American Psychological Association

Journal of Experimental Psychology:
Human Perception and Performance

ISSN: 0096-1523

2020, Vol. 46, No. 9, 1013–1028
http://dx.doi.org/10.1037/xhp0000804

The Visual System Does Not Compute a Single Mean but Summarizes a Distribution

MyoungAh Kim and Sang Chul Chong
Yonsei University

Ongoing discussions on perceptual averaging have the implicit assumption that individual representations are reduced into a single prototypical representation. However, some evidence suggests that the mean representation may be more complex. For example, studies that use a single item probe to estimate mean size often show biased estimations. To this end, we investigate whether the mean representation of size is reduced to a single mean or includes other properties of the set. Participants estimate the mean size of multiple circles in the display set by adjusting the mean size of the circles in the probe set that followed. Across 3 experiments, we vary the similarity of set-size, variance, and skewness between the display and probe sets and examine how property congruence affects mean estimation. Altogether, we find that keeping properties consistent between the 2 compared sets improves mean estimation accuracy. These results suggest that mean representation is not simply encoded as a single mean but includes properties such as numerosity, variance, and the shape of a distribution. Such multiplex nature of summary representation could be accounted for by a population summary that captures the distributional properties of a set rather than a single summary statistic.

Public Significance Statement

The key challenge of the visual system is to represent complex visual inputs in a way that is compact enough to not overflow limited processing capacity, yet descriptive enough to inform their important characteristics. Previous studies have shown that this task may be achieved via perceptually averaging multiple items into an ensemble mean representation and that various other statistical regularities are also represented, such as range, variance, and numerosity. However, exactly how these various statistics relate to each other has been unclear. We have found that keeping various statistical properties consistent between the 2 compared sets of items improves mean estimation. This result suggests that mean representation does not separately encode a single mean but includes various other statistical properties.

Keywords: ensemble representation, summary statistics, mean size, perceptual averaging, population coding

When given a set of similar items, research suggests that the visual system can form an overall statistical representation of the set (Ariely, 2001; Chong & Treisman, 2003). This naturally prompted questions about how we extract statistical properties from multiple items (Alvarez, 2011; Whitney & Yamanashi Leib, 2018). Some of the key issues were how individual item processing and ensemble processing are related (Banno & Saiki, 2012; Corbett & Oriet, 2011; Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008; Fischer & Whitney, 2011; Haberman, Brady, & Alvarez, 2015; Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2011; Hochstein, Pavlovskaya, Bonneh, & Soroker, 2015; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Sweeny, Wurnitsch, Gopnik, & Whitney, 2015; Ward, Bear, & Scholl, 2016; Leib, Landau, Baek, Chong, & Robertson, 2012; Leib, Puri, et al., 2012), how many items are used in the mean computation (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Alvarez & Oliva, 2008; Chong, Joo, Emmanouil, & Treisman, 2008; Chong & Treisman, 2005; Dakin, 2001; Florey, Clifford, Dakin, & Mareschal, 2016; Im & Halberda, 2013; Maule & Franklin, 2016; Myczek & Simons, 2008; Solomon, Morgan, & Chubb, 2011), and how items are weighted in the averaging process based on precision, attention, and recency (Albrecht & Scholl, 2010; de

This article was published Online First June 4, 2020.

MyoungAh Kim, Center for Cognitive Science, Yonsei University; Sang Chul Chong, Graduate Program in Cognitive Science and Department of Psychology, Yonsei University.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; NRF-2019R1A2B5B01070038). Data in this research were previously presented at the 19th annual meeting of the Vision Sciences Society (May 2019). The raw data for all experiments is available on the Open Science Framework (https://osf.io/h7afr).

We thank Chris Oriet and Chad Dubé for the helpful comments and discussion about this article.

Correspondence concerning this article should be addressed to Sang Chul Chong, Department of Psychology, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Korea. E-mail: scchong@yonsei.ac.kr

Fockert & Marchant, 2008; Hubert-Wallander & Boynton, 2015; Juni, Gureckis, & Maloney, 2012; Kanaya, Hayashi, & Whitney, 2018; Mareschal, Morgan, & Solomon, 2010; Tong, Dubé, & Sekuler, 2019).

The majority of these ongoing discussions on perceptual averaging deal with mean representation in terms of a single percept, somewhat like a "prototype" (e.g., Tong et al., 2019). For example, a set of different-sized circles is summarized as a single average size. This follows from the understanding of the basic mechanism of averaging, which is summarizing of the set through pooling. The idea is that noisy individual measurements are pooled together and averaged to a higher level of abstraction, during which the noise present in individual representations is canceled out. The resulting ensemble representation is an average signal that is more precisely represented than the individual representations in a set (Alvarez & Oliva, 2009; Galton, 1907; Sun & Chong, 2020). With this reduction process as a key subject of interest, a number of computational models have been proposed for mean orientation (e.g., Parkes et al., 2001) and mean size (e.g., Allik et al., 2013; Baek & Chong, 2020a; Solomon et al., 2011). Although these models have different assumptions in the incorporation of components, such as early noise, late noise, and attention, their basic computational scheme concerns reducing multiple inputs into one summary variable. This is intuitively understandable as well because if the visual system should combine multiple items into a summary, the most obvious strategy is to converge them into a representative value that contains essential information. In particular, when faced with the fundamental issue of how the brain deals with the overwhelming amount of sensory inputs, such reduction suggests a potential mechanism through which the brain overcomes its limited capacity (Im & Chong, 2014) because pruning the redundancies into a single average can be a quite efficient way to compress information.

The prevalence of this "one average" assumption can also be found in one of the most commonly used methods to assess mean perception: reporting the mean of multiple items (e.g., mean size) using a single item probe. For example, studies that investigate mean size perception show multiple items in the display set, followed by a single item as a probe. Observers are asked to identify whether the mean of the display set is larger or smaller than the probe, or they are asked to reproduce the mean by adjusting the probe's size to approximate the mean of the display set (Allik et al., 2013; Ariely, 2001; Bauer, 2009, 2017; Chong & Treisman, 2003; Lee, Baek, & Chong, 2016; Li & Yeh, 2017; Marchant, Simons, & de Fockert, 2013; Oriet & Hozempa, 2016).

However, contrary to this intuitive understanding, reporting the mean size in terms of a single mean often yields an inaccurate estimation, as observers tend to overestimate the mean size. For instance, numerous studies show a positive bias in error when the mean is reported using a single probe (e.g., Bauer, 2009, 2017; Dodgson & Raymond, 2020; Li & Yeh, 2017). This overestimation tends to increase when the stimulus has larger set-size (Kanaya et al., 2018; Lee et al., 2016; Marchant et al., 2013) and is more prominent when sizes are more irregular (Kanaya et al., 2018; Marchant et al., 2013; Oriet & Hozempa, 2016). Interestingly, such estimation bias disappears when the mean is reported using a probe set that consists of the same number of items as the display set (Baek & Chong, 2020a; Oriet & Brand, 2013). For instance, when observers were asked to adjust the length of multiple lines on the right half of the display to match the average length of the lines on the left half of the display (which consisted of an identical number of items), the net bias was close to zero (Oriet & Brand, 2013). Overall, the overarching trend of bias across studies shows that bias is related to the number of items in the probe, as well as to the set-size and variance of the display set being tested. Alternatively, it could be related to the similarity between the display and probe stimuli. The fact that bias diminishes when the mean is compared to a set with the same set-size, as opposed to a single item, suggests that mean representation includes more information about a set than a single average.

In fact, there is ample evidence that the ensemble representation embraces more information about a set than just mean information. For one, observers can compute other statistical properties from multiple items, such as range, variance, numerosity, and area (range: Hochstein, Pavlovskaya, Bonneh, & Soroker, 2018; Khayat & Hochstein, 2018; variance: Dakin & Watt, 1997; Haberman, Lee, & Whitney, 2015; Michael, de Gardelle, & Summerfield, 2014; Morgan, Chubb, & Solomon, 2008; Norman, Heywood, & Kentridge, 2015; Solomon et al., 2011; Tong, Ji, Chen, & Fu, 2015; numerosity and area: Halberda, Sires, & Feigenson, 2006; Hurewitz, Gelman, & Schnitzer, 2006). They can also encode the shape of the distribution of distractors in visual search implicitly (Chetverikov, Campana, & Kristjánsson, 2016, 2017). Second, these properties can be extracted simultaneously in parallel. For example, in Lee et al. (2016), observers were able to extract at least two different ensemble properties (numerosity and mean size) at once from the same set. Similarly, Utochkin and his colleagues found that mean and numerosity (Khvostov & Utochkin, 2019; Utochkin & Vostrikov, 2017), as well as mean and range (Khvostov & Utochkin, 2019), can be processed in parallel, as the judgment accuracy of the two statistics did not differ between the precue and postcue conditions nor between the single (extracting only one statistic) and dual tasks (extracting two summaries simultaneously). The fact that there is no cost in the performance of processing the mean along with other statistics indicates that these summary statistics are processed without having to compete for attentional resources with the mean computation. This suggests that a rich set of various summary statistics are represented at the same time with the mean information.

