Attention, Perception, & Psychophysics (2020) 82:63–79
https://doi.org/10.3758/s13414-019-01827-z

40 YEARS OF FEATURE INTEGRATION: SPECIAL ISSUE IN MEMORY OF ANNE TREISMAN

Distributed attention model of perceptual averaging

Jongsoo Baek[1] · Sang Chul Chong[2,3]

Published online: 25 July 2019

Abstract

The visual system efficiently processes complex and redundant information in a scene despite its limited capacity. One strategy for coping with the complexity and redundancy of a scene is to summarize it by using average information. However, despite its importance, the mechanism of averaging is not well understood. Here, a distributed attention model of averaging is proposed. Human percept for an object can be disturbed by various sources of internal noise, which can occur either before (early noise) or after (late noise) forming an ensemble perception. The model assumes these noises and reflects noise cancellation by averaging multiple items. The model predicts increased precision for more items with decelerated increments for large set-sizes resulting from late noise. Importantly, the model incorporates mechanisms of attention, which modulate each item's contribution to the averaging process. The attention in the model also results in saturation of performance increments for small set-sizes because the amount of attention allocated to each item is greater for small set-sizes than for large set-sizes. To evaluate the proposed model, a psychophysical experiment was conducted in which observers' ability to discriminate average sizes of two displays was measured. The observers' averaging performance increased at a decreasing rate with small set-sizes and it approached an asymptote for large set-sizes. The model accurately predicted the observed pattern of data. It provides a theoretical framework for interpreting behavioral data and leads to an understanding of the characteristics of ensemble perception.

Keywords Averaging · Distributed attention · Noisy percept · Observer model

Significance Ensemble perception is an important ability whereby information is compressed to efficiently process a complex scene. A computational model was developed to explain the empirical phenomena of mean size perception, where its mechanism is not well understood. The key idea in the model is that mean size perception is involved in noisy processes of coding sizes, noise cancellation by averaging, and the attentional modulation of coding sizes. The current model explains a wide range of phenomena, including (1) that distributed attention is better suited for computing mean sizes than focused attention, (2) that attention plays an important role in averaging, especially for a small number of items, and (3) that noisy percepts are responsible for a decelerated averaging performance for a large number of items.

✉ Sang Chul Chong
scchong@yonsei.ac.kr

[1] Yonsei Institute of Convergence Technology, Yonsei University, Incheon, Korea

[2] Graduate Program in Cognitive Science, Yonsei University, Seoul, Korea

[3] Department of Psychology, Yonsei University, 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Korea

Introduction

Visual environments are typically complex. Considering the limited capacity of the visual system (Broadbent, 1958; Luck & Vogel, 1997; Palmer, Fencsik, Flusberg, Horowitz, & Wolfe, 2011), it is necessary to efficiently process complex information. One strategy for efficient information processing is to use statistical regularities, such as the central tendency, to condense representations and, thus, promote more efficient processing of complex scenes (Alvarez, 2011; Ariely, 2001; Chong & Treisman, 2003). Another strategy is to filter out irrelevant information (Carrasco, 2011; Chun, Golomb, & Turk-Browne, 2011).

For the former strategy to be effective, the visual system should be able to form an accurate representation of statistical summaries. From research on a range of visual tasks, it is known that human vision can accurately compute averages from simple visual features, such as color (Maule, Witzel, & Franklin, 2014), direction (Williams & Sekuler, 1984), orientation (Dakin & Watt, 1997), size (Chong & Treisman, 2003), and motion speed (Watamaniuk & Duchon, 1992), to complex visual properties, such as animacy (Yamanashi Leib, Kosovicheva, & Whitney, 2016), gender (Haberman & Whitney, 2007), facial identity (de Fockert & Wolfenstein, 2009; Neumann, Schweinberger, & Burton, 2013), gaze direction (Sweeny & Whitney, 2014), and facial expression (Haberman & Whitney, 2007, 2009).

Despite the growing interest in the statistical summarization of visual information, the mechanism of averaging remains unclear (Alvarez, 2011; Whitney & Yamanashi Leib, 2018). To the authors' knowledge, there presently exist two observer models for averaging: the noise-and-selection model for averaging size (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013) and the orientation averaging model (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). These models were developed using several components, including early noise, late noise, averaging, selective attention, and decision structure. The early noise refers to the variability of internal responses for individual representations before the averaging process, whereas the late noise is the variability of responses during or after the averaging. The two models share components, i.e., early noise, averaging, and decision processes, but have different assumptions concerning selective attention and late noise. The noise-and-selection model assumes that coding the individual size of an item is noisy and that the visual system only samples a few items from the display, owing to its limited capacity (Allik et al., 2013). The model thus predicts the observers' performance in computing mean sizes by randomly selecting a few individual sizes with some noise and averaging them. In this model, internal noise arises before, but not after, the averaging process. Therefore, late noise is not taken into account. In addition, the selection process occurs before the averaging process, so that some of the inputs into the averaging process are filtered out. The orientation averaging model (Parkes et al., 2001) includes late noise in addition to the early noise, and it accurately predicts the orientation averaging performance in the periphery. Solomon and his colleagues further developed an observer model to incorporate the selection process and validated the model with a mean orientation discrimination task (Solomon, 2010) and a mean size discrimination task (Solomon, Morgan, & Chubb, 2011).

How does this averaging process interact with attention? Does attention select a few items to average and discard the rest (Myczek & Simons, 2008), as in the noise-and-selection model (Allik et al., 2013), or are the attended items' contributions to averaging greater than those of the unattended items (de Fockert & Marchant, 2008)? Alternatively, attention could enhance attended items' signals more or less equally, which is consistent with the suggestion that averages are computed more precisely under a distributed attention mode than under a focused attention mode (Chong & Evans, 2011; Chong & Treisman, 2005; Robitaille & Harris, 2011; Treisman, 2006). Thus, it is important to investigate how attention is involved in the process of averaging.

In this study, a distributed attention model of averaging is proposed. The model incorporates the components of early noise and late noise to reflect a noisy percept (Allik et al., 2013; Lu & Dosher, 1998, 1999, 2008; Parkes et al., 2001), averaging process to reflect the degree of noise cancellation, and signal detection theory to quantify the observers' discriminability and behavioral performance (Green & Swets, 1966; Macmillan & Creelman, 1991). Finally, the component of attention is added (Dosher & Lu, 2000b, 2000a; Lu & Dosher, 1998) to incorporate attentional effects on averaging. In a psychophysical validation of the model, two displays with circles of various sizes were sequentially presented, and observers were asked to compare the average size of the circles in the two displays. It was crucial that the number of items to be averaged, referred to as set-size, was varied in each display. Then, the predictions of the model were tested to determine whether they matched with the observed data for average size. The model was also tested using existing datasets in the literature (Allik et al., 2013) and comparing the fit of the noise-and-selection model to our dataset.

Observer models[1]

Noisy percept for individual items: Early noise

Human performance in information processing can be limited by various sources of internal noise (e.g., receptor sampling errors, randomness of neural responses, and loss of information during neural transmission). The model presented here assumes that individual sizes are encoded independently, which results in noisy internal representations for individual sizes. Because noises embedded in the system (whether early or late in the processing) are random variables, the magnitude of the internal response varies, even when the same stimulus is given to the system. Each stimulus passes through a feature detector, and the detector translates it into an internal response of a certain magnitude through a non-linear transducer function. The output of the detector is proportional to the exponent of the signal strength (Foley & Legge, 1981; Nachmias, 1981; Nachmias & Sansbury, 1974). Regarding size perception, it is known that the perceived size of a circle is related to its physical size by a power function with an exponent of 0.76 (Teghtsoonian, 1965). By combining the output of the detector with the early noise, a noisy internal response is generated for each stimulus (Fig. 1).

Averaging multiple items

When the formation of a summary representation is required, the visual system attempts to integrate all internal responses

[1] Refer to analytic models in Appendix 1.

