Comparing Global and Limited sampling Strategies in Size-averaging a Set of Items

Midori Tokita (Tokita.Midori@ocha.ac.jp)
Department of Letters and Education,
Ochanomizu University, Otsuka, Tokyo 112-8610 Japan

Akira Ishiguchi (Ishiguchi.Akira@ocha.ac.jp)
Department of Human Developmental Sciences,
Ochanomizu University, Otsuka, Tokyo 112-8610 Japan

Abstract

Many studies have shown that our visual system may construct a "statistical summary representation" over groups of visual objects. Although there is a general understanding that human observers can accurately represent sets of a variety of features, many questions on how the statistical summary is computed still remain unanswered. This study investigated sampling properties of visual information used by human observers in deriving an average representation of a set of items. We presented three models of ideal observers to perform a size averaging task: a global averaging model without item noise (GAM1), a global averaging model with item noise (GAM2), and a limited sampling model (LSM). We compared the performance of the ideal observer of each model to the performance of human observers using statistical efficiency analysis. Our results suggest that average size of items in a set may be computed without representing individual items, discarding the limited sampling model.

Keywords: statistical summary representation; Ideal observer analysis; size averaging; attention

Introduction

Many studies have shown that people accurately perceive and estimate the statistical properties of a set of items or events. For example, the visual system may construct a "statistical summary representation" over groups of visual objects (e.g., Alvarez, 2011; Ariely, 2001, 2008; Chong & Treisman, 2003). It has been shown that observers are able to quickly and accurately extract average values over a range of visual properties, including size (Chong & Treisman, 2005 ; Oriet & Brand, 2013), brightness (Bauer, 2009), orientation (Parkes, Liend, Angelucci, Solomon & Morgan, 2001), emotional expression (Haberman & Whitney, 2009, 2011), and others. Moreover, this ability is not limited to static and simultaneous events; it is observed in sequentially presented events （Oriet & Corbett, 2008; Whiting & Oriet, 2011）and dynamic objects, such as expanding and contracting circles (Albrecht & Scholl, 2010). In recent studies, it has been shown that the ability of representing statistical properties is not limited to visual properties but is also observed in auditory mechanisms such as extracting frequency information from sequences of sounds (Piazza, Sweeny, Wessel, Silver, & Whitney, 2013).

Although there is a general understanding that human observers can accurately represent sets of features, many questions on how the statistical summary is computed still remain unanswered. Three possibilities have been proposed: 1) representations of individual items are computed first and then combined to form a summary representation, 2) summary representations are computed without computing individual items, or 3) Only a couple of items in a set are sampled and included in the calculation of the average size. The first and second proposals predict that there are specialized statistical summary representation mechanisms that are separate from the mechanisms mediated to represent individual objects. Conforming to this argument, many studies have provided evidence that, when attention is distributed across a set of similar items, people can extract the average size of all the items without relying on focused attention to the individual items in the set (e.g., Chong & Treisman, 2003, 2005; Im & Halberda, 2013; Treisman, 2006). The third proposal claims that it is possible to accurately estimate the average size by sampling a couple of items in a set using focused attention. Modeling research has shown that a sampling strategy reasonably predicts the approximate levels of performance exhibited by observers in studies of average size perception (Myczek & Simons, 2008; Marchant, Simons & Fockert, 2013).

Neither the proponents of the summary representation mechanisms nor those of the limited sampling strategy have excluded or refuted the opposing argument. Rather, they prompt the necessity for further investigation on the processes of human performance of statistical summary mechanisms (Ariely, 2008; Simons & Myczek, 2008). Overall, it is necessary to examine if people estimate the average size by using the global information of all items in a set or of the limited number of items in a set.

The present study investigated sampling properties of visual information used by human observers in deriving an average representation of items in a set using the ideal observer (IO) analysis. We measured performance on a size averaging task for each ideal observer model and for human observers. Next, we compared the performance of the ideal observer of each model to the performance of human observers to evaluate which model could predict how human observers derive the average size of items in a set. While comparing, we used a statistical efficiency analysis that allows direct comparison of efficiencies among different models that represent different uses of information. Statistical efficiency is a relative index for the sampling rate of information in a given task. Many studies have utilized the efficiency to investigate how the visual system uses available information and revealed the characteristics of human performance (e.g. Tanaka & Ishiguchi, 2006; Ikeda & Ishiguchi, 2004; Watamaniuk, 1993). The fluctuation of efficiency has been also useful to explore the characteristics of the human sampling strategy.