Despite the abundant evidence indicating that various statistics other than the mean are represented, how these statistics are related to the mean has been unclear, or at times, conflicting. One line of evidence shows that each statistic is independent. For example, Yang, Tokita, and Ishiguchi (2018) found a lack of correlation between individual performances of mean and variance tasks, which was interpreted as an indication that different processes are involved in representing the mean and variance. Khvostov and Utochkin (2019) found the same pattern of results in terms of both individual differences and trial-by-trial correlations, in which the precision of mean size and range estimation showed no significant cross-correlation across observers nor correlation within a trial (also for mean size and numerosity). Contrary to the computational independence shown by correlation studies, another line of evidence shows a strong interrelationship between these statistics and the mean. For example, the mean is closely related to numerosity in that mean discrimination becomes less accurate when the number of items differs between the two displays (Chong & Treisman, 2005). Furthermore, the mean is related to variance and the shape of the distribution, as observers show more difficulty in representing the mean for high variance sets (Corbett, Wurnitsch, Schwartz, & Whitney, 2012; Dakin, 2001; Fouriezos, Rubenfeld, & Capstick, 2008; Haberman et al., 2015; Im & Halberda, 2013; Morgan et al., 2008; Oriet & Hozempa, 2016; Solomon et al., 2011; Utochkin & Tiurina, 2014) or when the shapes of the underlying distributions are different between the two sets being compared (e.g., comparing uniform vs. two-peaks or normal vs. homogenous as in Chong & Treisman, 2003). Conversely, variance perception is influenced by the mean in that providing a stable mean context over time facilitates variability perception (Tong et al., 2015). Such bidirectional influence between mean and variance representations was also found in an adaptation study, where the adaptation aftereffect of orientation variance influenced mean orientation discrimination and the adaptation aftereffect of mean orientation influenced orientation variance discrimination (Jeong & Chong, 2020). Given these results, it is unclear whether each summary statistic is supported by independent processors. However, it appears that mean perception is clearly influenced by other statistical properties of a set, which leads to the speculation that mean information is not simply encoded as a single value but may include a more complex set of statistical properties as part of the mean representation.

To this end, the current study investigated whether the mean representation of size was reduced to a single mean size or whether it included other statistical properties. We speculated that ensembles that have different properties could essentially have different summary representations even with the same mean. We approached this question by recreating the bias found in the previous studies in relation to the property similarity between the display and probe stimuli. We hypothesized that the bias found in single probe studies could be the result of the conversion cost involved in the reporting method, in which the representation was systematically altered in the process of translating multiple items into one item representation. This does not occur when two sets are directly compared because the same biases that influence the display set are present in the probe set as well. If this is the case, we should be able to eliminate estimation bias by reducing the differences between the display and probe set properties. To our knowledge, no study has examined mean estimation bias with respect to the similarity between set properties.

To test this hypothesis, we conducted a mean size estimation task using a modified version of the method-of-adjustment task where observers adjust the mean size of circles in the probe set to match the mean size of circles in the display set. Instead of using a single item as a probe, the probe set consisted of one or multiple items in which participants were asked to adjust the overall mean size to match the display set. We manipulated the property differences between the display and probe sets to see if increasing or decreasing the differences in set property would impact mean estimation bias. If property differences facilitate or hinder mean estimation, it would indicate that the mean representation is not simply encoded as a single value independent from these properties but that it is contingent upon these properties. We began by replicating the bias of single probe studies, testing our hypothesis that bias would increase as the set-size differences between the display and probe sets increase (Experiment 1). Next, we tested if mean estimation bias would vary depending on the difference in the variance of sets, even when the set-size is maintained equal (Experiment 2), as well as the difference in the skewness of distributions, even when the overall variance is maintained equal (Experiment 3). Altogether, the results showed that observers report more accurately when the probe set had a similar variance, set-size, and skewness to the display set.

Experiment 1a

Single probe studies often show an overestimation of mean size, whereas set probe studies do not. We suspected that this bias is a conversion cost caused by translating multiple items into one item representation. Hence, in Experiment 1, we examined whether larger set-size differences led to larger errors in mean size estimation. To investigate this, participants estimated the mean size of a set of circles in the display set by adjusting the size of the circle(s) in the probe set that immediately followed. Instead of always showing a single item as the probe, the probe set consisted of single or multiple items with which participants adjusted the overall mean size to match the display set. Using this paradigm, we manipulated the set-size disparity between the display (always 16 circles) and probe sets (1, 4, 9, and 16 circles) over four conditions: 16:1, 16:4, 16:9, and 16:16. We predicted that observers' estimation bias would become systematically larger as the set-size disparity increased between the display and probe sets.

Method

Participants. Eight participants (4 male, 4 female, M_{age} = 24.125 years, age range: 20–28) were recruited from the Yonsei University community board in exchange for monetary compensation. They proceeded with the experiment after they signed an informed consent form approved by the institutional review board of Yonsei University. All participants reported normal or corrected-to-normal vision. The sample size was adaptively determined according to an optional stopping rule based on the Bayes factor. We stopped collecting data when compelling evidence was found for, or against, our term of interest. Bayesian inference is not susceptible to the dangers of optional stopping that applies to frequentist inference (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). We used the Bayesian repeated-measures analyses of variance in the JASP statistical package Version 0.9.2 (JASP Team, 2019) to assess evidence for or against our variable of interest, which was the effect of set-size disparity on mean size estimation error. The repeated-measures ANOVA included one variable, set-size disparity (16:1, 16:4, 16:9, 16:16), and we derived the Bayes Factor (BF_k) for this term, which described how likely the correct model including set-size disparity was over the null model. We chose to stop data collection once compelling evidence was found either for ($BF > 3$) or against ($BF < 0.33$) the effect of set-size disparity, which correlates to "substantial evidence against the null hypothesis" according to Jeffrey's grades of evidence (Jeffreys, 1961). The evidence for set-size disparity was found quickly ($BF_{10} > 384.58$) after collecting data from eight participants, and hence data collection was then stopped.

Apparatus. All stimuli were presented in MATLAB using the Psychophysics Toolbox extension Version 3 (Brainard, 1997; Pelli, 1997) on a gamma-corrected 21–inch CRT monitor (HP P1230; resolution 1,600 × 1,200 pixels; refresh rate: 85 Hz). The experiment was conducted in a dark room where the monitor screen was the only source of light. Participants fixed their head on a chin-and-forehead rest at a viewing distance of 60 cm; one pixel subtended 0.0239° of visual angle at this distance.

Stimuli. A display set always consisted of 16 outlined circles presented in the center of the display. We varied the number of items in a probe set to determine whether the error of using a single item would gradually decrease as the number of items (1, 4, 9 16) neared the set-size of the display set. For this purpose, the number of items in the display set was always 16 as we did not find it necessary to examine any probe set-size that is larger than the display set, and any set-size smaller than 4 items is too small to be considered as an ensemble. For both displays, the circles were outlined in black (0.04 cd/m^2) and presented on a gray background (39.31 cd/m^2). The sizes of the 16 circles in the display set were randomly selected from a normal distribution in which the mean was between 1.13–2.26° and the standard deviation was one-eighth of its mean. This distribution was truncated so that the minimum and maximum size did not exceed two standard deviations from the mean. Each circle was located in each square of a 4 × 4 invisible grid (each square spanned 4° vertically and horizontally) and was randomly jittered within 0.35° without overlapping with another circle. The grid was located at the center of the screen. The probe set showed a different number of circles depending on the condition, which was presented within an area that occupied the same distance as the display set. For the 16:1 condition, only one circle was shown among nine possible locations. For the 16:4 condition, a 2 × 2 invisible grid with four circles was randomly presented among nine possible locations. For the 16:9 condition, a 3 × 3 invisible grid with nine circles was randomly presented among four possible locations. For the 16:16 condition, 16 circles were located in the 4 × 4 invisible grid which was centered on the screen. The size(s) of the circles initially presented on the probe set were selected in the same way as the display set, except the mean was selected from a slightly wider range than that of the display set (2 standard deviations larger than the maximum or smaller than the minimum of the display set mean range). When participants adjusted the probe set's mean size using the up/down arrows on the keyboard, the mean of the distribution was scaled by one pixel larger/smaller with each keypress, in which the sizes were resampled at random from this adjusted distribution. The location of each circle was also reselected with each keypress to avoid the gradual scaling of a fixed set. All sizes were sampled from the distribution based on the Teghtstoonian scale (Teghtsoonian, 1965).

Design and procedure. The trial procedure is illustrated in Figure 1A. A one-way (set-size disparity: 16:1 vs. 16:4 vs. 16:9 vs. 16:16) within-subjects design was used (Figure 2A). Each trial started with a black fixation cross presented for 500–1,000 ms. The display set consisting of 16 circles was then presented for 100 ms, followed by a 500 ms blank screen. Then the probe set was shown which consisted of one, four, nine, or 16 circles depending on the set-size disparity condition. The participants pressed up/down arrow keys to adjust the mean size of circles in the probe set (or the size of a single circle in the case of the 16:1 condition) so that it matched the mean size of the previously presented display set. They pressed the space key to confirm their response, after which the next trial began immediately. The four set-size disparity conditions were conducted with 50 trials each, blocked and ordered according to a Latin square, which was assigned to each participant according to the order of arrival. Participants conducted 12 practice trials (3 trials per set-size disparity condition) before proceeding to the main session. The main session consisted of 4 blocks, 50 trials each, and so 200 trials total. Breaks were inserted between each block and participants pressed the spacebar key to continue at their discretion. The entire session including practice trials took 20–25 min to complete.

Results and Discussion

The error was calculated as the difference between the reported and actual means, divided by the actual mean. The error distribution was derived for each set-size disparity condition per participant, from which we examined two measures: bias and response variance (Figure 1B). Bias was how much the mean of the error

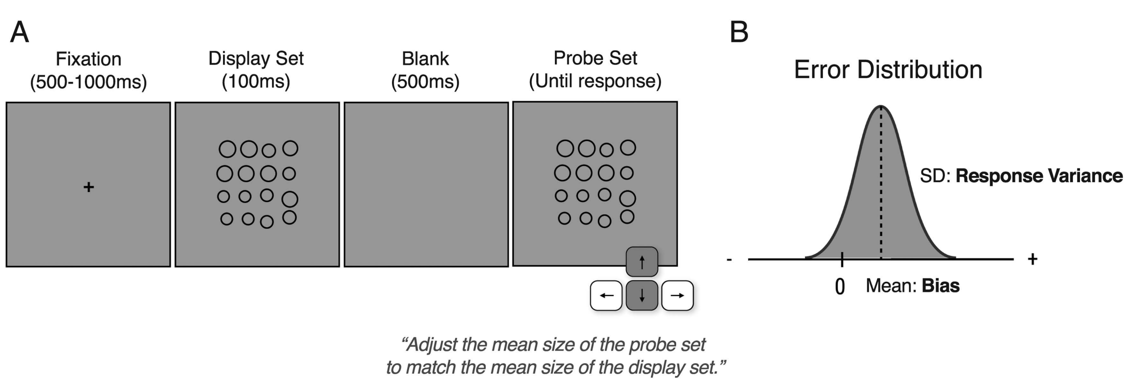

Figure 1. (A) Display sequence of a trial in all Experiments (the circle stimuli presented as display and probe sets are different for each experiment and its respective conditions). Participants were asked to report the perceived mean size of the display set by adjusting the mean size of the probe set. (B) Example of a participant's error distribution. Mean size estimation error was calculated as the difference between the reported and actual means, divided by the actual mean. An error distribution was derived for each condition per participant from which we looked at two measures: bias and response variance. Note that this is a conceptual illustration and not drawn to scale.

