How do Distributions of Item Sizes Affect the Precision and Bias in Representing Summary Statistics?

Midori Tokita (tokita@mejiro.ac.jp)
Faculty of Health Sciences, Mejiro University,
Ukiya, Iwatsukiku, Saitama 339-8501, Japan
Akira Ishiguchi (ishiguchi.akira@Ocha.ac.jp)
Faculty of Core Research, Ochanomizu University,
Otsuka, Tokyo 112-8610 Japan

Abstract

Many studies have shown that observers can accurately perceive and evaluate the statistical summary of presented objects' attribute values, such as the average, without attending to each object. However, it remains controversial how the visual system integrates the attribute values (e.g., information on size) of multiple items and computes the average value. In this study, we tested how distributions of item sizes affect the precision and bias in judging average values. We predicted that if observers utilize all of the available size information equally, the distribution would have no effect, and vice versa. Our results showed that, with novice observers, judgement precision differed among size distributions and that the observers overestimated the size of the average value compared to the actual size under all conditions. These results imply that observations of some items in a set could be weighted more easily than others, with the possibility that this process is easier for larger items than smaller ones. However, this was not the case for experienced observers, who showed no effects of distribution type on average assessment performance. Our findings imply that the process of representing the average value may not be explained by a single definitive mechanism and, is rather mediated by a mixture of multiple cognitive processes.

Keywords: average size; statistical summary representation; size distribution

Introduction

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 (Dakin and Watt, 1997; Parkes, Liend, Angelucci, Solomon & Morgan, 2001), emotional expression (Haberman & Whitney, 2009, 2011). This ability is not limited to static and simultaneous events; it has been observed in sequentially presented events (Albrecht, Scholl, & Chun, 2012; Corbett & Oriet, 2011; Hubert-Wallander & Boynton, 2015) and dynamic objects, such as expanding and contracting disks (Albrecht & Scholl, 2010). Moreover, the ability to represent statistical properties is not limited to visual cues but is also observed in perceptions from auditory input, such as extracting frequency information from sequences of sounds (Piazza, Sweeny, Wessel, Silver, & Whitney, 2013) and temporal details of sounds (McDermatt, Schemisch, & Simoncelli, 2013).

These representations of statistical summary representations (SSRs), have been proposed to assist our judgment and behavior and more efficiently than attending to each objects and/or events individually (e.g., Alvarez, 2011; Ariely, 2001, 2008; Chong & Treisman, 2003; Robitaille & Harris, 2011). Although there is a general understanding that human observers can accurately represent sets of features, the mechanism by which people extract summary statistics is yet to be fully understood.

One of the debates over the mechanism of SSRs is whether the average value is computed using the entire information on display or using a subset of items in a set; ideas on this have been classified into three types based on findings from and discussions in previous studies.

The first idea is based on the claim that SSRs are computed without computing individual items. Many studies have provided the evidence that people can estimate the average size of a set of items without relying on focused attention on individual items when attention is distributed across a set of similar items (e.g., Attarha, Moore, & Vecera, 2014; Chong & Treisman, 2003, 2005; Oriet & Brand, 2013; Oriet & Hozempa, 2016; Tokita, Ueda, & Ishiguchi, 2016).

In the second idea, the assertion is that all items in a set are processed but not all items contribute equally to the mean. This idea suggests that if some measures are very reliable and others are not, observers may give the more reliable measures more weight when combining them (Alvarez, 2011). For example, Haberman and Whitney (2010) tested how the deviant emotional expression was utilized in averaging emotion shown on multiple faces and suggested that people implicitly and unintentionally discount the emotional outliers, thereby computing a summary representation that involves the majority of the information present. Hubert-Wallander and Boynton (2015) tested how SSRs were computed when stimuli were sequentially presented across time and found that they do not incorporate all items equally.

The basis of the third idea is that average size is computed using a limited sampling strategy which does not necessitate an ensemble representation computed in parallel across all items on display (Myczek & Simons, 2008; Fockert & Marchant, 2008; Marchant, Simons, & Fockert, 2013). For example, Fockert and Marchant (2008) claimed that observers do not always accurately average together the entire set and that the average is either biased by features of the attended item or based on the practical strategy of extracting the mean of a smaller subset. In line with this argument, Marchant, Simons, and Fockert (2013) tested how the regularity of the item sizes affects the performance of average size perception and demonstrated that judgments of average size become less accurate with increases in the set size and heterogeneity of the item sizes.

