoM trials across the three encoding durations, we conducted a 2 (task: CoM vs. centroid) $\times$ 3 (duration: 50 vs. 100 vs. 200 ms) repeated measures ANOVA (Fig. 3a). We found a significant effect only for the main effect of task, $F(1, 21) = 6.32$, $MSE = .51$, $p = .02$, $\eta_p^2 = .23$. Unlike in Experiment 1, the CoM reports ($M = 1.49$, $SD = 0.61$) were less erroneous than the centroid reports ($M = 1.54$, $SD = 0.9$). There was a marginal effect of encoding duration, $F(2, 42) = 3.07$, $MSE = .044$, $p = .057$, $\eta_p^2 = .13$; however, none of the post-hoc Bonferroni-adjusted pairwise comparisons were significant, all $ps > .1$. The encoding duration by task type interaction did not reach significance, $F(2, 42) = 2.07$, $MSE = .037$, $p = .14$, $\eta_p^2 = .09$. The pattern did not vary on the basis of which task was completed first, $p > .5$. Critically, a Bayesian repeated measures ANOVA (see Wagenmakers et al., 2018, for

applications) did not provide critical support (only anecdotal support) for the model with only the main effect of task type, with a BF_{10} value of 1.21, suggesting that the difference between the two tasks may be negligible.

To ensure that we were able to replicate the pattern observed in Experiment 1 with backward masking, we analyzed the data from the 200-ms duration trials separately, and found that the CoM ($M = 2.54$, $SD = 0.67$) and the centroid ($M = 2.75$, $SD = 0.97$) reports were equally accurate, $t(21) = 1.38$, $p = .18$, $d = 0.32$. This null finding was also supported through Bayesian paired samples, with test results favoring the null, $BF_{10} = 0.22$. Taken together, these findings suggest that backward masking did not have a detrimental effect on either spatial summary, and that both were efficiently extracted within the first 50 ms. This is a very fast form of summary extraction. For mean size, for example, Whiting and Oriet (2011) found that 200 ms was needed to form accurate representations when there was backward masking. However, summarizing may be distinct for location versus nonlocation domains (Hubert-Wallander & Boynton, 2015), and spatial summaries might be extracted more rapidly by the visual system than are size summaries. Indeed, in a separate series of experiments in our lab, we had demonstrated that there was no difference in centroid errors for equally sized squares presented for 50, 75, and 100 ms⁴ (Mutlutürk & Boduroglu, 2013). Our Experiment 2 results extend these findings to CoM summaries.

As in Experiment 1, we also calculated bias scores to determine whether one spatial summary might be more likely to pull the response toward itself. As can be seen in Fig. 3b, all bias scores were above 1 (the no-bias point). We compared the bias in each condition against the no-bias value of 1 via one-sample t tests. As in Experiment 1, all bias scores were significantly different from 1, all $ps \leq .05$ for all conditions, indicating that viewers were correctly biased toward the probed spatial center. A 2 $\times$ 3 ANOVA on bias scores revealed a significant main effect only for task, $F(1, 21) = 34.83$, $MSE = .32$, $p < .001$, $\eta_p^2 = .62$. As in Experiment 1, the CoM reports were correctly pulled toward the CoM ($M = 1.83$, $SD = 0.4$) more than the centroid reports were correctly pulled toward the centroid ($M = 1.30$, $SD = 0.44$). The pattern of results did not change when task order (CoM first vs. centroid first) was entered into the analyses, $p > .13$. This bias difference between the CoM and centroid tasks was further confirmed with Bayesian ANOVA findings that yielded extremely strong evidence for the task main effect model, $BF_{10} = 6.534 \times 10^{11}$.

⁴ Viewers encoded a set of seven randomly positioned, same-sized colored squares and estimated the centroid across four encoding conditions. A comparison of errors in estimating the centroid yielded an effect of encoding duration, $F(3, 75) = 5.72$, $MSE = 14.24$, $p = .001$, $\eta_p^2 = .19$, due to lower errors in the 500-ms encoding condition ($M = 19.14$, $SD = 6.79$) than in the other encoding duration conditions (50 ms, $M = 22.46$, $SD = 5.38$; 75 ms, $M = 23.16$, $SD = 7.52$; 100 ms, $M = 22.08$, $SD = 6.05$), $p < .028$. Critically, the precision was not different in any of the remaining conditions, $p = 1$.