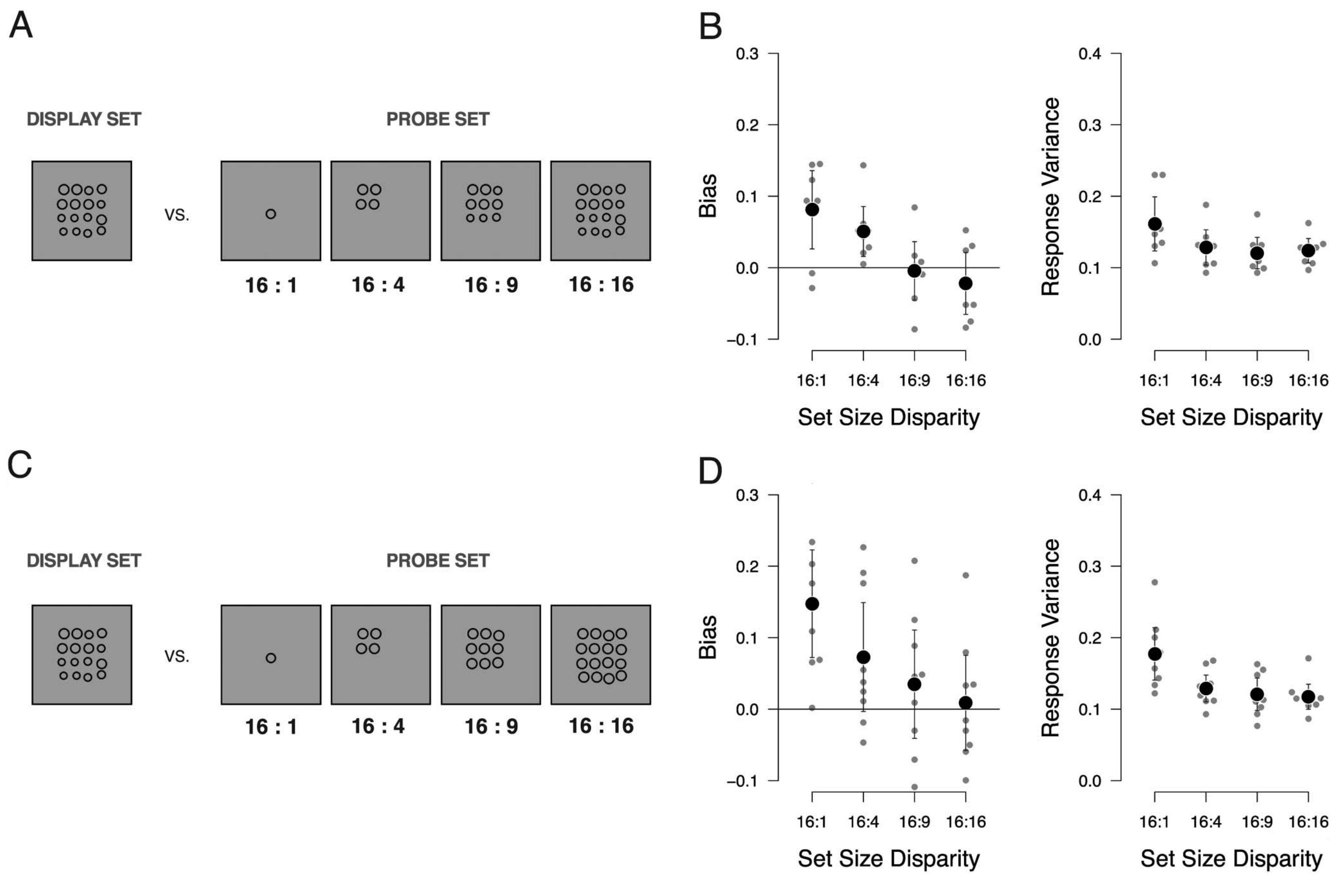

Figure 2. (A) Example of display and probe sets for each set-size disparity condition (display set:probe set) in Experiment 1a. The sizes of the circles in the probe set were randomly selected from a normal distribution as in the display set. (B) Bias and response variance of each set-size disparity condition in Experiment 1a. (C) Example of display and probe sets for each set-size disparity condition (display set:probe set) in Experiment 1b. The sizes of the circles in the probe set were homogenous. (D) Bias and response variance of each set-size disparity condition in Experiment 1b. All error bars indicate the 95% confidence interval.

distribution was shifted relative to 0, indicating how much the observer's estimation was over- or underestimated (positive bias indicates an overestimation of the mean size whereas negative bias indicates an underestimation). Response variance was defined by the standard deviation of the error distribution. Larger variances in responses indicated poor precision. A one-way (set-size disparity: 16:1, 16:4, 16:9, and 16:16) repeated-measures ANOVA was conducted for both bias and response variance.

Bias and response variance for all participants is shown in Figure 2B. The bias results showed a significant main effect of set-size disparity, $F(3, 21) = 11.53$, $p < .001$, $\eta_p^2 = .62$, $BF_{10} = 384.58$. A trend analysis showed that bias decreased linearly as set-size disparity decreased, indicating that there was larger overestimation of the mean with larger set-size differences between the display and the probe sets, $F(1, 7) = 34.81$, $MSE = .05$, $p = .001$, $\eta_p^2 = .83$. A one sample *t* test indicated that both the 16:1, $t(7) = 3.50$, $p = .010$, $d = 1.24$, $BF_{10} = 6.55$ and 16:4, $t(7) = 3.43$, $p = .011$, $d = 1.21$, $BF_{10} = 6.08$ conditions had bias significantly different from zero, while the 16:9 and 16:16 conditions did not (both $ps > .27$). This suggests that overestimation bias occurred when the set-size disparity ratio was larger than 16:9.

The response variance also showed a significant main effect of set-size disparity, $F(3, 21) = 4.20$, $p = .018$, $\eta_p^2 = .38$, $BF_{10} = 4.67$, with it being largest in the 16:1 condition and decreasing as set-size disparity decreased. Trend analysis confirmed that this decreasing trend was also linear, $F(1, 7) = 5.89$, $MSE = .01$, $p = .046$, $\eta_p^2 = .46$, indicating that precision became worse with larger set-size differences between the display and probe sets.

Thus, both bias and response variance results showed that observers reported more accurately when the number of items was similar between the two sets. This replicates the macro trend found across previous studies, where mean size estimation error tends to be greater in studies that use a single item probe (e.g., Bauer, 2009, 2017; Dodgson & Raymond, 2020; Kanaya et al., 2018; Lee et al., 2016; Li & Yeh, 2017; Marchant et al., 2013; Oriet & Hozempa, 2016), compared to those that use a set probe (Baek & Chong, 2020a; Oriet & Brand, 2013). Here we confirm this trend in one study by showing that bias decreases systematically with decreasing set-size difference, which suggests that the positive bias found in previous studies was due to the set-size difference between the display and probe sets. The fact that bias and response variance were largest in the single probe condition (16:1 condition) suggests that representing the mean size of multiple items in terms of a single average size is more difficult and may not be the representational form that size information is summarized in the visual system.

Experiment 1b

Experiment 1b was conducted to eliminate the variance difference that could have influenced performance in Experiment 1a. Since a different number of items were sampled from a fixed distribution, there could have been a possible difference in the variability of sizes between the four set-size disparity conditions. For example, sampling four items could have a smaller variance than sampling 16 items. As the variance of the sizes used in Experiment 1a was relatively low (standard deviation set to one-eighth of the mean), we did not expect this to be critical in producing the difference observed in Experiment 1a. The post analysis, however, showed that the sampled sizes in the probe set of four items did have a slightly smaller variance on average than nine or 16 items (although the difference was not statistically significant). Therefore, to control for this variance difference and see the sole effect of set-size disparity, we reconducted Experiment 1a using only homogenous items in the probe set. By keeping variance constant between the probe sets (zero variability) and only manipulating the set-size difference, the purpose was to confirm that any observed difference in performance was solely due to the set-size difference between the display and probe sets.

Method

Participants. Nine participants (4 male, 5 female, M_{age} = 29 years, age range: 27–34) were recruited and compensated based on the same criterion as in Experiment 1a. As in Experiment 1a, we chose to collect data until evidence was found for or against the effect of set-size disparity (16:1, 16:4, 16:9, 16:16) based on its Bayesian factor. The evidence for set-size disparity was found quickly ($BF_{10} > 957.57$) after collecting data from nine participants, and hence data collection was then stopped.

Apparatus, stimuli, design, and procedure. All experimental apparatus, design, and procedure were identical to Experiment 1a with only one difference: All circles in the probe set had an identical size (Figure 2C). When participants pressed the up/down arrow keys for size adjustment, the size of all circles was scaled by one pixel with each keypress. The location of each circle was reselected with each keypress to avoid the gradual scaling of a fixed set.