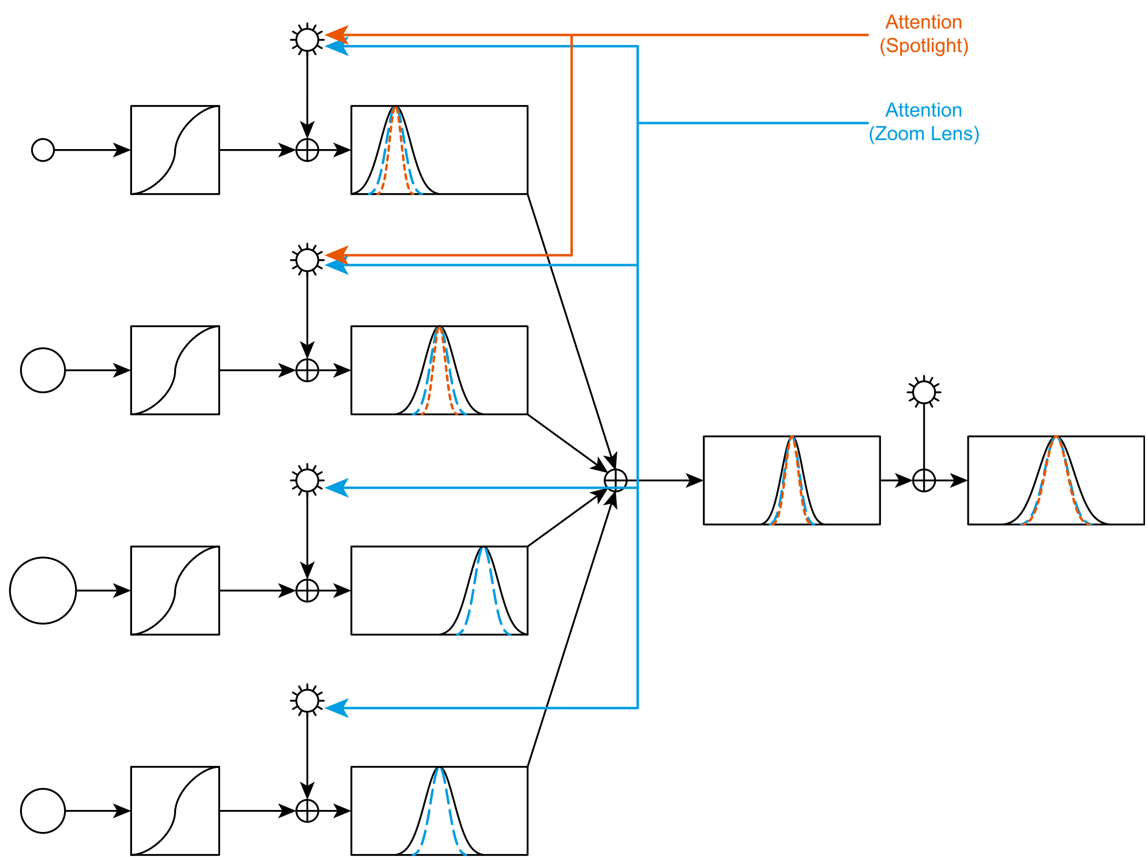

Fig. 1 Internal processes for averaging sizes. The internal representation is disturbed by both early and late noise (expressed by variability of the Gaussian probability distributions). In the zoom lens model (depicted in blue), attention reduces early noise by a large amount for small set-sizes and a small amount for large set-sizes. In the spotlight model (depicted in red), attention reduces the early noise for some items, but not for the others

generated from multiple inputs (Fig. 1). Because random noises for individual inputs can cancel each other out, increasing the number of inputs yields better estimates of the average size.

Noisy percept for the averaged representation: Late noise

Internal noise can occur across all information processing stages, including the stage representing average size after the integration process as well as the early perceptual stage before the integration process. Late internal noise is often thought to be associated with limited precision in internal representations (Heeley & Buchanan-Smith, 1996; Morgan, Ward, & Hole, 1990; Parkes et al., 2001). Because late noise is added to the representation of the averaged size, it does not vary with the number of items. Thus, when late noise is much greater than the averaged early noise (e.g., larger set-sizes, where the averaged early noise is relatively small due to noise cancellation), perceptual sensitivity does not change significantly with set-size.

Comparing average sizes: Signal detection theory

For a two-alternative forced-choice (2AFC) task, in which an observer compares the average sizes of two stimuli sets, i.e., a test and a standard, signal detection theory posits that the two stimulus sets generate two unique internal response distributions. For each trial, the observer compares the magnitudes of the two internal responses and decides which stimulus set generates the greatest internal response. The perceptual sensitivity *d'*, which can be defined as a signal-to-noise ratio, is a function of the difference in average sizes and the variability of the internal response for summary representations. A larger difference between two average sizes leads to a larger difference of internal responses on average, and, consequently, greater discriminability. Further, more precise internal responses (with less variability) for average sizes result in greater discriminability in the 2AFC average size comparison task.

Effect of attention

Attention is thought to enhance the representational precision for attended items (Dosher & Lu, 2000a, 2000b; Lee, Itti, Koch, & Braun, 1999; Lu & Dosher, 1998; but see also Schneider & Komlos, 2008). To model the attentional effects reported in ensemble perception literature (Chong & Treisman, 2005; de Fockert & Marchant, 2008), mechanisms of attention were included in the computational model. Here, the zoom lens model (Eriksen & St. James, 1986) and the spotlight model (Posner, 1980) were used to study attentional deployment over the space.

Zoom lens model: Distributed attention with equal effects on all items

The zoom lens model (Fig. 1 blue lines) assumes that attentional deployment can vary from a sharp focus to a broad window, and its processing capacity is in inverse proportion to the area of focus, similar to the zoom lens of a camera (Eriksen & St. James, 1986). This model assumes that the visual system has a limited processing capacity. Thus, an increase in the size of the attentional window (or set-size) leads to a decreased attentional allocation for each item. When attention is distributed over a broad area, relatively little attention is allocated to an individual item. By contrast, each item receives more attention when attention is focused on a small subset of a display. In a series of studies, Lu and Dosher reported that attention operates mainly by reducing internal noise (Dosher & Lu, 2000a, 2000b; Lu & Dosher, 1998). Attention makes internal responses more precise for attended items, resulting in an improvement of an observer's discriminability.

Spotlight model: Distributed attention with unequal effects on all items

The spotlight model (Fig. 1 red lines) posits that attention selects a fixed number of items, and that the visual system processes the selected items with high precision, with unselected items consigned to low precision representations (Posner, 1980).

Behavioral signatures of proposed models

Suppose an observer compares the average sizes of the circles in two ensemble sets, the *standard* and *test*, that contain equal numbers of circles.[2] Figure 2 shows the threshold of average size difference as a function of set-size that is predicted by four different models: (1) the no-late-noise model, (2) the no-attention model (with late noise), (3) the zoom lens model, and (4) the spotlight model. The threshold is the Weber fraction of mean sizes corresponding to the probability of choice (*pc*; the probability of reporting that the average size of circles was larger in the test display than in the standard display) = 0.76. Each row of Fig. 2 shows the predictions of observer models with different parameter values: The four panels in the first row (a–d) show the predictions of four different models for three different values of early noise (σ_1 = 0.08, 0.16, and 0.32; different colors in each panel), while the other parameters were set as constants. The panels in the second row (e–g) show the predictions of three models that have late noise in their components, for different late noise parameters (σ_2 = 0, 0.04, and 0.08). The panels in the third rows (h–i) show the predictions of two attention models for different attention factors (A_z or A_s = 0.1, 0.4, and 1; A_z for the zoom lens model and A_s for the spotlight model). Here, the $A_z = 1$ or $A_s = 1$ means that attention has no effect on the early noise (or perceptual quality on individual items), but $A_z = 0.1$ or $A_s = 0.1$ means that attention reduces the early noise down to 10% of its original value. The panel (j) in forth row shows the predictions of the spotlight, which have the attentional limit, *k*, in their components, for three *k* parameters (1, 3, and 256). All model predictions were analytically obtained (Eq. 20 with 4–7 in Appendix 1).

In all models, the thresholds decrease as the set-size increases. In the no-late-noise model, the threshold decreases linearly as the number of items increases, with a slope of –0.5 in the log–log axis that reflects a $\sqrt{n}$ relationship between the variability of the internal response and the set-size. In the no-attention model, the slope is smaller than −0.5 on the log-log axis and decelerates as set-size increases. The thresholds for large set-sizes (e.g., 32 or 64) exhibit only a minimal difference, as reported in the literature (Chong & Treisman, 2003; Lee, Baek, & Chong, 2016). Both the zoom lens and spotlight models can be distinguished from the no-attention models by the signatures in the small to middle range of set-sizes. The two attention models predict a deceleration for the slope, but their difference lies in the location of moderate slopes: the zoom lens model exhibits gradual effects of attention for small set-sizes, whereas the spotlight model predicts a sharp decrease in the threshold for small set-sizes ($n < k$) and a refraction around the set-size *k*.

The first row of Fig. 2 shows the expected thresholds with three different early noise levels in the four models. In all models, thresholds increase as early noise increases. This pattern is more clearly observable using small set-sizes in all models, including late noise. The second row shows expected thresholds with three different late noise levels. Threshold decrements tend to saturate at higher threshold levels when late noise increases. Indeed, when $\sigma_2 = 0$, all models predict a linear threshold decrement,

[2] Please note that the average size of the circles was defined using the arithmetic mean of apparent sizes (the diameters with an exponent of 1.52; Chong & Treisman, 2003; Teghtsoonian, 1965).