Fig. 3 Results for Experiment 2, showing (a) average errors and (b) average bias across the different tasks and encoding durations ($N = 22$). Error bars indicate 95% confidence intervals. * $p < .02$, ** $p < .001$

Also, our data were 13.25 times more likely under the task main effect model versus the null than under the second most successful, two-main-effect model versus the null, $BF_{10} = 4.931 \times 10^{10}$. Thus, at all durations, centroids were impacted by the size information, even though the size information was task-irrelevant.

All in all, Experiment 2 extended Experiment 1's findings to a shorter duration of 50 ms, and demonstrated that both types of spatial summaries can be extracted within a 50-ms window, even when stimuli are masked. Given the lack of costs at the shorter durations, object-based contributions to spatial summarization are unlikely. What remains noteworthy is that even at 50 ms, the centroid responses were impacted by the size distribution, suggesting that the size distribution is hard to ignore. This may be partly due to the more demanding nature of the task at the shortest encoding duration. Previous work has shown that, as tasks become more demanding, observers rely more on the automatic averaging process, which consequently results in a reduction of summarization error (e.g., Bauer, 2017; Dubé & Sekuler, 2015; Dubé, Zhor, Kahana, & Sekuler, 2014).

These findings could also be explained by the possibility that size and location are integral dimensions (Garner, 1974, 1976), and consequently that performance on tasks that require the combined processing of these two features may be facilitated. Adding to previous research that has claimed that certain features, such as mean size and orientation, are summarized in an obligatory fashion (e.g., Allik et al., 2014; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), our findings suggest that a pair of integral dimensions may be obligatorily processed.

Experiment 3

The first two experiments demonstrated that viewers can efficiently extract integrated summaries even at durations as short as 50 ms (masked), and that size information affects the centroid responses even when it is task-irrelevant. To better understand the impact of the size distribution on centroid

estimation and determine whether size information gets obligatorily processed during centroid extraction, we manipulated the context in which we asked viewers to estimate the centroid of a set of uniquely sized circles. In the single-task block, viewers reported the centroid only. In mixed-task block, they either reported the centroid or the location of one of the items, indicated by postcues (centroid, location of smallest item, or location of largest item). Thus, in the mixed blocks, the task demanded that viewers attend to at least two objects and the centroid simultaneously. In fact, to be able to identify the smallest and largest items, viewers most likely had to process the size distribution of the items. It is known that observers can implicitly code stimulus range (e.g., Khayat & Hochstein, 2018). If viewers' centroid responses were more biased toward the CoM during the mixed-task than during the single-task trials, this would suggest that the bias results in the earlier experiments might have been due to viewers paying attention to the size distribution and to particular item sizes. On the other hand, if centroid responses were similarly erroneous and the bias toward the CoM was equivalent across the two contexts, that would suggest that even when there is no incentive to process the size distribution (single-centroid-task block), the effective ensemble mechanism processes the centroid and CoM, possibly in parallel. Furthermore, if the errors in the mixed and single centroid blocks were found to be similar, this would also suggest that attending to the size distribution does not interfere with ensemble perception, adding to the evidence that selective attention may not be necessary for centroid extraction (Alvarez & Oliva, 2008).

Method

Participants We recruited a different set of 29 Bogazici University undergraduates (24 female, five male; $M_{\text{age}} = 21.38$ years, age range: 19–26 years) to take part in the experiment in exchange for course credit. All had normal or corrected-to-normal vision.