Results and Discussion

Even when homogenous circles were used as the probe set, the trend of bias and response variance showed the same set-size disparity effect observed in Experiment 1a (Figure 2D). A two-way mixed ANOVA (set-size disparity: 16:1, 16:4, 16:9, 16:16 × Experiment type: Exp 1a, Exp 1b) showed that the effect of set-size disparity had no significant difference between Experiments 1a and 1b for both bias and response variance (both $ps > .479$, Greenhouse–Geisser corrected for both due to violation of sphericity in Mauchly's test). This suggests that variance had a minimal influence in yielding the pattern of results observed in Experiment 1a. A one-way (set-size disparity: 16:1, 16:4, 16:9, and 16:16) repeated-measures ANOVA showed a significant main effect of set-size disparity for both bias, $F(3, 24) = 14.06$, $p < .001$, $\eta_p^2 = .64$, $BF_{10} = 957.57$, and response variance, $F(1.60, 12.82) = 10.98$, $p = .003$, $\eta_p^2 = .58$, $BF_{10} = 432.82$ (Greenhouse–Geisser corrected). The trend analysis showed that both the bias and response variance decreased linearly as set-size disparity decreased (bias: $F(1, 8) = 25.45$, $MSE = .09$, $p = .001$, $\eta_p^2 = .76$; response variance: $F(1, 8) = 13.19$, $MSE = .02$, $p = .007$, $\eta_p^2 = .62$). A one sample t test indicated that the bias of the 16:1 condition, $t(8) = 4.52$, $p = .002$, $d = 1.51$, $BF_{10} = 23.73$, was significantly different from zero while the 16:4 condition, $t(8) = 2.21$, $p = .058$, $d = 0.74$, $BF_{10} = 1.63$, was marginally significant. Both the 16:9 and 16:16 conditions did not significantly differ from zero (both $ps > .31$). The fact that the same pattern of results was shown even when there was zero variability in probe sets suggests that observer's mean size accuracy was influenced by the difference in set-size.

Experiment 1c

Experiments 1a and 1b showed that observers' estimation becomes more accurate with smaller set-size discrepancy between the display and probe sets. However, due to the way probe set was presented in Experiments 1a and 1b, larger set-size differences also had a smaller spatial overlap between the display and probe sets (and more variable locations as well). It is possible that the higher accuracy found in smaller set-size disparity conditions is not a reflection of similarity in representation per se but a result of larger spatial overlap. For instance, in the 16:16 condition, the 16 circles were always presented in the same location in the invisible grid, so the stimuli in the probe set had an almost exact spatial overlap with those of the display set. Therefore, the two sets could have been matched up based on item-by-item sensory memory, instead of their mean representations. To rule out this possibility, we reran the experiment for the 16:16 condition only, but presenting the stimuli on two different sides of the screen (left/right). Then we compared whether presenting the probe set on the same versus the different side as the display set produced any difference in mean estimation. If performance did not change when the probe was presented in a different location, we could safely exclude the possibility of spatial overlap producing the observed differences in Experiments 1a and 1b.

Method

Participants. Twenty-one participants (9 male, 12 female, M_{age} = 25.67 years, age range: 23–31) were recruited and compensated based on the same criterion as in Experiments 1a and 1b. As in Experiments 1a and 1b, we chose to collect data until evidence was found for or against the variable of interest, which was the location effect (same location, different location) based on its Bayesian factor. We ran Bayesian paired samples t tests between the same location condition versus the different location condition. Evidence against the location effect was found (BF_{10} = 0.23) after collecting data from 21 participants, hence data collection was then stopped.

Apparatus and stimuli. All experimental apparatus and stimuli presentation settings were identical to Experiment 1a but with the following difference: Both the display and probe sets consisted of 16 outlined circles presented in a 4 × 4 invisible grid and were presented on either the left or the right side of the display. The grid containing the circles was shifted 9.5° to the left or the right of the center of the display, depending on the side on which the set was presented. The sampling of circle sizes for both display and probe sets as well as the method of adjusting the probe set's mean size was identical to the 16:16 condition in Experiment 1a.

Design and procedure. Experiment 1b was a one-factor (location congruency: same location vs. different location) within-subjects design (Figure 3A). The procedure was identical to Experiment 1a with only two differences: First, both the display and probe sets had 16 circles. Second, for both sets, the stimuli were presented on either the left or the right side of the screen. In the same location condition, the probe set was presented on the same side as where the display set was presented (left & left, right & right). In the different location condition, the probe set was presented on the different side to where the display set was presented (left & right, right & left). All participants conducted 16 practice trials (8 trials of each condition) before proceeding to the main session. The main session consisted of 50 trials of each location condition (same location, different location) intermixed across four blocks, so 100 trials in total. Breaks were inserted between each block, and participants pressed the spacebar key to continue at their discretion. The entire session including the practice trials took 15–20 min to complete.

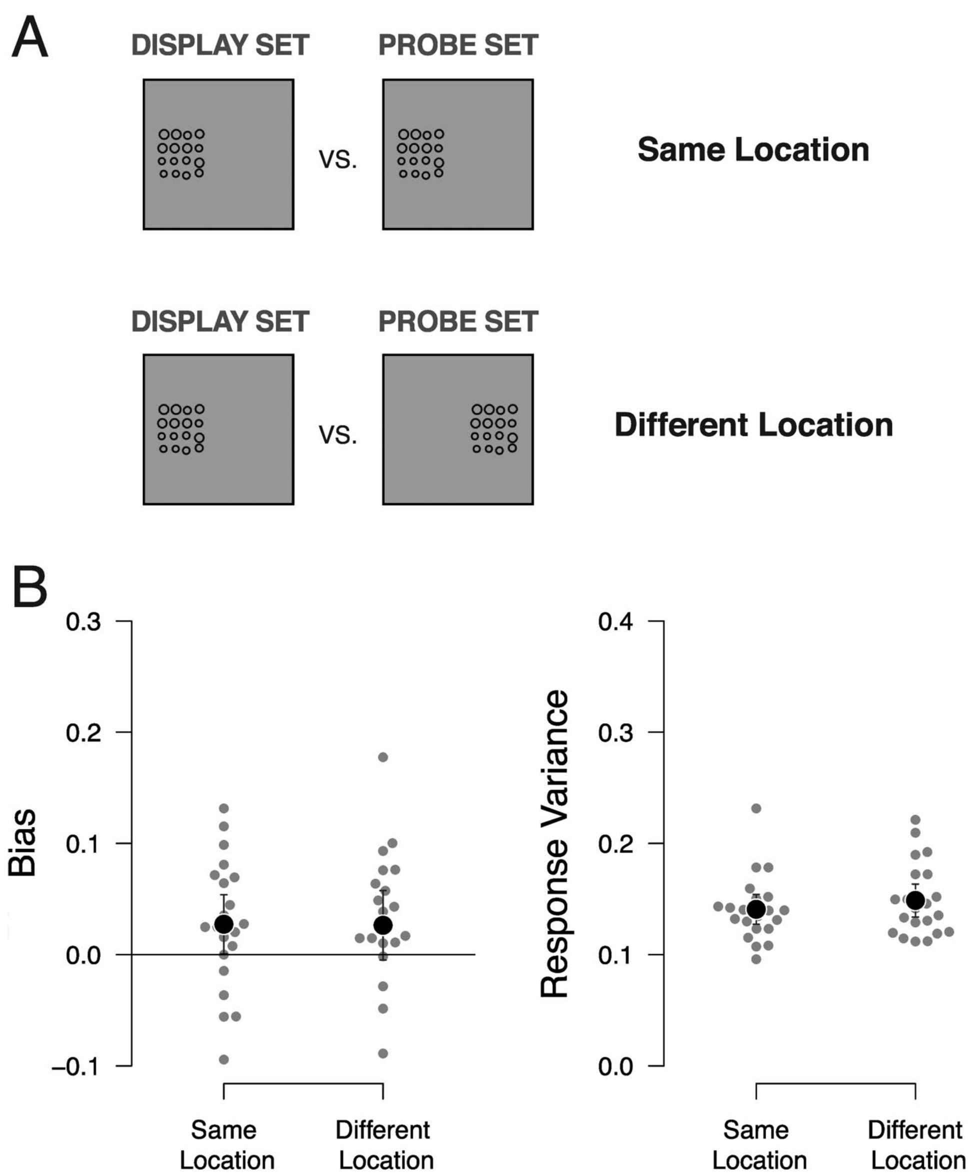

Figure 3. (A) Example of the display and probe sets for the same and different location conditions in Experiment 1c. (B) Bias and response variance of each location condition. All error bars indicate the 95% confidence interval.

Results and Discussion

Contrary to the concern, paired-samples *t* tests showed no significant difference between the same location and different location conditions for both bias and response variance (both $ps > .124$; see Figure 3B). Bias was significantly different from zero for the same location condition, $t(20) = 2.16$, $p = .043$, $d = 0.47$, $BF_{10} = 1.53$, but did not significantly differ from zero for the different location condition, $t(20) = 1.76$, $p = .093$, $d = 0.38$, $BF_{10} = 0.85$. However, Bayesian analysis still yielded no substantial evidence in support of either hypothesis. The result ensures that differences in spatial overlap were not the reason behind the bias difference found in Experiments 1a and 1b. Therefore, we can safely assume that observers did perform Experiments 1a and 1b based on ensemble representation.

Experiment 2

Experiment 1 showed that size representation of a set was not simply averaged into a single mean size and that accurate assessment is bound to the similarity in the number of items in the probe set used to report the mean. Another set property that could influence the mean assessment between two sets could be variance. Studies show that observers can represent information about the variability of features in a set (e.g., Solomon et al., 2011), in which case the difference in size variance between the two compared sets could influence mean estimation even with the same set-size. In Experiment 2, we tested this hypothesis by changing the size variance between the display and probe sets and examined whether the difference in size variance could influence mean size estimation accuracy.

Method

Participants. Twenty-seven participants (11 male, 16 female, $M_{age} = 25.33$ years, age range: 21–31) were recruited from the Yonsei University community board in exchange for monetary compensation. They proceeded with the experiment after they signed an informed consent form approved by the institutional review board of Yonsei University. All participants reported normal or corrected-to-normal vision. As in Experiment 1, we chose to collect data according to an optional stopping rule based on the Bayes factor. The repeated measures ANOVA included the factorial terms, variance congruency (congruent, incongruent) and display set variance type (low variance, high variance). Unfortunately, the *BF* of both model terms as well as interaction remained indecisive even after collecting data from 27 participants, which was three times the sample size of Experiment 1a, hence we stopped data collection (Edwards, Lindman, & Savage, 1963).