In this study, we further explored the mechanism of estimating the average size of items in a set by manipulating the frequency distribution of items. This is a useful approach for determining whether observers utilize information on all items equally, they weight the information of some items and less of others, or they use a limited number of items in a set (Chong & Treisman, 2003; Duffy, Huttenlocher, Hedges & Crawford, 2010; Marchant et al., 2013). For example, Duffy et al. (2010) used asymmetric (skewed) distributions in which there was more than one possible central value and demonstrated that observers adjusted estimates toward the category's running mean.

We tested how frequency distributions of item sizes affect the precision and bias in representing average size. Four types of the distributions were used: uniform, negatively skewed (i.e., Asym1), positively skewed (i.e., Asym2), and pseudo-normal distributions. Figure 1 shows the sizes and numbers of items in the stimulus set for each distribution. We measured the Weber fraction of the discrimination task to assess precision and also measured the point of subjective equality (PSE) to test the accuracy of the estimation.

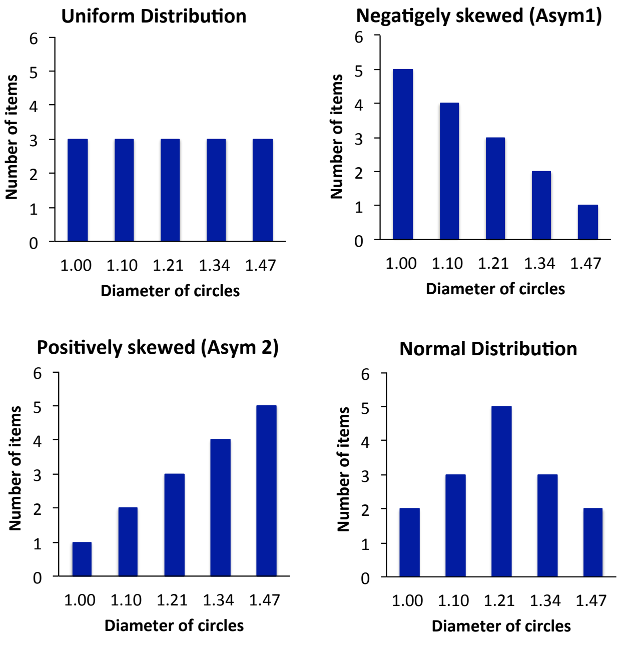

Figure 1: Four types of distributions were introduced: uniform, negatively skewed (i.e., Asym1), positively skewed (i.e., Asym2), and pseudo-normal distributions.

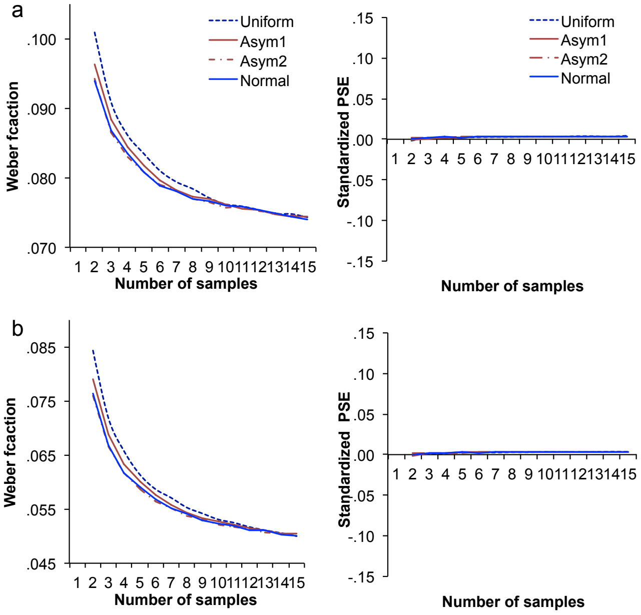

Figure 2: Simulated performances (Weber fraction and standardize PSE) for each distribution expressed as a function of the number of sampled items under different noise conditions: a. High internal noise, b. Low internal noise

To predict the performance of size average estimation, we used a computer simulation for calculating Weber fractions and the PSEs for each distribution expressed as a function of the number of sampled items, as shown in Figure 2. Two levels of noise were used in the simulation, with the values of the noise in each level being obtained from previous research (Tokita et al., 2016). As shown in the left-hand-side figures, when the average value was computed in parallel across all items in the display, performance was unaffected by the distribution of item size. When a limited number of items in the set were used to compute the average value, the performance was influenced by item distribution. Under Asym2 and normal distributions, precision was found to be better than that under uniform and Asym1 distributions. Precision under Asym1 was higher than under uniform. As shown in the right hand side of Figure 2, the PSE results show that observer accuracy was unaffected by distribution.