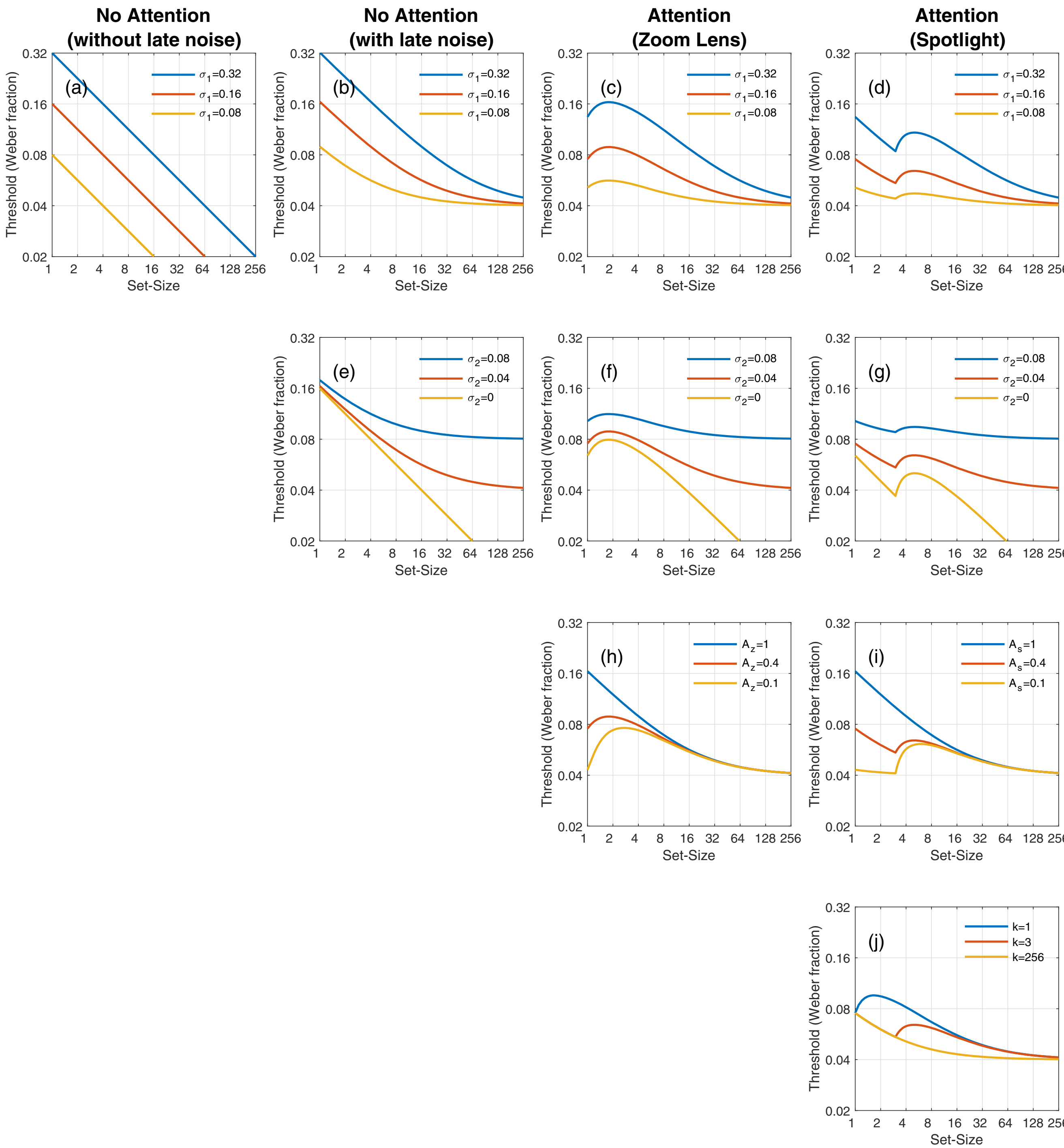

Fig. 2 Predictions of observer models with different parameter values. Each column represents predictions of different models. For all models, thresholds (Weber fractions) are predicted as a function of set-size at three levels (different colors in each panel) of early noise (σ_1), late noise (σ_2), attention factor (A_z or A_s), and attentional limit (k) in the first to fourth rows, respectively

particularly in larger set-sizes. The effects of late noise are exhibited by the asymptotic level in the threshold/set-size function. Attentional effects can be observed by the non-monotonicity of curves in smaller set-sizes (the third row in Fig. 2). In attention models (both the zoom lens and the spotlight models), thresholds for smaller set-sizes decrease with the attentional factor. For example, when $A_z = 0.1$ or $A_s = 0.1$, which implies that attention attenuates early noise down to 10%, perceptual precision is better (i.e., lower threshold) for set-size 1 than for set-sizes 4 or 8 in these figures. When $A_z = 1$ or $A_s = 1$, which implies that attention does not affect perceptual precision, models predict a monotonic and gradual threshold decrement. The fourth row in Fig. 2 shows model predictions for a different attentional limit, k. In the spotlight model, there is a dip and bump around set-size k in the threshold/set-size function. The threshold linearly decreases as set-size increases, reflecting the $\sqrt{n}$ relationship, and attentional effects for set-sizes smaller than k. However, when the set-size n is larger than k, a greater noise for the unselected $n-k$ items yields higher thresholds. Although noises are greater for the unselected $n-k$ items than for the selected k items, they could still cancel each other out, so thresholds decrease if the set-sizes are sufficiently large. Consequently, threshold decreases up to set-size k, increases after k, and gradually decreases again in larger set-sizes far after k. When k is close to 1, the predicted threshold function of the spotlight model (yellow curve in the panel j) is similar to the prediction of the zoom lens model (red curves in the panels c, f, and h). When k is sufficiently large (i.e., infinity), the prediction of the spotlight model (blue curve in the panel j) becomes the same as the prediction of the no-attention model (red curves in the panels b and e).

Empirical tests of the observer models: A new experiment

The goal of this experiment was to identify the nature of the averaging process for multiple inputs and the attentional effects occurring in the process. The proposed models were evaluated through a psychophysical experiment with a 2AFC average size comparison task, which has been commonly featured in previous studies on ensemble perception (Allik et al., 2013; Chong & Treisman, 2003; Gorea, Belkoura, & Solomon, 2014; Lee et al., 2016; Solomon et al., 2011). MATLAB codes for experimental program and data files are available on the Open Science Framework (https://osf.io/ad85r/).

Methods

Participants

Four observers, including the authors J.B. and S.C., participated in the experiment. We followed Smith and Little (2018)'s small-sample-size design and analyzed the data on the individual level. All observers had normal or corrected-to-normal vision, and the two non-authors were unaware of the purpose of the experiment. The study was approved by the Institutional Review Board of Yonsei University. All participants provided written informed consent.

Apparatus

The experiment was conducted on an IBM PC compatible computer, running MATLAB and Psychtoolbox extensions (Brainard, 1997; Pelli & Zhang, 1991). Stimuli were presented on a 21-in. CRT monitor (HP P1230) with 1,600 × 1,200 resolution at a refresh rate of 85 Hz. The viewing distance was approximately 175 cm.

Stimuli

Circles were randomly presented on an invisible 8 × 8 grid (8.94° × 8.94°), excluding the central 2 × 2 grid (2.24° × 2.24°). In each cell of the grid, the position of the circle was randomly jittered within a range of ±0.28° along the vertical and horizontal axes, independently on each trial. The display contained a set of black outlined circles (0.01 cd/m^2) with a stroke width of 0.03° on a gray background (14.43 cd/m^2).

The mean size of the circles was defined using the arithmetic mean of apparent sizes (the diameters with an exponent of 1.52; Chong & Treisman, 2003; Teghtsoonian, 1965). The mean size for the standard display was 556.41 pixel$^{1.52}$ (corresponding to 64 pixels or 0.56° of diameter) with a random jitter from a log-uniform distribution between −16 and +19%. To achieve a targeted mean size, a set of random circles was repeatedly generated, between −30 and +42 % of the targeted mean size, until the mean size of the set matched the targeted mean size.[3] In the set-size 1 trials, the targeted mean was the same as the individual circle size.

Design

The probability of choice (pc) was measured for each mean size difference × set-size conditions. The differences in mean sizes of the two displays were defined as Weber fraction, $\left(\hat{S}_t-\hat{S}_s\right)/\hat{S}_s$, where $\hat{S}_t$ and $\hat{S}_s$ denote apparent mean size of the test and standard sets. The Weber fraction varied over nine levels (−0.244, −0.131, −0.068, −0.034, 0, 0.036, 0.073, 0.150, and 0.323). There were six set-size conditions (1, 2, 4, 8, 16, and 32 items) and the set-size in the test stimulus was always the same as in the standard stimulus. For example, there were four items in both test and standard stimuli in a set-size 4 trial. All experimental conditions, with ten trials in each, were randomly interleaved in each session. Observers partook in ten sessions, each of which lasted approximately 40 min. Thus, the total number of trials was 5,400 (6 set-sizes × 9 mean size differences × 10 trials × 10 sessions) for each observer.

[3] For example, in a standard display of which targeted mean size is 560 pixel$^{1.52}$ in a set-size 4 trial, three circle sizes were generated first (i.e., 664, 414, and 442 pixel$^{1.52}$), then the size of the remaining one (i.e., 720 pixel$^{1.52}$) was determined to match the targeted mean size. Note all these circle sizes were between 395 and 795 pixel$^{1.52}$ (−30 and +42% of the targeted mean size 560 pixel$^{1.52}$).

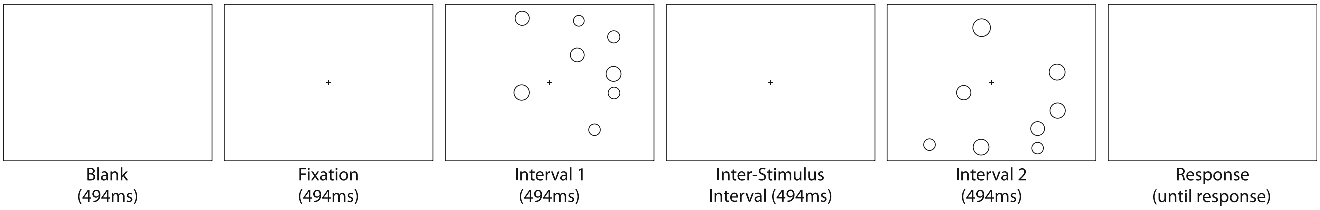

Fig. 3 Illustration of the experimental procedure (example of set-size 8 trial). Following a central fixation, two intervals (each display contains the same number of circles) and an inter-stimulus interval were presented. The observers' task was to determine the interval with the larger mean size

Procedure

An example of a trial sequence is shown in Fig. 3. Each trial began with a small crosshair at the center of the screen. After 494 ms, two stimulus displays were briefly presented (for 494 ms each) one after the other, with an inter-stimulus interval of 494 ms. The standard display was randomly presented in either the first or the second interval, and the test display in the other interval. Observers were asked to identify in which interval the mean size of circles was larger, i.e., a 2AFC paradigm. No feedback was provided after the response, to prevent the formation of bias (Bauer, 2009).

Results

Evaluation of model with late noise

The presence of late noise in the averaging process was evaluated by comparing the goodness of fit of the models with and without late noise: the reduced model with only early noise (σ_1), and the full model with late noise in addition to the early noise (σ_1 and σ_2).