Apparatus and stimuli. All experimental apparatus and stimuli presentation settings were identical to Experiment 1 except for a few changes to the display and probe sets. Both the display and probe sets consisted of 16 outlined circles, which were presented in the center of the display and could either have low size variance or high size variance. The sizes of the 16 circles were randomly selected from a normal distribution in which the standard deviation of the distribution was set to one-eighth of its mean for low variance and one half of its mean for high variance. The mean of this distribution was between 0.7–1.4° for the display set, while the mean for the probe set was selected from a slightly wider range (2 standard deviations larger than the maximum or smaller than the minimum of the display set mean range). The circles were randomly jittered within 0.4585° without overlapping with each other.

Design and procedure. Experiment 2 was a 2 (variance congruency: congruent variance vs. incongruent variance) × 2 (display set variance type: low variance vs. high variance) within-subjects design. The stimulus for each condition is illustrated in Figure 4A. The procedure of Experiment 2 was identical to Experiment 1 with the only difference being the stimulus used in the display and probe sets. In the display set, 16 circles were presented which had either low size variance or high size variance. The probe set that followed also consisted of 16 circles, which also had either low size variance or high size variance. In the congruent variance condition, the probe set had the same variance as the display set (low & low, high & high), whereas in the incongruent variance condition, the probe set had a different variance from the display set (low & high, high & low). All participants conducted 12 practice trials (3 trials of each condition) before proceeding to the main session. The main session had 50 trials of each display-probe variance pair (low & low, high & high, low & high, high & low) intermixed across four blocks, so 200 trials in total. Breaks were inserted between each block, and participants pressed the spacebar key to continue at their discretion. The entire session including the practice trials took 20–25 min to complete.

Results and Discussion

The bias and response variance from all participants averaged together are shown in Figure 4B. We conducted a 2 (variance congruency: same variance vs. different variance) × 2 (display set variance type: low variance vs. high variance) repeated-measures ANOVA for both bias and response variance.

Our primary interest was the effect of variance congruency. As expected, the bias results showed a main effect of variance congruency, $F(1, 26) = 15.80$, $p < .001$, $\eta_p^2 = .38$, $BF_{10} = 0.64$, with larger bias in incongruent variance conditions than in congruent variance conditions. We also found a significant main effect of display set variance type, $F(1, 26) = 33.15$, $p < .001$, $\eta_p^2 = .56$, $BF_{10} = 3.97 \times 10^6$, with larger bias in high variance display sets than in low variance display sets. The reason why the mean of high variance display sets was overestimated by a greater amount than low variance sets may be because high variance sets are likely to include larger items. Since large items are more salient than small items (e.g., Proulx, 2010; Proulx & Green, 2011), high variance sets are likely to be influenced by saliency-based-weighting of their larger items (Kanaya et al., 2018), resulting in a larger mean estimation bias. Despite no significant interaction effect, $F(1, 26) = 3.39$, $p = .077$, $\eta_p^2 = .12$, $BF_{inclusion} = 1.51$, simple main effects analysis revealed that there was a clear congruency effect for high variance display sets, $F(1, 26) = 12.94$, $p = .001$, $\eta_p^2 = .33$, $BF_{10} = 21.78$, suggesting that when the display set had high variance, using a low variance probe set (incongruent variance) yielded larger bias, whereas using a probe set with the same high variance reduced bias. This is interesting

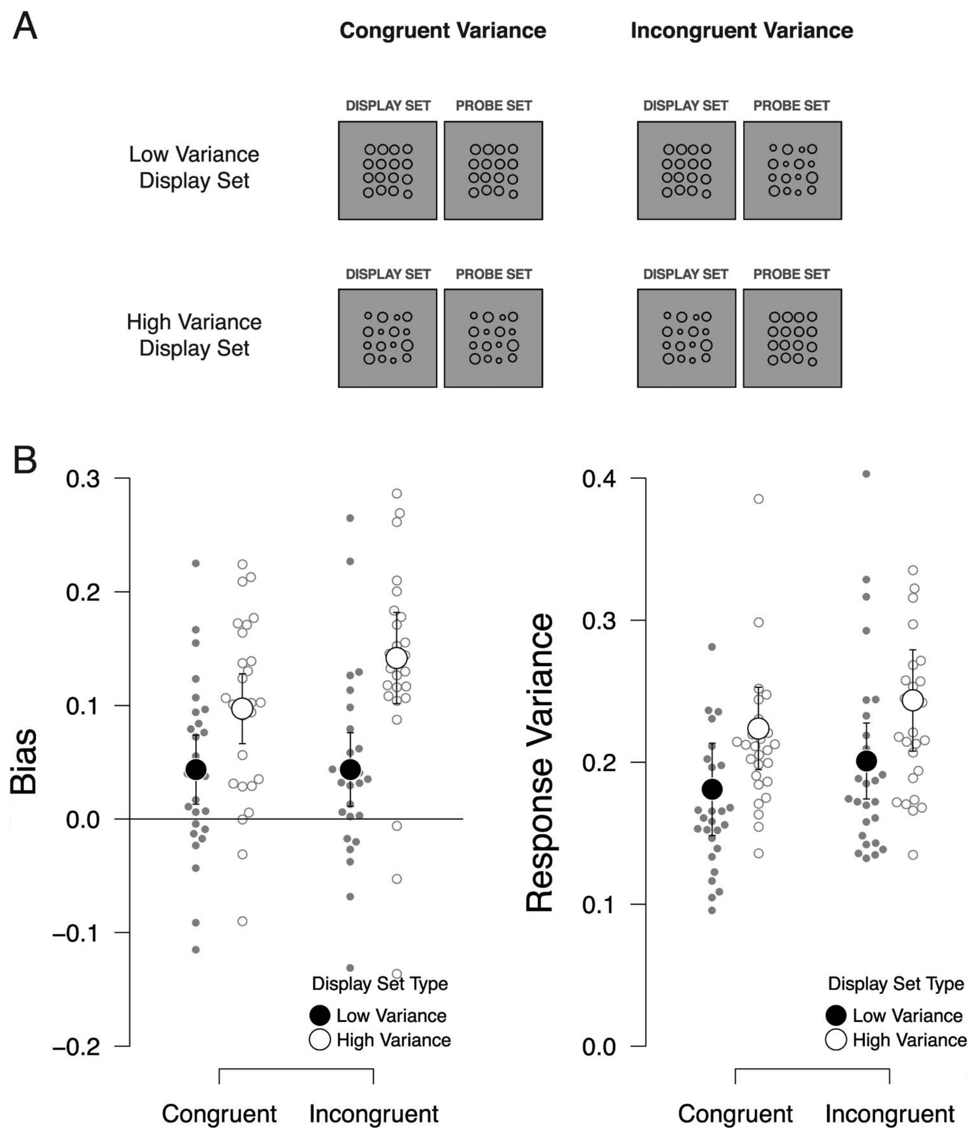

Figure 4. (A) Example of the four display-probe variance pairs in Experiment 2. In the congruent variance condition, both the display and probe sets had the same variance type (low & low, high & high). In the incongruent variance condition, the display and probe sets had different variance types (low & high, high & low). (B) Bias and response variance plotted for variance congruency and display set variance type. All error bars indicate the 95% confidence interval.

considering that previous studies have found that subjects have greater difficulty in reporting the mean as the variability between the items increases (e.g., Dakin, 2001; Solomon et al., 2011). If only a single mean was represented for a set of items, and if this was inherently difficult for high variance sets, then estimating the mean using two sets that are both high variance should have produced worse accuracy. However, our results showed that the mean accuracy even for high variance sets are improved when the probe set used to report the mean also has high variance. This suggests that the mean information is represented with variance information, such that keeping the variance property consistent between two ensembles facilitates accurate comparison of the mean. Unlike high variance, variance congruency did not influence low variance mean estimations, $F(1, 26) = 8.73 \times 10^{-5}$, $p = .993$, $\eta_p^2 = .00$, $BF_{10} = 0.27$, as participants performed well regardless of the probe set's variance. The reason why variance congruency did not show any effect in the low variance condition could be a ceiling effect. Previous studies have shown that mean representations can be formed more easily for low variance stimuli than for high variance stimuli (e.g., Chong & Treisman, 2005; Corbett et al., 2012; Dakin, 2001; Fouriezos et al., 2008; Haberman et al., 2015). Thus, performance for low variance display sets could have been saturated so as to obscure any effect of variance congruency.

Response variance also showed a significant main effect of variance congruency, $F(1, 26) = 26.56$, $p < .001$, $\eta_p^2 = .51$,

$BF_{10} = 3.89$, suggesting that using a probe set with incongruent variance increased response variance. There was also a main effect of display set variance type, $F(1, 26) = 57.37$, $p < .001$, $\eta_p^2 = .69$, $BF_{10} = 3.32 \times 10^6$, with response variance being larger for high variance display sets than for low variance display sets. There was no interaction effect of response variance, $F(1, 26) = 2.65 \times 10^{-4}$, $p = .987$, $\eta_p^2 = .00$, $BF_{\text{inclusion}} = 0.28$.

Altogether, the bias and response variance results showed that mean estimation became worse when the probe set's variance was different from the display set, particularly for high variance mean estimations. This suggests that mean size information of a set is not simply summarized as a single average size but also includes information about the variability of sizes.

Experiment 3

Experiment 2 showed that size information of a set is not simply summarized as a single average size and that accurate assessment is bound to the similarity in the size variance between the display and probe sets. In Experiment 3, we examined whether the mean assessment is influenced not only by a disparity in the general degree of dispersion (variance) but also in the shape of dispersion. Previous works in ensemble statistical learning have shown that observers can implicitly represent the shape of the underlying distribution in search displays (Chetverikov et al., 2016, 2017). In their studies, search times depended on the shape of the preceding distractor distribution, suggesting that observers can implicitly learn properties of distractor distributions. In Experiment 3, we examined whether the difference in the shape of the distribution between the display and probe sets could yield larger mean size estimation error even when the variance is the same. The shape of distribution was manipulated by changing the skewness of the size distribution (positively skewed, negatively skewed).