Materials and stimuli Most of the task details were the same as in Experiment 1, except for the following changes. First, the

circles in a set were always uniquely sized. Second, and most critical to the present experiment, three blocks of 80 trials were presented, in one of which participants were asked to report the centroid only. In the other two blocks, participants had to report either the centroid (total of 80 trials) or the location of the largest (40 trials) or the smallest (40 trials) circle; these trials were randomly intermixed. The instructions for the largest- and smallest-item-location trials emphasized that participants should target the circle's center. No more than three successive trials asked for the same task. In these mixed blocks, following the study display at the beginning of the response screen, we provided an auditory postcue indicating the required task for that trial. Thus, participants had to attend to the centroid and to the locations of the largest and smallest items during the viewing of the set. The two mixed blocks were always completed consecutively, and the order of the single centroid-estimation task and the mixed task was counterbalanced across participants. The encoding duration was fixed at 200 ms.

Results and discussion

Participants mostly reported “remembering” the study display while responding. Specifically, they said “remember” in 90% of the trials in the single-centroid block, and in 93% of the trials asking for the centroid in the mixed-task block. For the largest/smallest item location trials, participants reported remembering in 82% of trials. Pairwise comparisons showed that the “remember” percentages were similar for the centroid asked for in the mixed blocks and for the centroid asked for in the single-task block, $p = .49$. This indicates that people were similarly confident about their centroid estimations, regardless of whether the centroid task was intermixed with the largest/smallest item location estimation task. However, within the mixed blocks, the “remember” percentage was significantly greater for the centroid trials than for the largest/smallest item location trials, $p < .001$. This suggests that people were more confident about their centroid estimations than about their largest/smallest item location estimations. The remaining analyses were carried out on estimates qualified by a “remember” response.

For each trial, we calculated the error in centroid estimates as in Experiment 1; we used the same criteria to identify any outlier responses. First, and most relevant to our aim, we analyzed bias toward the CoM in the centroid reports. As can be seen in Fig. 4b, the bias scores were all above 1. One-sample t tests showed that bias was significantly different from 1 for both the single-task ($M = 1.62$, $SD = 0.52$) and the mixed-task ($M = 1.62$, $SD = 0.59$) conditions, both $ps < .001$, both $ds > 2.74$. That is, participants were always correctly pulled toward the centroid rather than the CoM. The biases in the single- and mixed-task conditions did not differ significantly, $t(28) = 0.03$, $p = .98$, $d = 0.004$. A Bayesian paired-sample t test on bias scores also provided support for no difference. We found that

our data were more likely to fit under the null hypothesis than under the alternative hypothesis, with a moderate level of evidence, $BF_{10} = 0.2$. Despite the fact that participants were encouraged to attend to size during the mixed blocks, there was no pull toward the CoM in those centroid reports, suggesting that observers were able to ignore the size distribution information. Thus, when people were not instructed to think about the CoM during the familiarization phase, they did not necessarily create spatial summaries incorporating size information, even though they were encouraged to detect the largest/smallest item. This suggests that the task specifics may impact whether or not the CoM is effortlessly extracted.

Second, we compared the errors in the centroid estimations, measured in visual degrees (Fig. 4a), across the single-task ($M = 2.29$, $SD = 0.91$) and mixed-task ($M = 2.21$, $SD = 0.67$) conditions, and found that the errors were similar, $t(28) = 0.85$, $p = .403$, $d = 0.17$. A Bayesian t test converged with these analyses; the odds favored the probability of the null hypothesis against the alternative hypothesis, $BF_{10} = 0.28$. This result suggests that the centroid can be extracted as attention is simultaneously directed to particular locations.

Finally, we analyzed the mixed-task condition separately, to see whether the centroid was represented better than the largest/smallest item locations. The centroid reports were less erroneous than the largest/smallest item location reports ($M = 4.11$, $SD = 0.74$), $t(28) = 10.85$, $p < .001$, $d = 2.02$. A very strong Bayes factor also indicated a highly likelihood that the error was different in the centroid reports than in the largest/smallest location reports, given our data: $BF_{10} = 5.842 \times 10^8$. Thus, our results were consistent with general findings that spatial summary statistics can be represented with distributed attention (Alvarez & Oliva, 2008, 2009), and that the precision is lower for item representations than for summary representations (Ariely, 2001; Corbett & Oriet, 2011). We further extended the previous evidence by showing that centroid extraction is also not costly when there is variance in the size of the elements and when attention is split between the centroid and the less precise item locations. Our findings are consistent with recent research suggesting that working memory load has no bearing on ensemble coding (Epstein & Emmanouil, 2017; but see also Bauer, 2017). Taking these results together with those from Experiment 2, we can conclude that the centroid is a simple summary that can be represented upon very brief views and also along with another spatial summary. Nonetheless, in none of the experiments was this simple, one-dimensional spatial summary, the centroid, able to bias CoM estimations.