To estimate parameters of models, the analytic models (Eq. 18 with 4 and 5) were fitted to *pc* as a function of mean size difference for the different set-sizes using the maximum likelihood method. The best fitting results are shown in Table 1 and Fig. 4. For all observers, the no-attention model achieved a better goodness of fit than the no-late-noise model (r^2 ranged from .91 to .95 for the no-late-noise model, and from .97 to .99 for the no-attention model).

To test the statistical significance of the improvement in goodness of fit with the additional parameter, the full model, with the late noise component, was compared with the reduced model, without the late noise component, by using the likelihood-ratio test with χ^2 statistics (Cox & Hinkley, 1974):

$$\chi^2(df) = 2\ln\left(\frac{\hat{L}_{full}}{\hat{L}_{reduced}}\right) \tag{1}$$

where $\hat{L}$ is the maximum likelihood, $df = K_{full} - K_{reduced}$ and K is the number of parameters in each model. The results showed that including late noise in the model significantly improved the goodness of fit for all observers (all p's < .001), which indicated that late noise is a significant limiting factor for the averaging process and should be included in the model.

Evaluation of models with attention

Two attention models (Eq. 18 with 6 and 7) were also fitted to the data. These included the attention parameters A_s and k for the spotlight model, and only A_z for the zoom lens model, in addition to σ_1 and σ_2. Both models with attention parameters provided a good explanation for the data. To quantitatively evaluate the observer model and the attentional effects involved, the goodness of fit for the two models were compared: The reduced model with only two parameters (σ_1 and σ_2) and the full model with additional attention parameters (A_z, A_s, and k). The χ^2 statistic for nested models (Eq. 1) was used for the comparison. The results showed that, after taking the increased number of parameters into account, the zoom lens model was superior to the no-attention model (p = .010 for observer JB and p < .001 for the others). The spotlight model also showed significantly better goodness of fit than the no-attention model (p = .035 for observer JB and p < .001 for the others). These results indicate that attention had effects on computing mean size when multiple items were presented in a display, and that reduced noise, attributed to attention for smaller set-sizes, is a major component that accounts for the behavioral performance in mean size perception.

To better distinguish attention models that were not nested to each other, the Akaike information criterion (AIC; Akaike, 1974; Burnham & Anderson, 2004) was computed. Both the likelihood ratio test (with the χ^2 statistic) and the AIC are frequently used for model selection. However, there are distinctions between the two; the likelihood ratio test can compare only nested models (e.g., the no-attention model vs. the zoom lens model, the no-attention model vs. the spotlight model, but not the zoom lens vs. the spotlight model), but the AIC allows the comparison of model quality, regardless of whether the models are nested; however, it does not provide a null hypothesis test. AIC can be computed by

$$AIC = 2k - 2\ln\left(\hat{L}\right) \tag{2}$$

Table 1 Best fitting parameters of proposed models (set-sizes 1–32)

Observer	Model	σ_1	σ_2	A	k	r^2	χ^2	p	AIC
SC	No late noise	0.17	-	-	-	0.95	228.44	<.001	427.99
	No attention	0.06	0.05	-	-	0.99	-	-	201.55
	Attention (zoom lens)	0.12	0.04	0.45	-	0.99	12.70	<.001	190.85
	Attention (spotlight)	0.14	0.04	0.42	1.52	0.99	13.73	<.001	191.82
JB	No late noise	0.29	-	-	-	0.91	212.47	<.001	440.25
	No attention	0.10	0.08	-	-	0.97	-	-	229.79
	Attention (zoom lens)	0.17	0.07	0.54	-	0.98	6.58	.010	225.20
	Attention (spotlight)	0.18	0.07	0.53	1.28	0.98	6.68	.035	227.10
JL	No late noise	0.38	-	-	-	0.92	123.43	<.001	367.76
	No attention	0.15	0.11	-	-	0.97	-	-	246.33
	Attention (zoom lens)	0.30	0.08	0.43	-	0.98	16.27	<.001	232.06
	Attention (spotlight)	0.33	0.07	0.41	1.53	0.98	17.47	<.001	232.87
GS	No late noise	0.31	-	-	-	0.94	181.35	<.001	442.61
	No attention	0.11	0.09	-	-	0.98	-	-	263.26
	Attention (zoom lens)	0.22	0.07	0.45	-	0.99	14.89	<.001	250.37
	Attention (spotlight)	0.26	0.07	0.41	1.64	0.99	17.35	<.001	249.91

χ^2 and its corresponding p value are used for comparing each model to the no-attention model

and the results are interpreted such that a model with a smaller AIC is more probable than the model with a greater AIC. The zoom lens model had a smaller AIC value than the spotlight model for all observers except GS. For the observer GS, the AIC was smaller for the spotlight model than for other models. The relative likelihoods of the zoom lens model to the spotlight model, $e^{((AIC_{spotlight}-AIC_{zoomlens})/2)}$, were 1.63, 2.59, 1.49, and 0.79

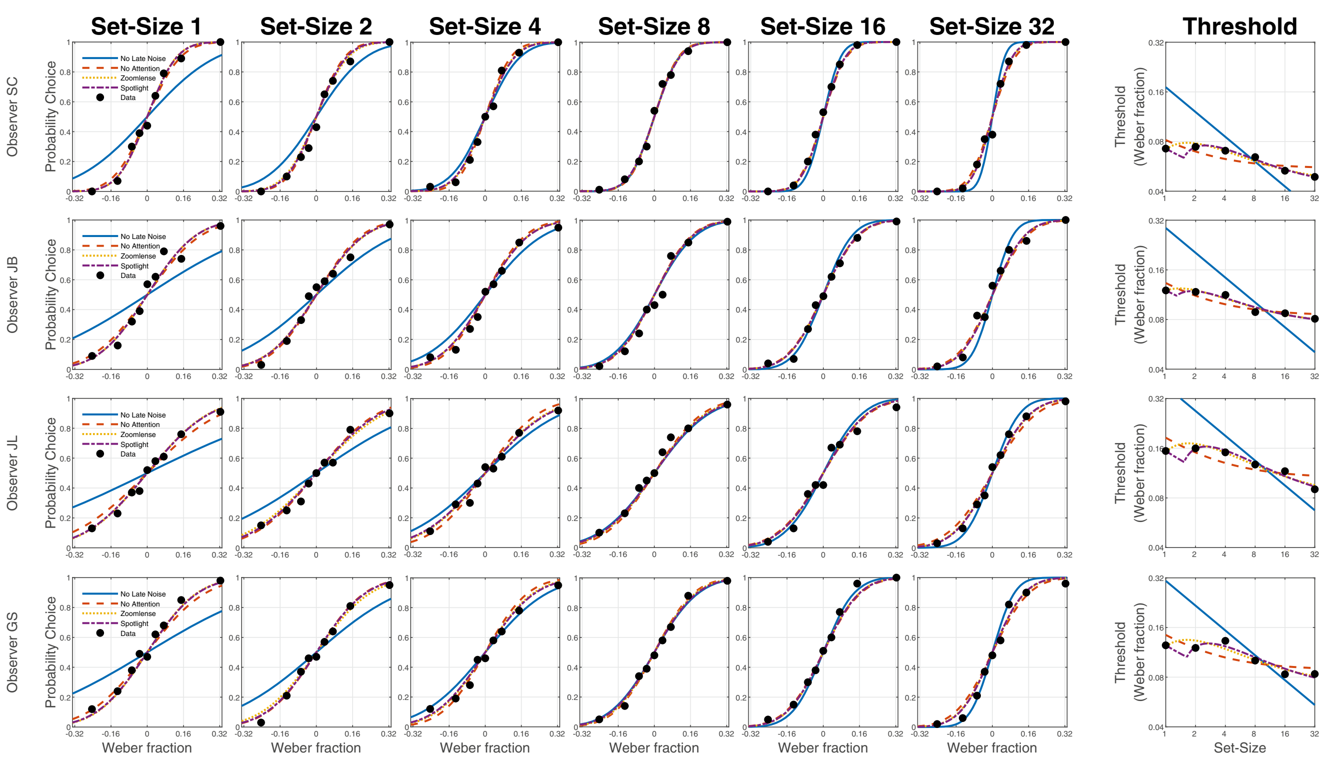

Fig. 4 Psychophysical data and the best fitting models (set-sizes 1–32). In each panel of columns 1–6, the probability of choice, pc (dots) was plotted as a function of test-standard mean size difference along with the best-fits of different models (curves). In the rightmost column, thresholds (dots), corresponding to pc = 0.76, were plotted as a function of set-size along with the best-fitting models for thresholds (curves)

for observers SC, JB, JL, and GS, respectively. That is, the zoom lens model was 1.63, 2.59, 1.49, and 0.79 times more probable than the spotlight model for observers.

Empirical tests of the observer models: Existing evidence

In the previous section, the theoretical predictions of the models were evaluated against the empirical data of our experiment. Often, observer models are developed for a dataset with a certain experimental setting and procedure, but are not suitable for analyzing another dataset. Thus, it would be a good practice to validate observer models with an existing dataset that can be obtained from experiments in literature. Here, the models were evaluated using an existing dataset, Allik et al. (2013),[4] in which the observers' performance in a mean size discrimination task was measured across four set-sizes.

Because most of the experimental settings in Experiment 1 by Allik et al. (2013) were the same as in the present study, our analytic models could be applied to fit their data. The only difference in the experimental procedures was that the standard display contained a single item in their experiment, whereas multiple items were to be averaged in the present experiment. Thus, a slight modification was made in the present analytical models. The analytical models for the 1-to-n comparison are described in Appendix 1 (Eq. 17 with 13–16).