Based on the three proposals mentioned above, we presented three models of ideal observers to perform the task: a global averaging model without item noise (GAM1), a global averaging model with item noise (GAM2), and a limited sampling model (LSM).

First, we described three ideal observer models and the statistical efficiency analysis in detail. Second, we tested the size discrimination threshold of human observers in each experimental condition to determine values for free parameters used in simulating the performance of the ideal observer models. Next, we conducted the size averaging task to the human observers and calculated the efficiency of the human observers to the ideal observers to evaluate the model appropriate for the averaging process.

Ideal observer models

An ideal observer is a theoretical device that performs a given task in an optimal manner with the available information and some specified constraints (e.g., Geisler, 2003; Yakushijin, 2007). We presented three ideal observer models to perform the size averaging task: a global averaging model without item noise (GAM1), a global averaging model with item noise (GAM2), and a limited sampling model (LSM). Figure 1 shows diagrams of each model. Each model comprises a sampling process, a summary representation process, and a decision process. There are two types of noises involved in a given process: an item noise added to each item in a set prior to summary representation (σ_{Item}) and a late noise added to an estimated average size (σ_{Late}).

GAM1 posits that summary representations are computed without computing individual items and used for calculating their average size. Thus no noise added to each item in a set and only late noise added to representation of average size. GAM2 posits that representations of individual items are computed first and then combined for forming a summary representation and subsequent calculation of their average size. At the representation of individual items, the noise is added to each item prior to form the summary representation. The value of the noise may depend on the set size, since it has been predicted the noise with which each individual item is represented increases as the number of items increases since each item receive less attention (Palmer, 1990; Franconeri, Alvarez, & Enns, 2007). A late noise is added to the average size of items in a set and compared to the test item.

LSM posits that people sample a couple of items randomly chosen from a set to calculate their average size. In this study, we randomly sampled two items from a set since it has been shown that the average of the set could be accurately estimated by sampling two items, estimating the average of those items alone (Myczek & Simons, 2008). The item noise added to each item prior to form the summary representation and then the late noise added to the average size prior to the decision process.

The performance of each model, discriminability d'_{Ideal}, was derived using the Monte Carlo method. Free parameters of the models, SD for an item noise of each set size (σ_{ItemN}) and the late noise (σ_{Late}), were determined by the experiment.