General discussion

In this study, our key finding is that viewers can efficiently extract spatial summaries integrating size and position information. Considerable research has demonstrated that featural

Fig. 4 Results of Experiment 3, showing (a) average errors and (b) average bias in the centroid reports across the single-task and mixed-task conditions ($N = 29$). Error bars indicate 95% confidence intervals

properties and spatial information can be summarized (e.g., Whitney & Yamanashi Leib, 2018); some research has further specified how features, like mean size and mean orientation, and display characteristics, like the featural range of a set, can be implicitly and obligatorily processed (Allik et al., 2014; Khayat & Hochstein, 2018; Parkes et al., 2001; for a review, see Dubé & Sekuler, 2015). However, to our knowledge, aside from one very recently published study, no research to date has shown that an integrated summary of features is possible (Rodriguez-Cintrón et al., 2019). Critically, the center of mass could not have been accurately identified by separate summaries of two distributions; instead, it required individuals to have some information about where the larger/smaller items were placed. Our results suggest that the CoM was not merely the centroid shifted toward the largest item, because the bias toward the shifted centroid was significantly less than the bias toward the required CoM. Furthermore, the fact that the CoM was estimated with the same level of accuracy as the centroid in the 50-ms condition rendered any contribution from individual items less likely.

The similar levels of accuracy and temporal efficiency for both spatial summaries, and the greater influence of the CoM on centroid estimates suggest that the viewers extracted something akin to a spatial distribution of size (see Alvarez & Oliva, 2009, for a similar argument on spatial patterns of orientation), very rapidly (within 50 ms). Our results also allow for the possibility that both spatial summary representations were concurrently extracted. There was no difference in the errors with which the centroid and CoM were reported; the bias observed could have been the result of the competition between these two types of spatial summaries during response generation. Our bias findings, then, would imply that the CoM representation was a stronger (more active) spatial summary representation than the centroid.

Our findings suggest that the CoM may be extracted preattentively. One possibility is that the visual system extracts the CoM during the initial registry of visual stimuli; during this initial phase outside of selective attention, features are thought to be conjoined implicitly (Treisman, 2006). However, even if this registry holds the CoM information, it

is unlikely that these implicit conjunctions could be accessed explicitly. Another possibility is that the CoM is estimated through texture processing mechanisms. Although the displays used in our experiments are not typical texture displays, the filled-in nature of the circles may have made surfaces salient and led viewers to process them as textures as opposed to sets of individual features (or feature distributions) or objects. If so, in the CoM task, viewers may have identified certain boundaries based on the continuity of textures through the density distribution within the texture and used that information to reach the required summary (Victor, Conte, & Chubb, 2017). Using uniquely sized circle displays very similar to our own, Cain and Cain (2018) recently demonstrated that ensemble perception of sizes may be supported by mid-level texture representations. In the average size tasks, it is known that viewers are biased to choose the set with more items to have a larger average; Cain and Cain (2018) found that only simulations of ideal observers trained on mid-level texture representations mimicked the bias observed in human data. The alternative model, focusing on low-level luminance did not explain this bias. These findings suggest that in our displays, CoM may also have been extracted by texture processing mechanisms; visual textures represent information gathered across spatial regions and allow for statistical inferences (Victor et al., 2017). Regardless of the exact mechanisms involved in extracting the CoM, our findings are consistent with efforts emphasizing the need to go beyond feature-based summary mechanisms (e.g., Jacoby, Kamke, & Mattingley, 2013). Future research is necessary to specify the exact mechanisms of how integrated summaries as the CoM is extracted.