Method

Participants. Twelve participants (3 male, 9 female, $M_{\text{age}} = 25.33$ years, age range: 19–28) were recruited from the Yonsei University community board in exchange for monetary compensation. They proceeded with the experiment after they signed an informed consent form approved by the institutional review board of Yonsei University. All participants reported normal or corrected-to-normal vision. As in previous experiments, we chose to collect data according to an optional stopping rule based on the Bayes factor. The repeated measures ANOVA included the factorial terms, skewness congruency (congruent skewness, incongruent skewness), and display set skewness type (positively skewed, negatively skewed). Evidence for a skewness congruency × display set skewness type interaction was found ($BF_{\text{inclusion}} = 6.65$) after collecting data from 12 participants, hence data collection was stopped.

Apparatus and stimuli. All experimental apparatus and stimuli presentation settings were identical to Experiment 2 with the following differences. The number of circles for both the display and probe sets was increased to 36 circles so that the sampled sizes would reflect skewness. Each circle was located in a 6 × 6 invisible grid in which each inner square spanned 3° vertically and horizontally. Instead of using a normal distribution, the sizes were randomly sampled from two types of triangular distribution (positively skewed, negatively skewed). A triangular distribution is defined by its range and peak. Because our task required a continuous adjustment of the mean size, we determined the range and peak of the distributions as proportional to the mean (Figure 5A). Specifically, the ratio between the mean and range was 1:1.5, while the distance from either end of the distribution to the peak was 1.35:0.15. The peak of the positively skewed distribution was closer to the smaller end, and the peak of the negatively skewed distribution was closer to the larger end. With such proportions, the variance of the resulting triangular distribution was approximately one third of the mean (which was between the low variance and high variance conditions of Experiment 2). The mean of the triangular distribution was between 0.5324–1.0648° for the display set, while the mean for the probe set was selected from a slightly wider range (2 standard deviations larger than the maximum or smaller than the minimum of the display set mean range). The circles were randomly jittered within 0.5689° without overlapping with each other.

Design and procedure. Experiment 3 had a 2 (skewness congruency: congruent skewness vs. incongruent skewness) × 2 (display set skewness type: positively skewed vs. negatively skewed) within-subjects design. The stimulus for each condition is illustrated in Figure 5B. The procedure was identical to previous experiments with the only difference being the stimulus used in the display and probe sets. In the congruent skewness condition, the display and probe sets had the same skewness (positive & positive, negative & negative). In the incongruent skewness condition, the display and probe sets had a different skewness (positive & negative, negative & positive). All participants conducted 16 practice trials (4 trials of each condition) before proceeding to the main session. The main session had 50 trials of each display-probe skewness pair (positive & positive, negative & negative, positive & negative, negative & positive) intermixed across four blocks, so 200 trials in total. Breaks were inserted between each block, and participants pressed the spacebar key to continue at their discretion. The entire session including the practice trials took 20–25 min to complete.

Results and Discussion

The bias and response variance from all participants are shown in Figure 5C. We conducted a 2 (skew congruency: congruent skewness, incongruent skewness) × 2 (display set type: negatively skewed, positively skewed) repeated-measures ANOVA for both bias and response variance.

In terms of bias, the skewness congruency influenced the two display set skewness types differently (interaction, $F(1, 11) = 5.92$, $p = .033$, $\eta_p^2 = .35$, $BF_{\text{inclusion}} = 6.65$). Simple main effects analysis showed that for positively skewed display sets, using a probe set with different skewness yielded large bias, but using the same positively skewed probe set significantly decreased bias, $F(1, 11) = 10.08$, $p = .009$, $\eta_p^2 = .48$, $BF_{10} = 4.90$. However, this was not the case for negatively skewed display sets; there was no significant difference between using probe sets with the same or different skewness, $F(1, 11) = 1.69$, $p = .220$, $\eta_p^2 = .13$, $BF_{10} = 0.66$. The main effect for skewness congruency was not significant, $F(1, 11) = 1.83$, $p = .204$, $\eta_p^2 = .14$, $BF_{10} = 0.34$, while the main effect of display set type

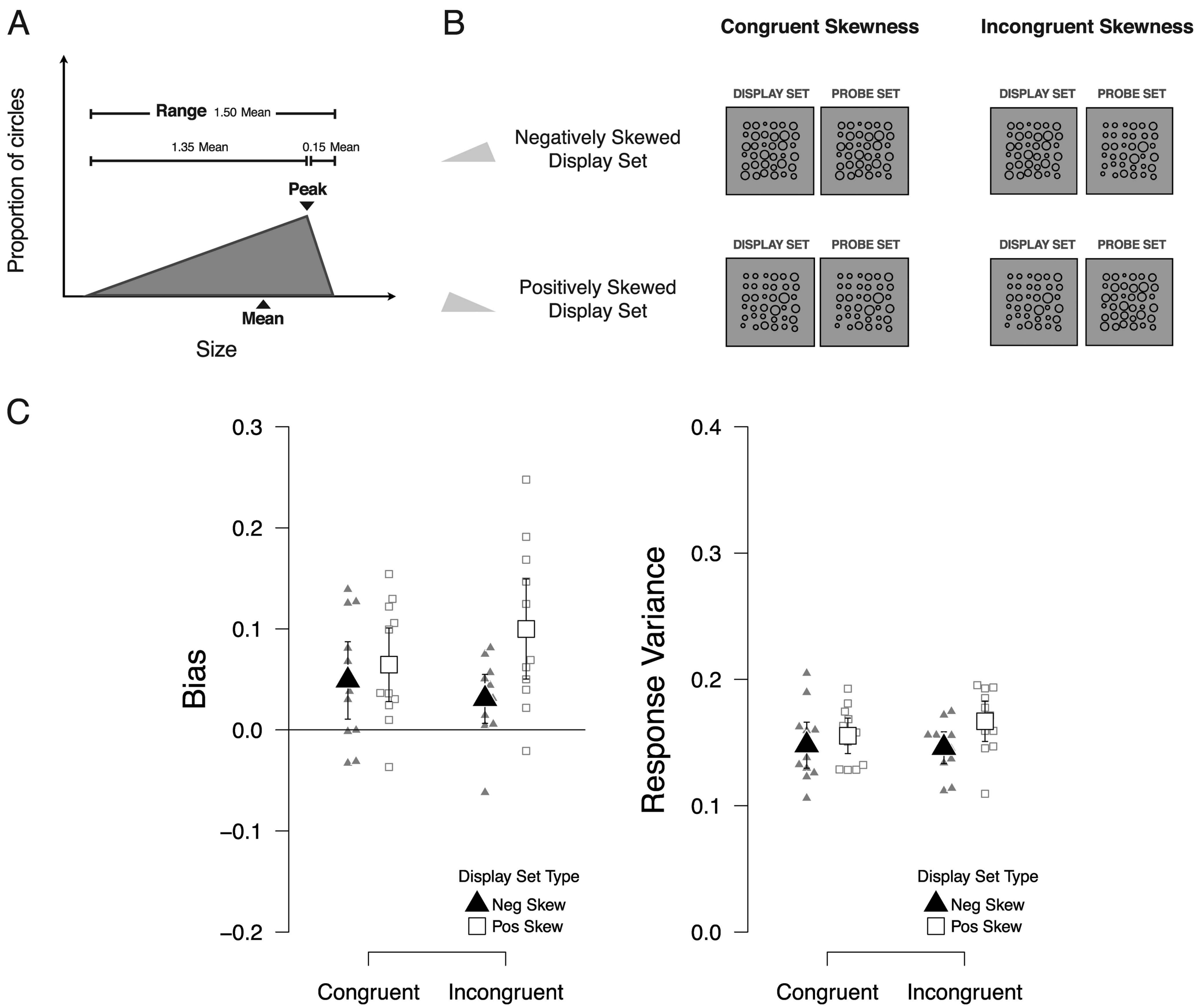

Figure 5. (A) Example of a size distribution for sampling stimuli in Experiment 3. The sizes were sampled from a triangular distribution in which the range and the location of the peak were determined in a fixed ratio to the mean. Two types of triangular distributions were used (positively skewed, negatively skewed). The picture is an example of a negatively skewed distribution. (B) Example of the four display-probe skewness pairs in Experiment 3. In the congruent skewness condition, the display and probe sets both had the same skewness (negative & negative, positive & positive). In the incongruent skewness condition, the display and probe sets had different skewness (negative & positive, positive & negative). (C) Bias and response variance plotted for variance congruency and display set skewness type. All error bars indicate the 95% confidence interval.

was significant, $F(1, 11) = 20.11$, $p < .001$, $\eta_p^2 = .65$, $BF_{10} = 130.22$, with larger bias in positively skewed display sets than in negatively skewed display sets.

Response variance did not show any effect of skewness congruency, $F(1, 11) = 0.35$, $p = .566$, $\eta_p^2 = .03$, $BF_{10} = 0.36$, display set type, $F(1, 11) = 4.129$, $p = .067$, $\eta_p^2 = .27$, $BF_{10} = 2.30$, nor an interaction effect $F(1, 11) = 3.29$, $p = .097$, $\eta_p^2 = .23$, $BF_{\text{inclusion}} = 0.58$.

Similar to Experiment 2, the reason why skewness congruency affected positively skewed display sets but not negatively skewed display sets could be due to differences in task difficulty; mean size computation could have been easier for negatively skewed sets compared to positively skewed sets. Because large items are more salient than small items (Proulx, 2010; Proulx & Green, 2011), a set with few small items mixed among the large items (negatively skewed) makes the set perceptually homogenous and therefore easier to extract the mean.

General Discussion

The current study investigated whether mean size was represented as a single size or whether it included other statistical properties about a set. We conducted a mean size estimation task while varying set-size, variance, and skewness between the display and probe sets and examined whether these property differences yielded estimation bias. In Experiment 1, we found that both bias and response variance increased linearly as the set-size gap increased between the two sets. In Experiment 2, bias and response variance increased when size variance was incongruent between the two sets. Although observers largely overestimated the mean for high variance sets, consistent with their poor accuracy seen in previous studies, their estimations nevertheless improved significantly when it was reported using the same high variance set. Experiment 3 showed that bias increased for positively skewed displays when skewness was incongruent. Altogether, the results show that keeping set-size, variance, and skewness properties consistent between the two compared displays facilitated accurate mean estimation. Differences in these properties biased mean perception and sometimes even impaired precision as well.