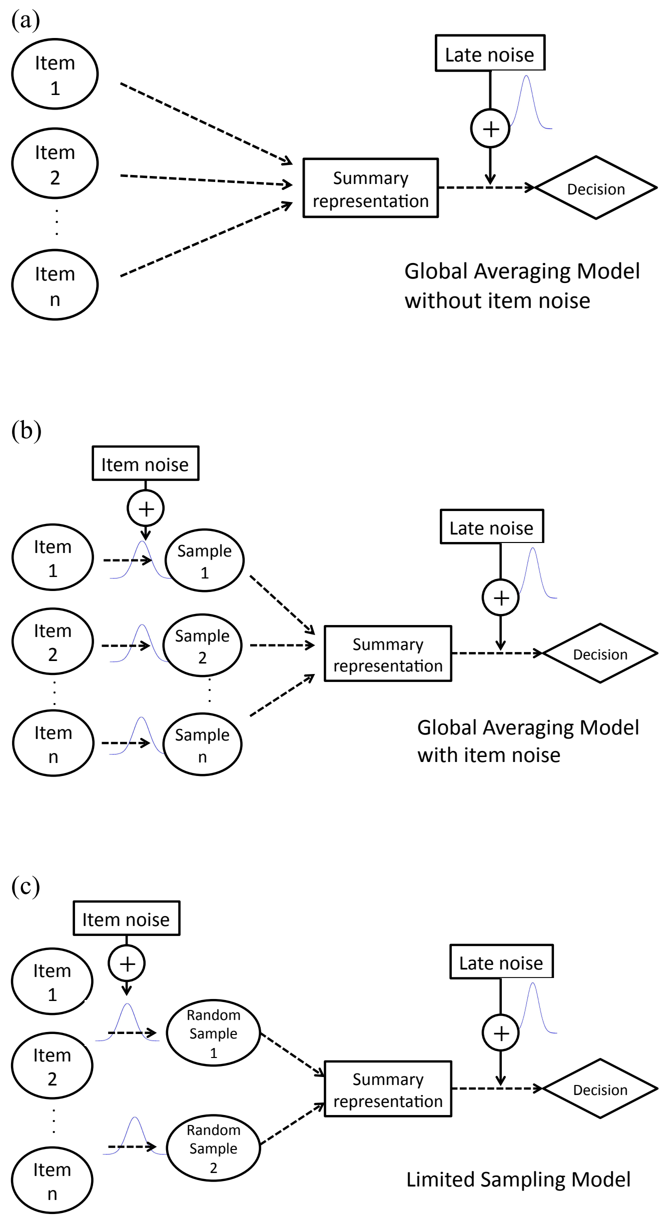

Figure 1: Description of three ideal observer models; Global Averaging Model without item noise (GAM1); Global Averaging Model with item noise (GAM2); Limited Sampling Model (LSM).

Calculation of statistical efficiency

We calculated statistical efficiencies using the following method. First, we prepared two stimuli: a set of multiple items and a single test item. Both stimuli consisted of solid circles with a given diameter. The diameter of the test item was larger or smaller than the average size of items in a set. We added independent lognormal noise $lnN(0, \sigma_c^2)$ to the diameter of individual items in the set. No noise was added to the test item. The observer's task was to determine which item, the average size of items in a set or the test item, was lager in size, using the two alternative forced choice (2AFC) procedure. Ideal observer of each model sampled the values of information available in each model. The discriminability for each ideal observer (d'_{Ideal}) is obtained by performing monte carlo simulation. The discriminability for human observer (d'_H) was determined by performing the same 2AFC task and calculated as follows:

$$d'_H = \sqrt{2} z_c, \qquad (1)$$

where z_c is the z-value transformed from the observer's percent correct. The statistical efficiency F was defined by the square of the ratio of these two scores as

$$F = (d'_H / d'_{Ideal})^2. \qquad (2)$$

The details of calculation of statistical efficiency were discussed in Barlow (1978) and Watamaniuk (1993).

Many studies have found that the sampling capacity of visual system is limited. If the statistical efficiency is larger than 100%, we could infer that the model may not describe the appropriate process to perform the task.

Size Discrimination Experiment

The purpose of this experiment is to obtain the value of free parameters σ_{ItemN} and σ_{Late}. We measured accuracies for size discrimination in a single item condition and four set size conditions. The discrimination threshold in a single item condition corresponds to σ_{Late}; those of four set size condition correspond to the σ_{ItemN}. Figure 2 shows examples of stimuli presentation in each condition and a schematic view of a trial sequence.

Method

Participants There were four observers, author MT and three experienced psychophysical observers, TT, YA, and SU. All had normal or corrected-to-normal vision.

Design There are five conditions: a single item and four item set sizes, 2, 4, 9, and 16. A set of items was presented in the first interval and a test item was presented in the second interval. One of the items was a target item, which needed to be compared with the test item. The QUEST procedure (Watson & Pelli, 1983) adaptively determined the JND at which the observer was 75% correct. Thus, the size of the target in each trial was calculated by QUEST.

Apparatus The stimuli were presented on the screen of a Mitsubishi 17 in. Monitor. The monitor was driven by a Mac Pro computer which also performed all timing functions and controlled the course of the experiment. Display resolution was 1024 × 756 pixels. Participants viewed the screen with both eyes and were seated approximately 115 cm from the monitor, fixed with in position with a chin rest.

Stimuli The items consisted of light gray dots on a dark gray background. The set of items, consisting of given number of dots was presented in the first temporal interval of two intervals trial. The test item, consisting of one dot, was presented in the second interval. In each trial, all of the dots shown were randomly scaled by a small multiplicative factor to discourage the participants from basing their judgments on previously seen stimuli. Three multiplicative factors (1, 1.1, 1.2) were used and the same factor scaled all items in any one trial.

The items were arranged on the array. The array was divided into m × m matrix: 2 × 2 (set size=2, 4), 3 × 3 (set size=9), and 4 × 4 (set size=16). Each item was displayed at the center of each cell with a position jitter. A lognormal Gaussian noise $lnN(0, \sigma_c^2)$ was added to the diameter of each item of a set independently. It has been assured that a lognormal distribution of circle diameters will produce a Gaussian distribution of discriminable sizes after logarithmic transduction (Solomon, Morgan & Chubb, 2011). Thus, in attempt to create normal distributions of transduced size, we use lognormal distributions of circle diameter. In this experiment, σ_c was set to 0.2.