In addition, we conducted experiments with to two types of observer: novice and experienced. Since some studies have suggested that there are considerable individual differences in simultaneous visual tasks (Tokita & Ishiguchi, 2010; Herbert & Whitney, 2015), we considered it important to test how the effect of distribution differs between novice and experienced observers. Note that "the observers" refer to novice observers in this paper.

Experiment

We tested how the distributions of item sizes affect performance and bias in representing the average sizes. To examine the performance, we obtained Weber fractions that would indicate the precision in the participants' estimation of average sizes. To examine bias, we obtained PSEs, which indicate the constant error in estimation. In deriving the Weber fractions and PSEs, we used the method of constant stimuli, in which the observers in each trial decided which size of stimuli—the average size of a displayed item set (i.e., standard stimuli) or a single item size (i.e., comparison stimuli)—had larger.

Method

Participant As for novice observers, fourteen undergraduate volunteers from Mejiro University participated in exchange of course credit. All observers were naïve as to the purpose of the study. Three experienced observers including one of the authors, who have participated in many psychophysical experiments such as average estimation experiments were added. Two of the observers did not know the purpose of the study. All had normal or corrected-to-normal vision.

Apparatus Stimuli were displayed on the iMac desktop computer monitor (21-inch) controlled by a Macintosh computer (Mac OS X). Stimuli were generated using the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.4, Mathworks, MA). Participants viewed the screen with both eyes and seated approximately 65 cm from the screen.

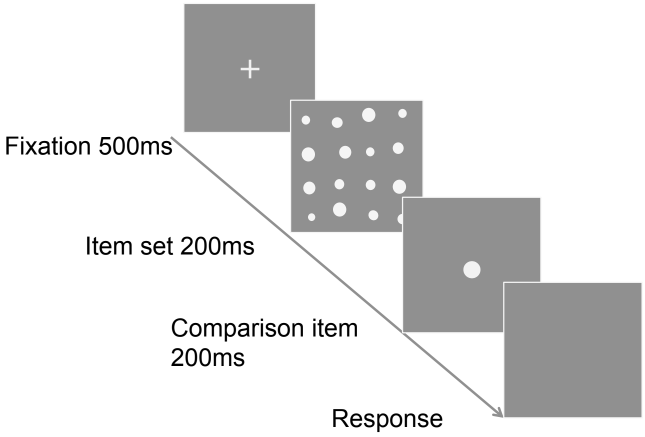

Figure 3: A schematic view of the stimulus presentation. Each trial started with a fixation cross for 500ms. The items in a set were presented first for 200 ms. The comparison item was presented for 200ms after blank for 500 ms, and then a blank screen until response.

Design Distributions of item sizes were manipulated. There are four distribution types; uniform distribution, negatively skewed distribution (Asym1), positively skewed distribution (Asym2), and pseudo-normal distribution. A set of items (i.e., standard stimuli) was presented in the first interval and a comparison item was presented in the second interval.

Stimuli The standard stimuli consisted of fifteen of filled light gray disks of various seizes, which were presented on dark gray background. The disk sizes were equally spaced on a log scale separated by a factor of 1.25. The comparison stimuli was a single disk with a given levels. There were five comparison levels, -0.14, -0.07, 0, 0.07, and 0.14 diameter differences on the power function scale. Three of participants needed wider range of stepwise levels (i.e., -0.16, -0.08, 0, 0.08, and 0.16) due to the low accuracy.

The items were arranged on the array. The array was divided into 4 × 4 matrix. Each item was displayed at the center of each cell with a position jitter.

In each trial, all of the disks shown were randomly scaled by a small multiplicative factor to discourage the participants from basing heir judgments on previously seen items. Three multiplicative factors (0.9, 1, 1.1) were used and the same factor scaled all items in any one trial.