As clearly shown in Fig. 5 and Table 2, all models well accounted for the data, but there were individual differences between models. For observers KA and MT, the no-attention model explained the dataset better than the no-late-noise model (χ^2 = 27.79, p <.001 for KA; χ^2 = 50.01, p <.001 for MT), but worse than the zoom lens model (χ^2 = 5.18, p =.023 for KA; χ^2 = 52.79, p <.001 for MT) and the spotlight model (χ^2 = 5.51, p =.064 for KA; χ^2 = 53.73, p <.001 for MT). Therefore, it could be concluded that late noise and attention are important components for mean size perception in these observers, as in all observers in our experiments.

However, for observer JA, the goodness of fit showed a significant difference between the no-late-noise model and the no-attention model (χ^2 = 4.00, p = .045), but no significant difference between the no-attention model and the attention models (χ^2 = 0.00, p = 1.000 for the zoom lens model; χ^2 = 0.10, p = .951 for the spotlight model).

All observers' AIC showed a preference for the zoom lens model over the spotlight model, and the relative likelihood of the zoom lens model to the spotlight model was 2.58, 2.30, and 1.86 for observers JA, KA, and MT, respectively.

[4] The dataset was extracted from a figure in the original article (Allik et al., 2013).

Empirical tests of an existing observer model

It would also be important to quantitatively compare the present models to existing models in order to gain insights into the advantages of the present models. Thus, we fitted the noise-and-selection model to our dataset.

Since the noise-and-selection model suggested by Allik et al. (2013) is stochastic and does not have an analytic form of the model, we were not able to fit the model to our data directly. Instead, we approximated the model parameters by the grid search algorithm: we simulated observers' responses 100,000 times for each data point with different k values (ranging from 1 to n, with a sampling grain of 0.01 for each set-size n), then selected the k value producing the minimum sum of squared error between the simulated psychometric function and the data as the parameter value for each set size. In the noise-and-selection model, k represented the number of items selected and processed in the averaging process, and ς the standard deviation of internal representation. The parameter ς was estimated by fitting a cumulative Gaussian function to set-size 1 data using a maximum likelihood estimation, as in the study by Allik et al. (2013).

As shown in Table 3 and Fig. 6, the noise-and-selection model provides an excellent account for our dataset (r^2 > .97). However, the AIC values were greater for the noise-and-selection model than for our attention models. The relative likelihood of the zoom lens model to the noise-and-selection model was 17.13, 6.60, 9.51, and 56.73 for observers SC, JB, JL, and GS, respectively. These results suggested that the zoom lens model was superior to the noise-and-selection model, considering the trade-off between the goodness of fit and the simplicity (six parameters, five k and one ς, for the noise-and-selection model vs. three parameters, $\sigma 1$, $\sigma 2$, and A_z, for the zoom lens model) of the models.

Discussion

A distributed attention model was proposed for size averaging. This model assumed the involvement of noise in the early stages of processing for individual size coding, the integration of information from individual items, and the involvement of noise in a later stage for internal representation of the average size. In addition, the components of attention were added to reflect the previous findings regarding the attentional effects in averaging size perception. In the psychophysical validation of the proposed model, the number of items to be averaged in a display was varied, and observers were asked to discern the average size of the display. The distributed attention model accounted well for the observers' averaging performance. Furthermore, it accounted well for the existing mean size discrimination dataset (Allik et al., 2013) and was superior to the noise-and-selection model for explaining our results.

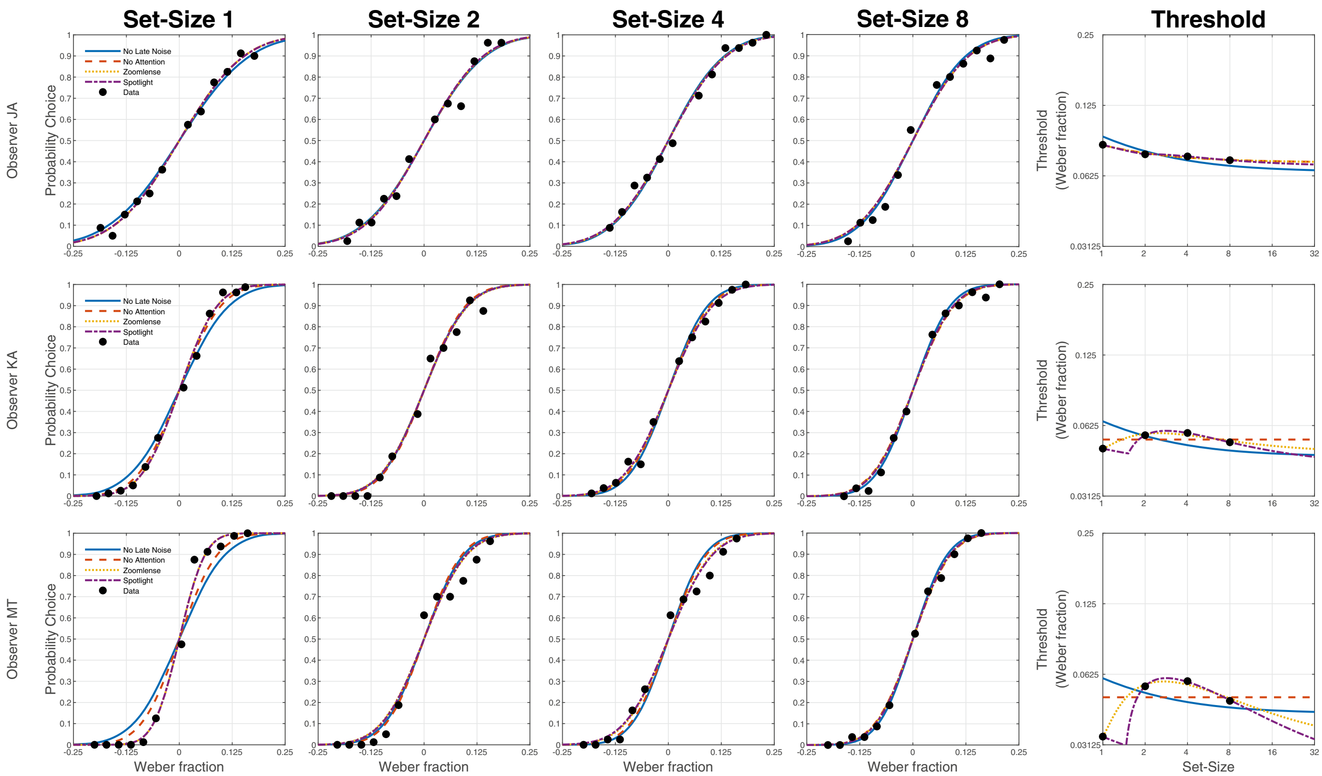

Fig. 5 Psychophysical data from Experiment 1 by Allik et al. (2013) and the best fitting models. In each panel of columns 1–4, probability of choice (dots) was plotted as a function of test-standard mean size difference along with the best-fits of different models (curves). In the rightmost column, thresholds (dots), corresponding to $pc = 0.76$, were plotted as a function of set-size along with the best-fitting models for thresholds (curves)

The strength of the proposed model lies in the component of attention. Although attentional effects in computing average sizes are well known (Chong & Treisman, 2005; de Fockert & Marchant, 2008; McNair, Goodbourn, Shone, & Harris, 2017), some existing models of averaging (Parkes et al., 2001; Setic, Svegar, & Domijan, 2007) do not account for them. In the proposed model, all items contributed to the averaging process, but their contribution varied depending on set-size. Items can contribute to the process equally (zoom lens model) or unequally (spotlight model). Items seem to contribute to the averaging process equally, as the zoom lens model outperformed the spotlight and noise-and-selection models. These results are consistent with the notion that ensembles are better computed under the mode of distributed attention than under focused attention (Chong & Evans, 2011; Treisman, 2006).

Table 2 Best fitting parameters of proposed models: data from Experiment 1 by Allik et al. (2013)

Observer	Model	σ_1	σ_2	A	k	r^2	χ^2	p	AIC
JA	No late noise	0.09	-	-	-	0.99	4.00	.045	244.69
	No attention	0.06	0.06	-	-	0.99	-	-	242.69
	Attention (zoom lens)	0.06	0.06	1.00	-	0.99	0.00	1.000	244.69
	Attention (spotlight)	0.11	0.05	0.67	1.99	0.99	0.10	.951	246.58
KA	No late noise	0.07	-	-	-	0.99	27.79	<.001	245.94
	No attention	0.00	0.05	-	-	0.99	-	-	220.15
	Attention (zoom lens)	0.11	0.05	0.17	-	0.99	5.18	.023	216.98
	Attention (spotlight)	0.14	0.03	0.26	1.52	0.99	5.51	.064	218.64
MT	No late noise	0.06	-	-	-	0.97	50.01	<.001	320.54
	No attention	0.00	0.05	-	-	0.98	-	-	272.54
	Attention (zoom lens)	0.17	0.03	0.10	-	0.99	52.97	<.001	221.56
	Attention (spotlight)	0.19	0.00	0.18	1.47	0.99	53.73	<.001	222.80

Table 3 Best fitting parameters of the noise-and-selection model: data from our experiment

Observer	ς	k_2	k_4	k_8	k_{16}	k_{32}	r^2	AIC
SC	0.07	1.00	1.32	1.73	2.26	2.59	.99	196.53
JB	0.12	1.00	1.05	1.70	2.27	2.51	.98	228.98
JL	0.15	1.00	1.17	1.15	1.75	2.19	.98	236.57
GS	0.12	1.00	1.08	1.14	1.85	2.87	.99	258.45

The proposed model also confirmed the existence of a late noise component in the averaging process (Parkes et al., 2001; Solomon, 2010; Solomon et al., 2011). Even if sufficient items are presented, observers will still make some errors in perceiving the average size. That is, for a set-size close to infinity, the threshold will not drop to zero, but rather become saturated at a certain level. The limit of the threshold is often explained by late noise. Several observer models have been constructed with a late noise component to characterize internal responses in tasks that require feature precision (Heeley & Buchanan-Smith, 1996; Morgan et al., 1990; Parkes et al., 2001) and contrast sensitivity (Burgess, Wagner, Jennings, & Barlow, 1981; Eckstein, Ahumada, & Watson, 1997; Lu & Dosher, 1999; Pelli, 1985).