The fact that accuracy of mean perception is contingent on the congruency of other statistical properties is strong evidence that the mean is not simply reduced into a single average but is represented with other properties of a set, including numerosity, variance, and skewness. Although encoding of these properties was not explicitly required by the task, differences in them nevertheless yielded estimation bias. Applying a similar logic as Garner (1974)'s interference paradigm[1], this suggests that these properties are processed as an inseparable dimension with the mean, or more fundamentally, that they are part of the same representation. If perceptual averaging simply reduced the set into a single average value without other property information, mean perception should not be biased by a difference in property nor should they be estimated with lower precision. Instead, the results of this study showed that the pooled representation also embraces information about set-size, variance, and the shape of a distribution. As observers consult a representation that is distributed in nature when evaluating the average, their estimates are more precise when the probe used to test for the percept shares a similar distribution. However, when the task requires pitting against different distributions to evaluate the average, the representation should go through a conversion (reduction of irrelevant dimensions), resulting in an estimation bias.[2]

Previous studies have shown that mean representation becomes more difficult when items are heterogeneous (Corbett et al., 2012; Dakin, 2001; Fouriezos et al., 2008; Haberman et al., 2015; Im & Halberda, 2013; Kanaya et al., 2018; Morgan et al., 2008; Oriet & Hozempa, 2016; Solomon et al., 2011; Utochkin & Tiurina, 2014). However, our results showed that the mean accuracy even for high variance sets can be improved when the probe used to report the mean also has high variance. This indicates that the mean is not simply represented in low precision when the variance is high but that the mean is represented with variance information, such that keeping the variance property consistent between two ensembles facilitates accurate comparison of the mean.

This multiplex nature of summary representation is consistent with the finding that the visual system can implicitly represent the range and shape of distributions. Khayat and Hochstein (2018) showed that in deciding set membership, observers preferentially chose items near the mean while easily excluding items outside of the set range, suggesting that participants can automatically and implicitly perceive both mean and range. Utochkin and Yurevich (2016) found that target search duration was affected by the segmentability of the underlying feature distribution, in that a larger number of segmented distractor distributions slowed search speed even when the distance between the target and the nearest distractor stayed the same. Similar results were also shown by Chetverikov et al. (2016, 2017) where search times depended on the shape of the preceding distractor distribution, suggesting that observers implicitly encoded properties of distractor distributions over time. In these studies, the range and distribution were not explicitly required by the tasks but were nevertheless encoded implicitly.

The question that remains is how these various properties are represented as a unit. One possibility is that the visual system represents the summary of the population responses across individual representations instead of explicitly forming the representation of a mean (Baek & Chong, 2020b). Baek and Chong (2020b) proposed that the key with which the visual system can access so many items over its limited capacity could well be explained by the population coding of individual items (Georgopoulos, Schwartz, & Kettner, 1986) across the relevant stages of visual processing. Instead of independently encoding the individual values and averaging them, the incoming visual inputs can be integrated into a population response that reflects their magnitude and quality depending on location. To grasp the statistical summary of multiple items spread over a broad region, the brain can distribute attention over the region and extract the distributional summary from the noisy population activation by normalizing to the closest fitting Gaussian-shaped activity profiles (e.g., Deneve, Latham, & Pouget, 1999; Treue, Hol, & Rauber, 2000). This Gaussian profile summarizes the overall shape of the underlying distribution. When faced with the task of estimating mean, variance, or range, the visual system can read-out the property estimates from this distributional summary. For example, the mean is identified as the peak, while the variance and range are identified by the edges of the distribution.

The idea that the population code is used to represent statistical summaries has been suggested in previous studies. Zohary, Scase,

[1] Garner and colleagues studied the nature of encoding more than one feature and distinguished between integral and separable dimensions. If two features are separable dimensions, they can each be attended/processed without interference from the other irrelevant dimension (e.g., color and shape). In contrast, integral dimensions cannot each be selectively processed without the interference from the second dimension, and it is highly likely that they are processed together (e.g., lightness and saturation, width and height of a rectangle). Applying a similar logic, the fact that processing the mean is influenced by properties that are irrelevant to the task suggests that these properties are not processed as separable dimensions from the mean (variance may not strictly be irrelevant given that it is related to task difficulty; therefore, observers may not try to ignore variance in the same way they ignore the width when judging the height of the rectangle).

[2] The idea that similarity in encoding impacts difficulty in retrieval is also related to encoding specificity principle (Tulving & Thomson, 1973). According to the encoding specificity principle, memory is improved when information available at encoding is also available at retrieval as it supplies effective cues to facilitate retrieval. However, whereas the encoding specificity account would consider the irrelevant properties (numerosity, variance, skewness) as "cues" to retrieve a separately encoded representation (mean), our perspective focuses on the fact that the mean may not even be encoded in the first place as a single value representation to be retrieved but a property that should be "read-out" from a distributional representation.

and Braddick (1996) suggested that perceptual judgments of overall direction of motion is not based on a rigid algorithm generating a single-valued output, but that the visual system has access to the entire distribution of activity where it is able to flexibly access the mean or the modal direction of motion. Relatedly, Hochstein et al. (2018) have suggested that mean discrimination and outlier detection can both be explained by the vector sum and range encoded in the population responses. Similarly, Webb, Ledgeway, and McGraw (2007) showed that population coding algorithms that read-out from directionally tuned activity (e.g., maximum-likelihood) predicted human motion perception better than statistical estimates of central tendency, while Brezis, Bronfman, and Usher (2015) used a population coding-based model to approximate numerical averaging.

Statistical summary as a population response calls into question the assumption of subsampling models where only a few items are used to compute the mean (Allik et al., 2013; Myczek & Simons, 2008; Solomon et al., 2011; Solomon et al., 2011). Although a few item samples will be able to predict the mean of the population to a reliable degree for any type of distribution, it is hard to make sense of how distributional information can still be preserved when only a few items are selected and used in the encoding of ensemble information. Moreover, the estimation error of inferring the mean from a few samples is expected to be the worst when two high variance sets are to be compared/discriminated, which was not the case in our study, as it was better than comparing two unmatched variances even if one had a smaller variance. Thus, it seems that the ensemble information encoded by the visual system does not involve single-valued statistics inferred from a few samples of the population, but the entire population is summarized in a way that preserves its distribution. Relatedly, the fact that the mean percept is influenced by implicitly represented distributional properties indicate that it is important for future models to consider how such distributional properties may influence mean computation. When a set is integrated as a population response, depending on the shape of the underlying feature distribution, it may be more difficult to identify the mean (peak) in some distributions than others (e.g., in extreme cases such as two-peak distributions, it may be better summarized as two means rather than a one global mean that averages the two peaks). Set properties such as numerosity, variance, or the specific shape of the distribution should affect the difficulty of integrating measurements into a distinguishable mean (peak), and such integration difficulty (rather than a sampling error) could be incorporated to better predict mean estimation performance.

The evidence that the mean is contingent on other statistical properties seemingly supports that a single processor may underlie the computation of these different statistics. However, findings in this study cannot suggest whether these summary statistics are computed by a single processing mechanism or whether each depends on an independent processor, as we did not examine the direct relationship between them. What we have found is that the mean judgment is influenced by task-irrelevant, implicitly coded properties. Most importantly, studies have shown that observers cannot discriminate high-order statistics such as skewness (Atchley & Andersen, 1995; Dakin & Watt, 1997), which suggests that the set properties that influence the mean computation may not even have designated processors for its explicit computation. Thus, this study does not suggest anything about whether the numerosity or variance computations are supported by the same processor as the mean or whether they are independent. In fact, it is possible that mean, numerosity, and variance originate from the same early representation (distributional summary) but are extracted by independent mechanisms. This goes in line with why mean and range/variance (e.g., Khvostov & Utochkin, 2019; Khayat & Hochstein, 2018) or mean and numerosity (Khvostov & Utochkin, 2019; Lee et al., 2016; Utochkin & Vostrikov, 2017) are processed in parallel, while their computations do not show correlation (e.g., Khvostov & Utochkin, 2019; Utochkin & Vostrikov, 2017; Yang, Tokita, & Ishiguchi, 2018) and also why mean and variance are influenced by nonoverlapping sources of late noise (Solomon et al., 2011) and yet interact with each other (Jeong & Chong, 2020).

One further question that remains is why the direction of bias is positive? Our results showed that the mean size is overestimated when the probe differed greatly from the display set. If representation conversion should find an equilibrium point between two population responses with different set-size or variance, why does it systematically lead to overestimation in more numerous or high-variance sets as opposed to underestimation? We believe the answer is related to saliency-weighted averaging that assigns greater weight to larger items in mean size estimation (Kanaya et al., 2018). Much evidence shows that larger items are more salient than smaller items. For instance, detecting a large target among small distractors is easier and faster than vice versa, showing a search asymmetry for a larger target (Treisman & Gormican, 1988). Moreover, the set with the largest mean size attracts selective attention in the same way as larger individual objects (Im, Park, & Chong, 2015; Proulx, 2010), and a preference for larger objects is observed from early infancy (Newman, Atkinson, & Braddick, 2001). This relates to why having few large items capturing attention among the more numerous small items, rather than vice versa, renders stimuli perceptually heterogeneous (as seen in Experiment 3). However, we do not think that the mean was inflated because larger items were selected with priority in the mean computation. As a matter of fact, in the process of summarizing the population activation, there is no selecting or discarding of individual items (as proposed by Myczek & Simons, 2008; and as modeled by Allik et al., 2013; Solomon et al., 2011), but rather all items are integrated into the population activation. Instead, the weighting is likely to be applied in the registering process of individual representations, so that the larger items are registered with higher precision, and therefore integrated into the population response with less noise, resulting in a population activation that is skewed toward the larger end. Consequently, the peak of the activation is biased toward bigger items, resulting in biased mean readouts. In the same sense, sets with bigger set-size or high variance, which are more likely to include larger items, result in more biased readouts. However, specifically how the mean is read-out in a biased manner for different property changes is a subject for further investigation. The bottom line is that the bias reflects a representational discrepancy between two sets. Future studies should look into how individual representations are registered into the population activation and how the brain normalizes and summarizes these population activations.