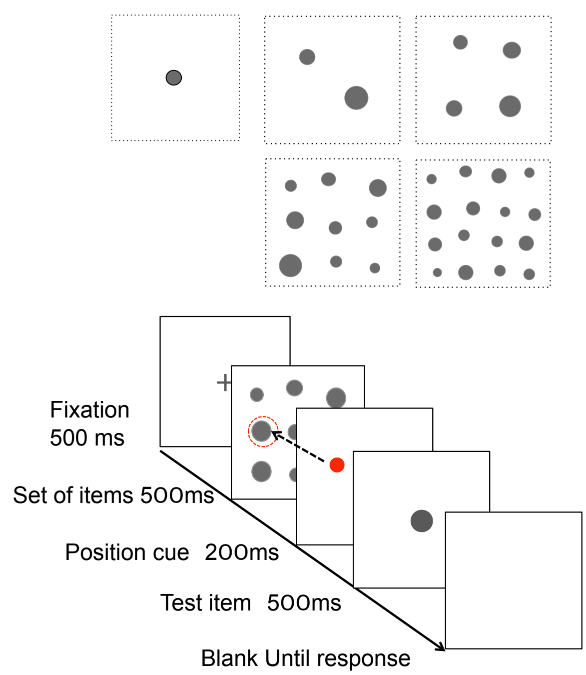

Figure 2: Example of item set of each set size (above) and schematic view of a trial sequence in Size Discrimination Experiment.

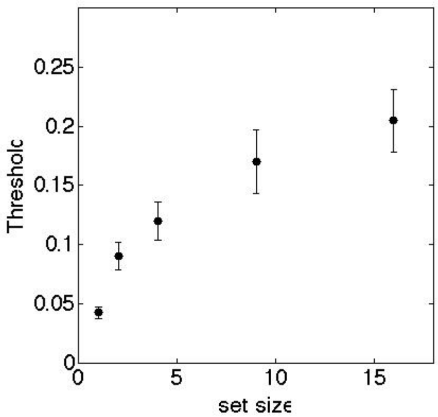

Figure 3: Means of size discrimination threshold as a function of the set size in Size Discrimination Experiment. Error bars represent standard deviations.

Procedure Each trial started with a fixation cross for 500ms. The items in a set were presented first for 500 ms. A red dot position cue, indicating the target, was presented for 200 ms after the item set. The test item was presented for 500ms after the ISI of 400 ms. The observers' task was to decide which item, the target item in a set or the test item, was larger in size. A 2AFC procedure was used. When they thought that the target item in a set was larger than the test item, they pressed '1'. When they thought that the test item was larger than the target item, they pressed '3'. No feedback about the correctness of responses was provided.

Results

Results are shown in Figure 3. The mean discrimination threshold for a single item condition was 0.043. This value was used for the parameter of the late noise (σ_{Late}). The mean discrimination threshold for target in each set size, 2, 4, 9, and 16 was 0.09, 0.12, 0.17 and 0.21, respectively. Each value was used for the parameters of the item noise (σ_{Item}) for each set size.

Size Averaging Experiment

In this experiment, we measured the discrimination performance between the estimated average size of a set and a test item. A schematic view of the stimulus presentation is shown in Figure 4. The observers, apparatus and stimuli were the same as in Size Discrimination Experiment except that no position cue was presented between the item set and following test item.

Method

Design There were two independent variables in the experiment that were varied within participants. The first variable was the number of items in a set; there are four set size, 2, 4, 9, and 16. The second variable was the level of difference between average size of a set and the test item. There were two levels: ± 0.08 (hard) and ± 0.12 (easy) relative to the average diameter. A set of items was presented in the first interval and a test item was presented in the second interval.

Each condition had 200 trials, resulting in 1600 trials in total. Each block had 160 trials (10 repetitions × 4 set size × 2 levels × 2 directions of test size (smaller or larger)) with 10 blocks in total. The participants performed five blocks in each experimental session, two sessions in total. The set size and the level of difference were blocked and the order of trials were randomly mixed. There were trials in the practice blocks.

Procedure Each trial started with a fixation cross for 500ms. The items in a set were presented first for 500 ms and the test item for 500ms after the intermission of 400 ms. The observers' task was to decide whether the test item was larger or smaller than the average size of item in a set. A two-alternative (larger or smaller) forced choice procedure was used. When they thought that the test item is smaller than the average size of items in a set, they pressed '1'. When they thought that the test item is larger than the mean size of a target set, they pressed '3'. No feedback about the correctness of responses was provided.