Similar to other visual tasks, the average computation has limitations. It could be inaccurate because (1) noise is involved in the averaging process (the current paper; Allik et al., 2013; Solomon et al., 2011), (2) only selected items contribute to it (Allik et al., 2013; Myczek & Simons, 2008; Solomon et al., 2011), or (3) both the noisy percept and the selection process reduce the precision of mean computation (Allik et al., 2013; Solomon et al., 2011). The proposed attention models focused on the noisy percept and its interaction with attention because attention is believed to reduce noise in the perceptual processes (Dosher & Lu, 2000b, 2000a; Lu & Dosher, 1998). Furthermore, it is unlikely that the visual system uses only selected items for averaging because unattended information contributes to visual processing, even up to a semantic level (Treisman, 1969; Wolford & Morrison, 1980). There are also empirical data that all items contribute to the averaging process. Chong et al. (Chong, Joo,

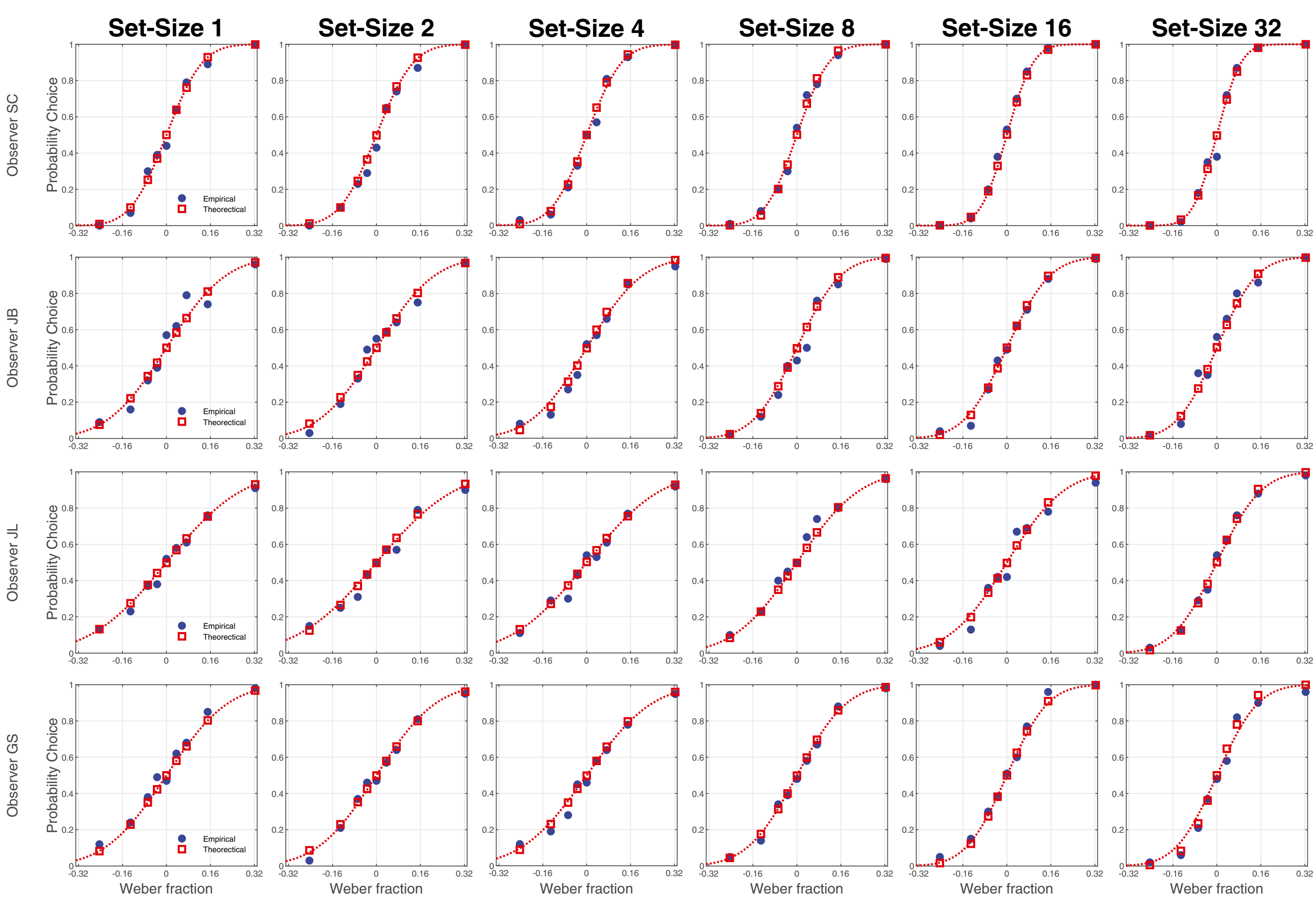

Fig. 6 Psychophysical data and the best fit of the noise-and-selection model. Probability of choice (dots) was plotted as a function of test-standard mean size difference along with the best-fits of the noise-and-selection model (red squares). Observers' theoretical responses were simulated with 100,000 trials for each data point. The best-fitting cumulative Gaussian distribution was presented with a red dotted curve in each panel

Emmanouil, & Treisman, 2008) found that the performance was inferior when only a portion of the entire display was presented compared to when the entire display was presented. The averaging performance decreased as the number of items to be included in the averaging decreased because they became invisible (Choo & Franconeri, 2010; Joo, Shin, Chong, & Blake, 2009). Our findings supported the idea that the visual system integrates information from unselected items, at least to some extent. However, we must acknowledge that other observer models with a filtering-out mechanism could also explain the existing averaging performance (Allik et al., 2013; Solomon et al., 2011). The noise-and-selection model also explained the results of the current study, although inferior to the zoom lens model. Further studies are necessary to differentiate between the two potential mechanisms of attention in the averaging process: enhancing the representational precision for attended items or filtering out unattended items.

The attention models adopt different assumptions from the noise-and-selection model regarding the size of the attentional window, which represents the limited capacity of an observer. In the proposed attention models, it was assumed that there was a fixed number for all set-sizes, rather than a varying number depending on the set-size, because an observer's intrinsic capacity limit should not change for different inputs (e.g., set-size; but see also Solomon, 2010; Solomon et al., 2011). This is unlimited in the zoom lens model and represented by a constant, *k,* in the spotlight model. By contrast, the noise-and-selection model (Allik et al., 2013) does not assume a fixed size of the attentional window for different set sizes. Another important difference between the two models lies in the existence of late noise. The present data clearly demonstrated the saturation of the threshold decrement in large set-sizes, and this decrement could be well accounted for by the late noise component.

Although the proposed attention model explains the observers' perception of mean size, it has limitations in accounting for some phenomena. For example, Solomon et al. (2011) found that the size discrimination obeys Weber's law, which the proposed models did not incorporate. In addition, we did not control the precision of individual sizes and assumed that they are encoded with equal precision. The precision of individual sizes might have decreased with increasing set-size because smaller separations between items in larger set-sizes could have produced crowding (van den Berg, Roerdink, & Cornelissen, 2007). Note that Solomon et al. (2011) did not find crowding in discriminating individual sizes among eight items.

In summary, a distributed attention model of averaging was developed and tested by varying the number of items to be averaged. It predicted the observed data almost perfectly. The components of attention and late noise played an important role not only in predicting the observed data but also in explaining existing data on averaging. The computational components in the proposed model (early noise, late noise, and attentional modulation) are sufficiently general that they can be applied to other feature averaging applications.

Acknowledgements This research was supported by the Brain Research Program (NRF-2017M3C7A1029658) through the National Research Foundation of Korea (NRF) to SC and the ICT Consilience Creative program (IITP-2017-2017-0-01015) through the Institute for Information and Communications Technology Promotion (IITP) funded by the Ministry of Science and ICT. Some of the data was presented at the 2018 Vision Sciences Society. For helpful comments and discussion about this manuscript, we thank Jüri Allik, Randolph Blake, Matthew Inverso, Min-Suk Kang, Joshua A. Solomon, and an anonymous reviewer. Anne Treisman was a wonderful mentor and role model for SC, and SC is eternally grateful for her guidance.

Open Practices Statement The data and materials for all experiments are available at the Open Science Framework (https://osf.io/ad85r/), and none of the experiments was preregistered.