Although the majority of our findings highlight the ease with which CoM information is extracted, the findings from Experiment 3 suggest that CoM information is not determined obligatorily. In Experiment 3, observers reported either the centroid or the location of the largest/smallest item. Critically, there was no bias toward the CoM in the centroid estimates. One key difference between the first two and the third experiment was that in the latter, during the familiarization phase, there was no mention of the CoM. Thus, observers

may have inadvertently inhibited the extraction of an integrated spatial summary. Previously in tasks requiring global motion detection, it has been shown that observers can pay attention to different aspects of a display and/or subsample according to different task demands (Raidvee, Averin, Kreegipuu & Allik, 2011; Zohary, Scase, & Braddick, 1996). In our case, observers may have chosen to focus on the centroid rather than the CoM. This suggests that CoM extraction may be automatic, but not necessarily obligatory.

Beyond the specifics of how spatial summaries are formed and whether integrated summaries of size and location are possible, the present work also adds to the research that has investigated attentional demands on ensemble perception, further providing support for the existence of multiple summary mechanisms. Although some research has shown that costs are associated with summarizing identical features (e.g., Brand et al., 2012; Yildirim et al., 2018; but also see Chong & Treisman, 2003), other research has shown that there are no attentional costs for summarizing different features (e.g., Attarha et al., 2014). Because our experiments did not have a postcue condition in which viewers extracted both the centroid and the CoM in mixed-trial blocks, we cannot directly compare our findings with those from these earlier studies. However, indirectly, the bias results showing how the CoM impacted the centroid performance suggest that these two summaries may have been extracted concurrently. In particular, the similar levels of bias toward the CoM in the single-task and mixed-task conditions in the last experiment suggest that the size distribution was processed even when there was no obvious incentive to do so. Thus, our findings suggest the possibility that two different spatial summaries could be efficiently extracted from a single set of uniquely sized circles, possibly in parallel. The efficiency with which the two spatial summaries were extracted, in terms of both accuracy and temporal window, further supports the claims that there may be multiple summary mechanisms (e.g., Haberman et al., 2015; Yörük & Boduroglu, 2019) and that global attentional mechanisms may be sufficient for spatial summarization (Treisman, 2006).

Acknowledgements This project was supported by a TUBITAK (113K217) grant awarded to A.B. We thank D. Boz and O. Uner for their assistance with data collection, and E. Oral for assistance with the data analyses and interpretation.