Results

The performance of each observer is shown in figure 5. The discriminability d'_H were calculated using the equation (1) and plotted as a function of the set size. As shown in the figure 5, the discriminability appeared to be unaffected by the number of items in a set, being consistent with the findings in previous studies. All observers showed higher discriminability when the difference between the average size of items and the test item is larger, which is ±0.12.

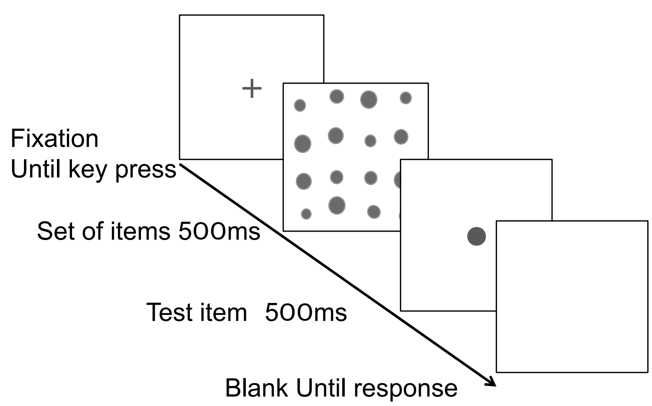

Figure 4: schematic view of a trial sequence in Size Averaging Experiment.

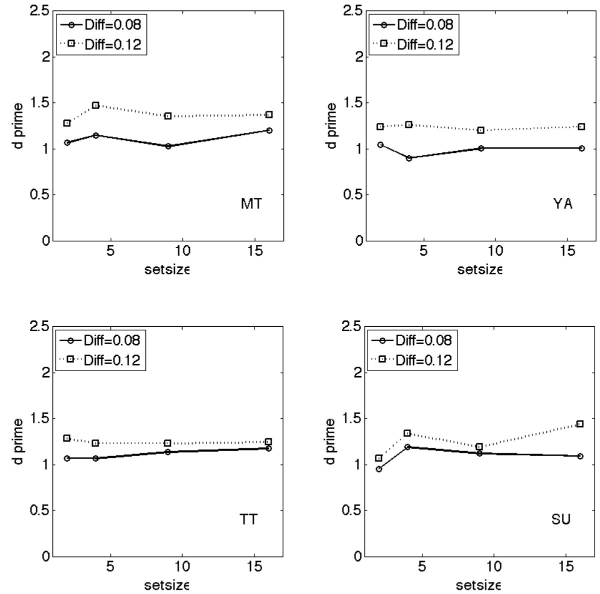

Figure 5: Discriminability of each observer as a function of the set size in Size Averaging task.

Statistical Efficiencies and evaluation of models

The ideal observer's discriminabilities of each model were shown in Figure 6 (top left). We calculated statistical efficiencies for the mean data of four observers using the equation (2). The efficiencies are presented in Figure 6 as a function of the set size. The efficiencies of Global Averaging Model without item noise (GAM1) were varied across set sizes. When the set sizes are 2 and 4, the efficiencies approximated 100%, whereas when they are 9 and 16, the efficiencies approximated 70%. The efficiencies of Global Averaging Model with item noise (GAM2) exceeded 100%, especially when the difference between set average and test size was smaller, indicating that the human observers did not adopt these strategies. The efficiencies of the LSM far exceeded 100%, indicating that the performance of human observers was higher than those of the models. This means that the human observers did not adopt this strategy.

Discussion

This study investigated sampling properties of visual information used by human observers in deriving an average representation of item set introducing three ideal observer models. We measured the performance of the Size Averaging Task for each ideal observer model and for human observers. Then, we compared the performance of the ideal observer of each model to the performance of human observers to evaluate which model could predict the human behavior.

As the statistical efficiencies of GAM2 and LSM far exceeded 100% and that the performance of human observers was higher than those of the models, we could assume that the human observers did not adopt these strategies. On the other hand, the efficiencies of GAM1 (Global Averaging model without item noise) approximated 100% when they are two and four and approximated they 70% when they are 9 and 16. Thus, we could assume that GAM1 may be appropriate model for the human observers in deriving the average of item sets. This suggests that the statistical summary representation might derive average size of items in a set without representing the size of individual items. The results were consistent with the claim that observers can estimate with high accuracy the average size of a set of items, even when they seem unable to report the size of individual items in the set (e.g, Airely, 2001). In other words, the results disagree with the claim that the average of the set could be accurately estimated by sampling as few as one or two items, and estimating the average of those items; observers are not strategically subsampling when they compute the mean size, especially in the case that the number of items is large such as 9 and 16.