Appendix 1. Analytic model of mean size perception

Noisy percept for individual items: Early noise

Each stimulus passes through a feature detector, which translates it into an internal response of a certain magnitude. The output of the detector is proportional to the exponent of the signal strength (Foley & Legge, 1981; Nachmias, 1981; Nachmias & Sansbury, 1974). Regarding size perception, it is known that the perceived size of a circle is related to its physical size by a power function with an exponent of 0.76 (Teghtsoonian, 1965). By combining the output of the detector with the early noise, a noisy internal response is produced for each stimulus. An internal response for a circle size with diameter d is described as a random variable from a Gaussian probability density function, $\mathcal{N}\left(\mu, \zeta^2\right)$, where μ is the apparent size of the circle $S = (d^2)^{.76} = d^{1.52}$, and ζ is the standard deviation of the Gaussian early noise for each input, σ_I.[5]

Averaging multiple items

Then, the visual system attempts to integrate all internal responses generated from multiple inputs. The percept of the average size is a random variable with a Gaussian probability distribution, $\mathcal{N}\left(\mu, \zeta^2\right)$, whose center is at the mean of the circle sizes and whose standard deviation decreases with the square root of set-size, $\sqrt{n}$. Therefore, the expected value of the internal response for the average size is proportional to the arithmetic mean of apparent sizes (Chong & Treisman, 2003):

[5] It should be noted that the apparent size, *S,* has a physical unit (e.g., $pixel^{1.52}$) whereas μ is an internal quantity with an arbitrary unit.

$$\mu = \hat{S} = \frac{\sum d_i^{1.52}}{n} \quad (3)$$

where d_i is the diameter of i-th item. The standard deviation of the internal response distribution for the average size is inversely proportional to the square root of the set-size:

$$\zeta = \frac{\sigma_1}{\sqrt{n}} \quad (4)$$

Noisy percept for the averaged representation: Late noise

Late noise is added to the internal response for the average size of multiple items. Because late noise and the internal response for average size are both random variables with certain variabilities, the variance of the noisy internal response distribution after late noise can be calculated as follows:

$$\zeta = \sqrt{\frac{{\sigma_1}^2}{n} + {\sigma_2}^2} \quad (5)$$

where σ_2 is the standard deviation of the zero-mean Gaussian late noise.

Zoom lens model: Distributed attention with equal effects on all items

In the zoom lens model, attention parameter, A_z, is a noise-reduction factor multiplied to early noise, σ_1. That is, a small A_z value signifies that attention greatly reduces early noise. The model posits that attentional benefits, 1 - A_z, are inversely related to set-size. For example, if the attention parameter A_z is 0.7, which implies that attention reduces a random noise by a factor of 0.3 (= 1 − 0.7), 0.15, 0.075, 0.0375, 0.0188, and 0.0094 for set-sizes 1, 2, 4, 8, 16, and 32, respectively. Thus, the multiplication factors are 0.7, 0.85, 0.925, 0.9625, 0.9812, and 0.9906 for set-sizes 1, 2, 4, 8, 16, and 32, respectively. Therefore, by replacing σ_1 with $\sigma_1\left(\frac{n-1+A_z}{n}\right)$ in Eq. 5, the variability of the internal response for the average size should be

$$\zeta = \sqrt{\frac{(n-1+A_z)^2}{n^3}{\sigma_1}^2 + {\sigma_2}^2} \quad (6)$$

Spotlight model: Distributed attention with unequal effects on all items

The spotlight model posits that attention selects a fixed number of items and that the visual system processes the selected items with high precision, with the unselected items consigned to low precision representations (Posner, 1980). Therefore, in the proposed analytic model, the early noise for k attended items is $A_s \times \sigma_1$, but is only σ_1 for the other $n-k$ items. By replacing σ_1 with $\sigma_1\sqrt{\frac{{A_s}^2k+n-k}{n}}$ (when $n > k$) or $A_s\sigma_1$ (when $n \leq k$) in Eq. 4, the variability of the internal response for the average size should be

$$\begin{cases} \zeta = \sqrt{\frac{({A_s}^2k+n-k)}{n^2}{\sigma_1}^2 + {\sigma_2}^2}, \text{if } n > k \\ \zeta = \sqrt{\frac{{A_s}^2}{n}{\sigma_1}^2 + {\sigma_2}^2}, \text{if } n \leq k \end{cases} \quad (7)$$

Comparing average sizes: Signal detection theory

For a two-alternative forced-choice (2AFC) task, the visual system compares two random samples, one from each internal response distribution: $\mathcal{N}(\mu_t, {\zeta_t}^2)$ for the test and $\mathcal{N}(\mu_s, {\zeta_s}^2)$ for the standard. Since μ_t and μ_t are the expected responses for the standard and test ensemble sets, the expected value of difference of the internal responses for the test and standard sets is proportional to the Weber fraction between apparent mean sizes: $w = (\hat{S}_t - \hat{S}_s)/\hat{S}_s$. In comparison, the perceptual sensitivity d' can be defined as a signal-to-noise ratio, namely,

$$d' = \frac{w}{\sqrt{\frac{\zeta_t^2 + \zeta_s^2}{2}}} \quad (8)$$

Comparing *N*-to-*N*

When an observer compares the average sizes of two displays with the same set-size, that is, $\zeta_t = \zeta_s = \zeta$, Eq. 8 can be simplified as:

$$d' = \frac{w}{\zeta} \quad (9)$$

By substituting Eqs. 4–7 into Eq. 9, d' can be calculated by

$$d' = \frac{w}{\sqrt{\frac{{\sigma_1}^2}{n} + {\sigma_2}^2}} \quad (10)$$

$$d' = \frac{w}{\sqrt{\frac{(n-1+A_z)^2}{n^3}{\sigma_1}^2 + {\sigma_2}^2}} \quad (11)$$

$$\begin{cases} d' = \frac{w}{\sqrt{\frac{({A_s}^2k+n-k)}{n^2}{\sigma_1}^2 + {\sigma_2}^2}}, \text{if } n > k \\ d' = \frac{w}{\sqrt{\frac{{A_s}^2}{n}{\sigma_1}^2 + {\sigma_2}^2}}, \text{if } n \leq k \end{cases} \quad (12)$$

for the no-late-noise model, the no-attention model, the zoom lens model, and the spotlight model, respectively.

Comparing 1-to-*N*

When an observer is required to compare a standard display containing only a single item to a test display containing multiple items, the internal response distributions are different. For the test display, in which there are n items, the variability of internal response distributions σ_t can be computed by Eqs. 4–7. For the standard display, the variability σ_s can be computed by setting n to 1 in Eqs. 4–7. By substituting σ_s and σ_t into Eq. 8, d' function can be simplified as:

$$d' = \frac{w}{\sqrt{\frac{n+1}{2n}\sigma_1^2}} \tag{13}$$

$$d' = \frac{w}{\sqrt{\frac{n+1}{2n}\sigma_1^2 + \sigma_2^2}} \tag{14}$$

$$d' = \frac{w}{\sqrt{\frac{{A_z}^2 n^3 + (n-1+A_z)^2}{2n^3}\sigma_1^2 + \sigma_2^2}} \tag{15}$$

$$\begin{cases} d' = \dfrac{w}{\sqrt{\frac{{A_s}^2(n^2+k)+n-k}{2n^2}\sigma_1^2 + \sigma_2^2}}, \text{if } n > k \\ d' = \dfrac{w}{\sqrt{\frac{(n+1){A_s}^2}{2n}\sigma_1^2 + \sigma_2^2}}, \text{if } n \le k \end{cases} \tag{16}$$

Behavioral signatures of proposed models

An observer's probability of choice, pc, can be determined by the sensitivity and response bias (toward the first or second intervals, or toward the test or standard displays). By the signal detection theory, this can be calculated as:

$$pc = \Phi\left(\frac{d'}{\sqrt{2}}\right) \tag{17}$$

in a 2AFC task, in which the response bias is known to be minimal. For the comparison of n-to-n items, it can be calculated by plugging in Eq. 9 into 17,

$$pc = \Phi\left(\frac{d'}{\sqrt{2}}\right) = \Phi\left(\frac{w}{\zeta\sqrt{2}}\right) \tag{18}$$

where $\Phi(x)$ is the standard normal cumulative density function. The threshold for a targeted probability choice, t is:

$$\theta_{pc=t} = \Phi^{-1}(t) \times \zeta\sqrt{2} \tag{19}$$

where Φ^{-1} is the inversed standard normal cumulative density function. For example,

$$\theta_{pc=0.76} = \zeta \tag{20}$$

The threshold can be also defined as the reciprocal of the sensitivity d'. Thus, when there is no response bias, the threshold for a certain level of sensitivity, t, can be described as

$$\theta_{d'=t} = t \times \zeta \tag{21}$$

For example,

$$\theta_{d'=1} = \zeta \tag{22}$$

In a special case comparing 1-to-n, the pc can be computed by plugging 13–16 into 17.

Appendix 2. Evaluation of models with set-sizes 2–32

One might argue that no averaging process is necessary when an observer is asked to report the average size of a single item. Clearly, the average size of a single item is the actual size. Therefore, it would be reasonable that the averaging process for the display with set-size 1 should be distinguished from the process for larger set-size displays. To evaluate the observer model and the attentional effects while excluding the special property of the averaging process for a display with set-size 1, we fitted models to pc of set-sizes 2–32.

The same data analysis as for set-sizes 1–32 (see the main text) was performed for this analysis. The best fitting results are shown in Table 4 and Fig. 7. For all observers, the comparison of models with and without late noise showed that the goodness of fit of the observer models was significantly improved by including a late noise component (for all observers, $p < .001$). In the comparison between the no-attention model and the attention models, the results showed that the additional attention parameters significantly improved the goodness of fit for all observers except JB (for JB, $\chi^2 = 1.00$, $p = .316$ for the zoom lens model, and $\chi^2 = 1.06$, $p = .588$ for the spotlight model). The AIC value was also smaller for the zoom lens model than for the spotlight model, indicating that the former was more probable than the latter. The relative likelihood of the spotlight model to the zoom lens model was 2.54, 2.64, 2.38, and 2.07 for observers SC, JB, JL, and GS, respectively.