In summary, we investigated whether the mean size of a set is simply represented as a single size or includes more information about other statistical properties of the set. Across three experiments, we varied the similarity of set-size, variance, and skewness between the test and probe displays and examined how the property congruence between the two displays affected mean estimation bias. We found that observers report more accurately when the probe set had a similar set-size, variance, and skewness, as the display set, suggesting that ensemble representation of sizes is not simply reduced to a single average but is bound to various statistical properties of a set. Such multiplex nature of summary representation could be understood by the summary of population response.

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. http://dx.doi.org/10.1177/0956797610363543

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83,* 25–39. http://dx.doi.org/10.1016/j.visres.2013.02.018

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

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19,* 392–398. http://dx.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 of the United States of America, 106,* 7345–7350. http://dx.doi.org/10.1073/pnas.0808981106

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

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

Baek, J., & Chong, S. C. (2020a). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics, 82,* 63–79. http://dx.doi.org/10.3758/s13414-019-01827-z

Baek, J., & Chong, S. C. (2020b). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin & Review*. Advance online publication. http://dx.doi.org/10.3758/s13423-020-01718-7

Banno, H., & Saiki, J. (2012). Calculation of the mean circle size does not circumvent the bottleneck of crowding. *Journal of Vision, 12*(11), 13. http://dx.doi.org/10.1167/12.11.13

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

Bauer, B. (2017). Perceptual averaging of line length: Effects of concurrent digit memory load. *Attention, Perception, & Psychophysics, 79,* 2510–2522. http://dx.doi.org/10.3758/s13414-017-1388-4

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

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports, 5,* 10415. http://dx.doi.org/10.1038/srep10415

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition, 153,* 196–210. http://dx.doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Representing color ensembles. *Psychological Science, 28,* 1510–1517. http://dx.doi.org/10.1177/0956797617713787

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

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404. http://dx.doi.org/10.1016/S0042-6989(02)00596-5

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

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. http://dx.doi.org/10.1016/j.actpsy.2011.08.002

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition, 20,* 211–231. http://dx.doi.org/10.1080/13506285.2012.657261

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 1016–1026. http://dx.doi.org/10.1364/JOSAA.18.001016

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

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

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

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience, 2,* 740–745. http://dx.doi.org/10.1038/11205

Dodgson, D. B., & Raymond, J. E. (2020). Value associations bias ensemble perception. *Attention, Perception, & Psychophysics, 82,* 109–117. http://dx.doi.org/10.3758/s13414-019-01744-1

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. http://dx.doi.org/10.1037/h0044139

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106,* 1389–1398. http://dx.doi.org/10.1152/jn.00904.2010

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports, 6,* 32210. http://dx.doi.org/10.1038/srep32210

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics, 70,* 456–464. http://dx.doi.org/10.3758/PP.70.3.456

Galton, F. (1907). One vote, one value. *Nature, 75,* 414. http://dx.doi.org/10.1038/075414a0

Garner, W. R. (1974). The processing of information and structure. *The Processing of Information and Structure*. Oxford, England: Erlbaum.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233,* 1416–1419. http://dx.doi.org/10.1126/science.3749885

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. http://dx.doi.org/10.1037/xge0000053

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1. http://dx.doi.org/10.1167/9.11.1

Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision, 15*(4), 16. http://dx.doi.org/10.1167/15.4.16

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review, 18,* 855–859. http://dx.doi.org/10.3758/s13423-011-0125-6

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

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2015). Global statistics are not neglected. *Journal of Vision, 15*(4), 7. http://dx.doi.org/10.1167/15.4.7

Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Soroker, N. (2018). Comparing set summary statistics and outlier pop out in vision. *Journal of Vision, 18*(13), 12. http://dx.doi.org/10.1167/18.13.12

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. http://dx.doi.org/10.1167/15.4.5

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 19599–19604. http://dx.doi.org/10.1073/pnas.0609485103

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

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics, 75,* 278–286. http://dx.doi.org/10.3758/s13414-012-0399-4

Im, H. Y., Park, W. J., & Chong, S. C. (2015). Ensemble statistics as units of selection. *Journal of Cognitive Psychology, 27,* 114–127. http://dx.doi.org/10.1080/20445911.2014.985301

JASP Team. (2019). JASP (Version 0.10.2) [Computer software]. Retrieved from https://jasp-stats.org/

Jeffreys, H. (1961). *The Theory of Probability* (3rd ed.). New York, NY: Oxford University Press.

Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research, 167,* 46–53. http://dx.doi.org/10.1016/j.visres.2020.01.002

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2012). Effective integration of serially presented stochastic cues. *Journal of Vision, 12*(8), 12. http://dx.doi.org/10.1167/12.8.12

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences, 285,* 20172770. http://dx.doi.org/10.1098/rspb.2017.2770

Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision, 18*(9), 23. http://dx.doi.org/10.1167/18.9.23

Khvostov, V., & Utochkin, I. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision, 19*(9), 3. http://dx.doi.org/10.1167/19.9.3

Lee, H., Baek, J., & Chong, S. C. (2016). Perceived magnitude of visual displays: Area, numerosity, and mean size. *Journal of Vision, 16*(3), 12. http://dx.doi.org/10.1167/16.3.12

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia, 50,* 1698–1707. http://dx.doi.org/10.1016/j.neuropsychologia.2012.03.026

Leib, A., Landau, A. N., Baek, Y., Chong, S. C., & Robertson, L. (2012). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. *Frontiers in Human Neuroscience, 6,* 267. http://dx.doi.org/10.3389/fnhum.2012.00267

Li, K.-A., & Yeh, S.-L. (2017). Mean size estimation yields left-side bias: Role of attention on perceptual averaging. *Attention, Perception, & Psychophysics, 79,* 2538–2551. http://dx.doi.org/10.3758/s13414-017-1409-3

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142,* 245–250. http://dx.doi.org/10.1016/j.actpsy.2012.11.002

Mareschal, I., Morgan, M. J., & Solomon, J. A. (2010). Attentional modulation of crowding. *Vision Research, 50,* 805–809. http://dx.doi.org/10.1016/j.visres.2010.01.022

Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 33,* A22–A29. http://dx.doi.org/10.1364/JOSAA.33.000A22

Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 7873–7878. http://dx.doi.org/10.1073/pnas.1308674111

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision, 8*(11), 9. http://dx.doi.org/10.1167/8.11.9

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. http://dx.doi.org/10.3758/PP.70.5.772

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

Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision, 15*(4), 3. http://dx.doi.org/10.1167/15.4.3

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research, 79,* 8–16. http://dx.doi.org/10.1016/j.visres.2012.12.004

Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision, 16*(3), 3. http://dx.doi.org/10.1167/16.3.3

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. http://dx.doi.org/10.1038/89532

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

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

Proulx, M. J., & Green, M. (2011). Does apparent size capture attention in visual search? Evidence from the Muller-Lyer illusion. *Journal of Vision, 11*(13), 21. http://dx.doi.org/10.1167/11.13.21

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision, 11*(12), 13. http://dx.doi.org/10.1167/11.12.13

Sun, J., & Chong, S. C. (2020). Power of averaging: Noise reduction by ensemble coding of multiple faces. *Journal of Experimental Psychology: General, 149,* 550–563. http://dx.doi.org/10.1037/xge0000667

Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4–5-year-old children. *Developmental Science, 18,* 556–568. http://dx.doi.org/10.1111/desc.12239

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

Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception & Psychophysics, 81,* 1962–1978. http://dx.doi.org/10.3758/s13414-019-01697-5

Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision, 15*(4), 15. http://dx.doi.org/10.1167/15.4.15

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

Treue, S., Hol, K., & Rauber, H.-J. (2000). Seeing multiple directions of motion-physiology and psychophysics. *Nature Neuroscience, 3,* 270–276. http://dx.doi.org/10.1038/72985

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373. http://dx.doi.org/10.1037/h0020071

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica, 146,* 7–18. http://dx.doi.org/10.1016/j.actpsy.2013.11.012

Utochkin, I. S., & Vostrikov, K. O. (2017). The numerosity and mean size of multiple objects are perceived independently and in parallel. *PLoS ONE, 12,* e0185452. http://dx.doi.org/10.1371/journal.pone.0185452

Utochkin, I. S., & Yurevich, M. A. (2016). Similarity and heterogeneity effects in visual search are mediated by "segmentability". *Journal of Experimental Psychology: Human Perception and Performance, 42,* 995–1007. http://dx.doi.org/10.1037/xhp0000203

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology, 60,* 158–189. http://dx.doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). *An agenda for purely confirmatory*. http://dx.doi.org/10.1177/1745691612463078

Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition, 152,* 78–86. http://dx.doi.org/10.1016/j.cognition.2016.01.010

Webb, B. S., Ledgeway, T., & McGraw, P. V. (2007). Cortical pooling algorithms for judging global motion direction. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 3532–3537. http://dx.doi.org/10.1073/pnas.0611288104

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69,* 105–129. http://dx.doi.org/10.1146/annurev-psych-010416-044232

Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is there a common summary statistical process for representing the mean and variance? A study using illustrations of familiar items. *i-Perception, 9,* 2041669517747297. http://dx.doi.org/10.1177/2041669517747297

Zohary, E., Scase, M. O., & Braddick, O. J. (1996). Integration across directions in dynamic random dot displays: Vector summation or winner take all? *Vision Research, 36,* 2321–2331. http://dx.doi.org/10.1016/0042-6989(95)00287-1

Received October 4, 2019
Revision received May 1, 2020
Accepted May 8, 2020 ■