There might be individual differences in the way observers access averages. Some observers are inclined to use a global process, whereas others a limited sampling process. A further study of how the sampling strategy is determined and whether the precision in the averaging could be improved should be conducted.

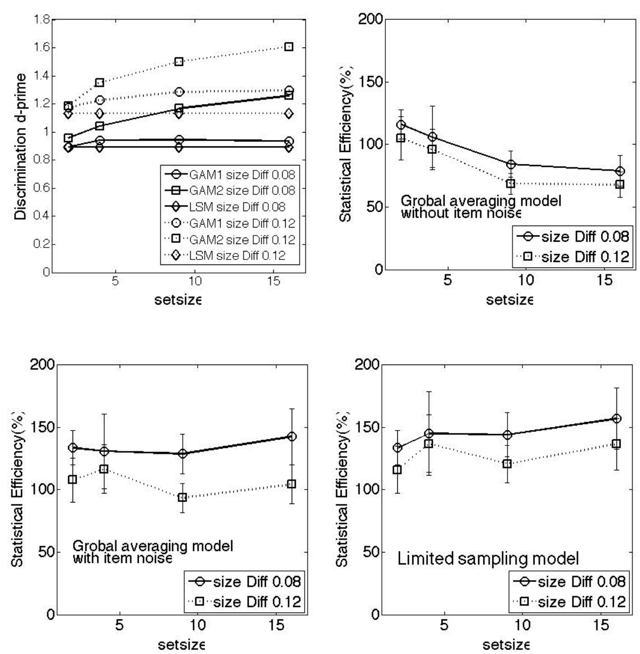

Figure 6: The ideal observer's discriminabilities (top left) and human efficiencies of each model as a function of the set size. The efficiency score is reported in a percentage.

Our results imply that the strategy depend on the number of items in a set, since the efficiencies of GAM1 fluctuate across set size. The implication is consistent with the finding of Tanaka & Ishiguchi (2006). In their study, the sampling strategy is varied across the number of items in a set; the efficiency decreased at the number of lines increase until a range of stable efficiency.

This study provides the evidence for the statistical summary representation mechanisms may not necessarily require the focused attention and the representation of individual objects. The processes of extracting statistical summary of number of items are relevant to the possible cognitive mechanisms of categorization, recognition, learning, and others. Further investigation is necessary to reveal the averaging process in more detail.

References

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

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychol Sci*, 19, 392-398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychol Sci,* 12, 157-162.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Percept Psychophys*, 70, 1325-1326; discussion 1335-1326.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. V*ision Research,* 18, 637-650.

Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19, 783-793.

Bauer, B. (2009). Does stevens's law for brightness extend to perceptual brightness averaging? *The Psychological Records, 59,* 171-186.

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

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

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 Psychol ogia,* 138, 289-301.

Franconeri, S. L., Alvarez, G. A., & Enns, J. T. (2007). How many locations can be selected once? *Journal of Experimental Psyhcology,: HPP*, 33, 1003-1012.

Geisler, W. S. (2003). "Ideal Observer Analysis". In Chalupa, Leo M.; Werner, John S. *The Visual Neurosciences.* MIT Press. pp. 825–837.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psyhcology,: HPP*, 35, 718-734.

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychon Bull Rev*, 18, 855-859. doi: 10.3758/s13423-011-0125-6

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Atten Percept Psychophys*, 75(2), 278-286.

Marchant, Alexander P., Simons, Daniel J., & de Fockert, Jan W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica,* 142, 245-250.

Ikeda, M., & Ishiguchi, A. (2004). Human efficiency for symmetry detection on 3D bumpy surface . *The Japanese Journal of Psycholog*y, 75 407-414.

Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Percept Psychophys,* 70, 772-788.

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research,* 79, 8-16.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. J*ournal of Experimental Psyhcology: Human Perception and Performance*, 16, 332-350.

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science,* 24(8), 1389-1397.

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

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *J Vis,* 11, 13.

Tanaka, M., & Ishiguchi, A. (2006). Discriminating the integrated orientation of line segments and its statistical efficiencies. *Japanese Psychological Research,* 48, 233-244.

Watamaniuk, S. N. (1993). Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *J Opt Soc Am A*, 10, 16-28.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113-120.

Yakushijin, R. (2007). Independence of sampling of motion parallax and binocular disparity in discrimination of depth, *Japanese Psychological Research*, 49, 248-260.