Table 4 Best fitting parameters of proposed models (set-sizes 2–32)

Observer	Model	σ_1	σ_2	A	k	r^2	χ^2	p	AIC
SC	No late noise	0.19	-	-	-	0.97	115.20	<.001	276.91
	No attention	0.09	0.05	-	-	0.99	-	-	163.71
	Attention (zoom lens)	0.15	0.04	0.18	-	0.99	4.61	.032	161.11
	Attention (spotlight)	0.16	0.04	0.56	2.68	0.99	4.74	.093	162.97
JB	No late noise	0.31	-	-	-	0.94	124.21	<.001	296.17
	No attention	0.14	0.08	-	-	0.98	-	-	173.95
	Attention (zoom lens)	0.18	0.07	0.46	-	0.98	1.00	.316	174.95
	Attention (spotlight)	0.18	0.07	0.28	1.00	0.98	1.06	.588	176.89
JL	No late noise	0.40	-	-	-	0.95	62.00	<.001	265.34
	No attention	0.22	0.09	-	-	0.98	-	-	205.34
	Attention (zoom lens)	0.35	0.07	0.15	-	0.98	5.81	.016	201.53
	Attention (spotlight)	0.38	0.07	0.54	2.81	0.98	6.07	.048	203.26
GS	No late noise	0.33	-	-	-	0.96	99.08	<.001	311.27
	No attention	0.16	0.08	-	-	0.98	-	-	214.19
	Attention (zoom lens)	0.26	0.07	0.12	-	0.99	6.65	.010	209.53
	Attention (spotlight)	0.26	0.07	0.16	1.39	0.99	7.19	.027	210.99

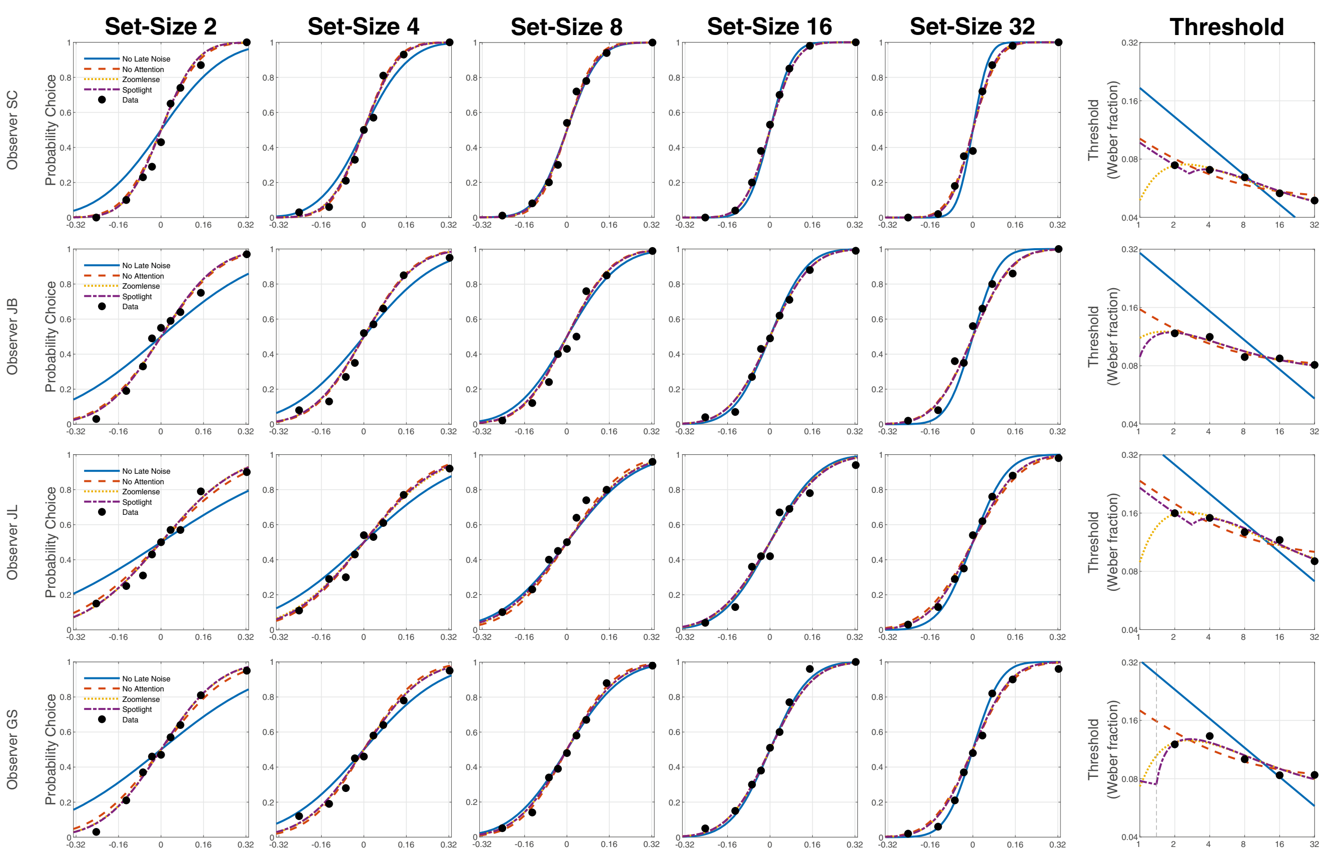

Fig. 7 Psychophysical data and the best fitting models (set-sizes 2–32). In each panel of columns 1–6, probability of choice, *pc* (dots) was plotted as a function of test-standard mean size difference along with the best-fits of different models (curves). In the right-most column, thresholds (dots), corresponding to *pc* = 0.76, were plotted as a function of set-size along with the best-fitting models for thresholds (curves)

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactionson Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39. https://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*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception & Psychophysics*, *71*(3), 655–665. https://doi.org/10.3758/APP.71.3.655

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, *214*(4516), 93–94. https://doi.org/10.1126/science.7280685

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model selection and multimodel inference*. New York, NY: Springer. https://doi.org/10.1007/b97636

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Chong, S. C., & Evans, K. K. (2011). Distributed vs. focused attention (count vs. estimate). *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(6), 634–638. https://doi.org/10.1002/wcs.136

Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334. https://doi.org/10.3758/PP.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12535996

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15912869

Choo, H., & Franconeri, S. L. (2010). Objects with reduced visibility still contribute to size averaging. *Attention, Perception, & Psychophysics, 72*(1), 86–99. https://doi.org/10.3758/APP.72.1.86

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. London: Chapman and Hall/CRC.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://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*(5), 789–794. https://doi.org/10.3758/PP.70.5.789

de Fockert, J. W., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

Dosher, B. A., & Lu, Z.-L. (2000a). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10–12), 1269–1292. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10788639

Dosher, B. A., & Lu, Z.-L. (2000b). Noise exclusion in spatial attention. *Psychological Science*, *11*(2), 139–146. https://doi.org/10.1111/1467-9280.00229

Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *14*(9), 2406–2419. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9291610

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, *21*(7), 1041–1053. https://doi.org/10.1016/0042-6989(81)90009-2

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision, 14*(9), 22–22. https://doi.org/10.1167/14.9.22

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology : CB*, *17*(17), R751-3. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Heeley, D. W., & Buchanan-Smith, H. M. (1996). Mechanisms specialized for the perception of image geometry. *Vision Research*, *36*(22), 3607–3627. https://doi.org/10.1016/0042-6989(96)00077-6

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision, 9*(9), 1–12. https://doi.org/10.1167/9.9.7

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, *2*(4), 375–381. https://doi.org/10.1038/7286

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

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198. https://doi.org/10.1016/S0042-6989(97)00273-3

Lu, Z.-L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, *16*(3), 764. https://doi.org/10.1364/JOSAA.16.000764

Lu, Z.-L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*(1), 44–82. https://doi.org/10.1037/0033-295X.115.1.44

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide* (1st ed.). New York: Cambridge Univesity Press.

Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *31*(4), A93-102. https://doi.org/10.1364/JOSAA.31.000A93

McNair, N. A., Goodbourn, P. T., Shone, L. T., & Harris, I. M. (2017). Summary statistics in the attentional blink. *Attention, Perception & Psychophysics*, *79*(1), 100–116. https://doi.org/10.3758/s13414-016-1216-2

Morgan, M. J., Ward, R. M., & Hole, G. J. (1990). Evidence for positional coding in hyperacuity. *Journal of the Optical Society of America. A, Optics and Image Science*, *7*(2), 297–304. https://doi.org/10.1364/JOSAA.7.000297

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & psychophysics, 70*(5), 772-788. https://doi.org/10.3758/PP.70.5.772

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, *21*(2), 215–223. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7269298

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research*, *14*(10), 1039–1042. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4432385

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56–63. https://doi.org/10.1016/j.cognition.2013.03.006

Palmer, E. M., Fencsik, D. E., Flusberg, S. J., Horowitz, T. S., & Wolfe, J. M. (2011). Signal detection evidence for limited capacity in visual search. *Attention, Perception & Psychophysics*, *73*(8), 2413–2424. https://doi.org/10.3758/s13414-011-0199-2

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. https://doi.org/10.1038/89532

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America. A, Optics and Image Science*, *2*(9), 1508–1532. https://doi.org/10.1364/JOSAA.2.001508

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on micro-computer displays. *Vision Research*, *31*(7–8), 1337–1350. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1891822

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(12), 1–8. https://doi.org/10.1167/11.12.18

Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of Vision, 8*(15), 3–3. https://doi.org/10.1167/8.15.3

Setic, M., Svegar, D., & Domijan, D. (2007). Modelling the statistical processing of visual information. *Neurocomputing*, *70*(10–12), 1808–1812. https://doi.org/10.1016/j.neucom.2006.10.069

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review, 25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14), 19–19. https://doi.org/10.1167/10.14.19

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

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, *78*(3), 392. https://doi.org/10.2307/1420573

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/h0027242

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision, 7*(2), 14. https://doi.org/10.1167/7.2.14

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-I

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*(1), annurev-psych-010416-044232. https://doi.org/10.1146/annurev-psych-010416-044232

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*(1), 55–62. https://doi.org/10.1016/0042-6989(84)90144-5

Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, *8*(6), 521–527. https://doi.org/10.3758/BF03213771

Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186. https://doi.org/10.1038/ncomms13186

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