References

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, 101, 34–40. doi:<https://doi.org/10.1016/j.visres.2014.05.003>
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, 19, 392–398. doi:<https://doi.org/10.1111/j.1467-9280.2008.02098.x>
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, 106, 7345–7350. doi:<https://doi.org/10.1073/pnas.0808981106>
- Anscombe, F. J. (1973). Graphs in statistical analysis. *American Statistician*, 27, 17–21. doi:<https://doi.org/10.1080/00031305.1973.10478966>
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157–162.
- Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, 77, 1116–1131.
- Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1440–1449. doi:<https://doi.org/10.1037/a0036206>
- Boduroglu, A., & Shah, P. (2014). Configural representations in spatial working memory. *Visual Cognition*, 1, 997–1007.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22, 384–392. doi:<https://doi.org/10.1177/0956797610397956>
- Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, 120, 85–109.
- Brand, J., Oriet, C., & Tottenham, L. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology*, 66, 63–69.
- Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: “Cost free” color phenomenality outside focal attention. *Psychological Science*, 25, 1394–1403.
- Cain, S. S., & Cain, M. S. (2018). *Texture statistics: The mechanism behind ensemble perception in human vision*. Conference on Cognitive Computational Neuroscience, Philadelphia, Pennsylvania. Retrieved from <https://ccneuro.org/2018/proceedings/1157.pdf>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43, 393–404. doi:[https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, 67, 1–13. doi:<https://doi.org/10.3758/BF03195009>
- Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, 45, 891–900. doi:<https://doi.org/10.1016/j.visres.2004.10.004>
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). Ensemble perception, summary statistics, and perceptual awareness: a response. *Trends in Cognitive Sciences*, 20, 643–644.
- Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, 138, 289–301. doi:<https://doi.org/10.1016/j.actpsy.2011.08.002>
- de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, 70, 789–794.
- Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & Psychophysics*, 70, 946–954. doi:<https://doi.org/10.3758/PP.70.6.946>
- Epstein, M. L., & Emmanouil, T. A. (2017). Ensemble coding remains accurate under object and spatial visual working memory load. *Attention, Perception, & Psychophysics*, 79, 2088–2097.
- Fan, J. E., Turk-Browne, N. B., & Taylor, J. A. (2016). Error-driven learning in statistical summary perception. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 42, 266–280.
- Garner, W. R. (1974). The processing of information and structure. Potomac, MD: Erlbaum.
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, 144, 432–446.
- Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, 15(4), 5. doi:<https://doi.org/10.1167/15.4.5>
- Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, 48, 149–160.
- Jacoby, O., Kamke, M. R., & Mattingley, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 233–244. doi:<https://doi.org/10.1037/a0028762>
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24, 175–219. doi:[https://doi.org/10.1016/0010-0285\(92\)90007-0](https://doi.org/10.1016/0010-0285(92)90007-0)
- Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, 18(9), 23. doi:<https://doi.org/10.1167/18.9.23>
- Luck, S. J. (2008). Visual short-term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (pp. 43–85). New York, NY: Oxford University Press.
- Mutlutürk, A., & Boduroglu, A. (2013). *Ensemble representations in spatial working memory*. Poster presented at the Annual Workshop of Object, Perception, Attention and Memory, Toronto, Canada.
- Mutlutürk, A., & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, & Psychophysics*, 76, 2276–2285.
- Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, 7(1–3), 93–106.
- Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, 79, 8–16. doi:<https://doi.org/10.1016/j.visres.2012.12.004>
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744. doi:<https://doi.org/10.1038/89532>
- Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, 13(8), 6:1–11. doi:<https://doi.org/10.1167/13.8.6>
- Raidvee, A., Averin, K., Kreegipuu, K., & Allik, J. (2011). Pooling elementary motion signals into perception of global motion direction. *Vision Research*, 51, 1949–1957. doi:<https://doi.org/10.1016/j.visres.2011.07.004>
- Rodriguez-Cintrón, L. M., Wright, C. E., Chubb, C., & Sperling, G. (2019). How can observers use perceived size? Centroid versus mean-size judgments. *Journal of Vision*, 19(3), 3. doi:<https://doi.org/10.1167/19.3.3>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237. doi:<https://doi.org/10.3758/PBR.16.2.225>
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80, 1–46. doi:[https://doi.org/10.1016/S0010-0277\(00\)00152-9](https://doi.org/10.1016/S0010-0277(00)00152-9)
- Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology*, 78, 392–402. doi:<https://doi.org/10.2307/1420573>
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A, 201–237. doi:<https://doi.org/10.1080/02724988843000104>
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6, 171–178. doi:[https://doi.org/10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411–443. doi:<https://doi.org/10.1080/13506280500195250>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. doi:[https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Victor, J. D., Conte, M. M., & Chubb, C. F. (2017). Textures as probes of visual processing. *Annual Review of Vision Science*, 3, 275–296.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1436–1451. doi:<https://doi.org/10.1037/0096-1523.32.6.1436>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. doi:<https://doi.org/10.3758/s13423-017-1323-7>
- Westfall, P. H., Johnson, W. O., & Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. *Biometrika*, 84, 419–427.
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, 18, 484–489. doi:<https://doi.org/10.3758/s13423-011-0071-3>
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11:1120–1135. doi:<https://doi.org/10.1167/4.12.11>
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440, 91–95. doi:<https://doi.org/10.1038/nature04262>
- Yildirim, I., Ögreden, O., & Boduroglu, A. (2018). Impact of spatial grouping on mean size estimation. *Attention, Perception, & Psychophysics*, 80, 1847–1862. doi:<https://doi.org/10.3758/s13414-018-1560-5>
- Yörük, H., & Boduroglu, A. (2019). *Feature-specificity in visual statistical summary processing*. Manuscript submitted for publication.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.