Procedure A schematic view of the stimulus presentation is shown in Figure 3. Observers completed one 65-min session that consisted of a practice block of 24 trials, followed by five experiment blocks of 100 trials each (4 distribution types × 5 comparison level × 5 repetitions). There are 500 trials in total. The distribution types and the comparison level and the order of trials were all randomly mixed.

Each trial started with a fixation cross for 500ms. The items in a set were presented first for 200 ms. The comparison item was presented for 200ms after blank for 500 ms, and then a blank screen until response. The next trial automatically began 500ms after the response.

Observers' task was to decide whether the comparison 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 comparison item is smaller than the average size of items in a set, they pressed '1', otherwise, they pressed '3'. No feedback about the correctness of responses was provided.

Analysis The PSEs and Weber fractions were measured using the method of constant stimuli. First, the relative sizes for the comparison item were plotted on the x-axis, and the proportion of greater responses for each comparison stimulus was plotted on the y-axis, and fits were done for individual data. The plotted data points constructed the psychometric function approximated by a cumulative Gaussian function for individual data.

This discrimination threshold was defined as the smallest amount of the stimuli number change, for which a correct response rate of 75% was achieved. The PSEs were obtained as the values of the locations on the psychometric function at which the standard and comparative choice probabilities were equal to 50%.

Results

The fits of the data points to the psychometric functions were generally good, and the Pearson product–moment correlation coefficient exceeded .9 in most cases. The data of three novice observers were not satisfactory fitted to the psychometric functions, thus, we excluded their data form further analysis.

Figure 4a shows the results of Weber fractions in each distribution types. The precision of representing average size appeared to be partly affected by the size distribution. Figure 4b shows the standardized PSEs in each distribution condition. The average size of the set of items seems to be overestimated as compared with the comparison item in all distribution conditions with the novice observers.

Novice observers To test whether and how the size distributions affect the precision of representing average values, one way (4 distribution conditions) repeated measures analysis of variance was conducted on the individual Weber fraction. This yielded significant main effect of distribution, *F*(3,11)= 2.94, *p*<.05. Bonfferoni Post-hoc analysis revealed that the judgment in Asym2 condition (positively skewed distribution) gave higher precision than uniform, *p*< .05, and Asym1 (negatively skewed distribution) *p*< .05 conditions.

In a similar way, to test how the distribution affected the accuracy of representing average sizes, one way (4 distribution conditions) repeated measures analysis of variance was conducted on the individual PSE. This yields no main effect of distribution, *F*(3,11)= .554, *p*>.1. This suggests that the accuracy of representing average size was not affected by the size distribution. As average of PSE seems larger than the 0, we conducted a one-sample t test to compare the mean standardized PSEs of each condition with a PSE of 0. The analysis revealed that the mean of the PSE was significantly larger than 0 at the standard stimuli in the uniform, *t*(10) = 4.99, *p* < .01, Asym1, *t*(10) = 2.63, *p* < .05, Asym2, *t*(10) = 4.40, *p*< .01, and normal distribution, *t*(10) = 3.30, *p* < .01. This suggests that average size were overestimated as compared with the actual size in all distribution condition.

Experienced observers To test whether and how the size distributions affect the precision of representing average values, one way (4 distribution conditions) repeated measures analysis of variance was conducted on the individual Weber fraction of the experienced observers. This analysis revealed no significant effect of distribution, *F*(3, 2)= .523, *p* >.1.

To test how the distribution affected the accuracy of representing average sizes, one way (4 distribution conditions) repeated measures analysis of variance was conducted on the individual PSE. This yields no main effect of distribution, *F*(3, 2)= 1.342, *p* >.1. This suggests that the accuracy of representing average size was not affected by the size distribution. This suggests that in experienced observers there is no sign of bias average size were overestimated as compared with the actual size in all distribution condition.

Discussion

We tested how the distribution of item size would affect the precision and bias in the judgement of average size of items in a set by observers. Four types of distributions, namely uniform, negatively skewed, positively skewed, and pseudo-normal, were examined. We predicted that, if all items in a set were processed equally, the performance of the observers in judging the average would be unaffected by item distribution conditions. Our results demonstrated three significant findings. First, precision was significantly higher under Asym2 (i.e., positively skewed distribution) compared to the others, among which there were no significant differences.

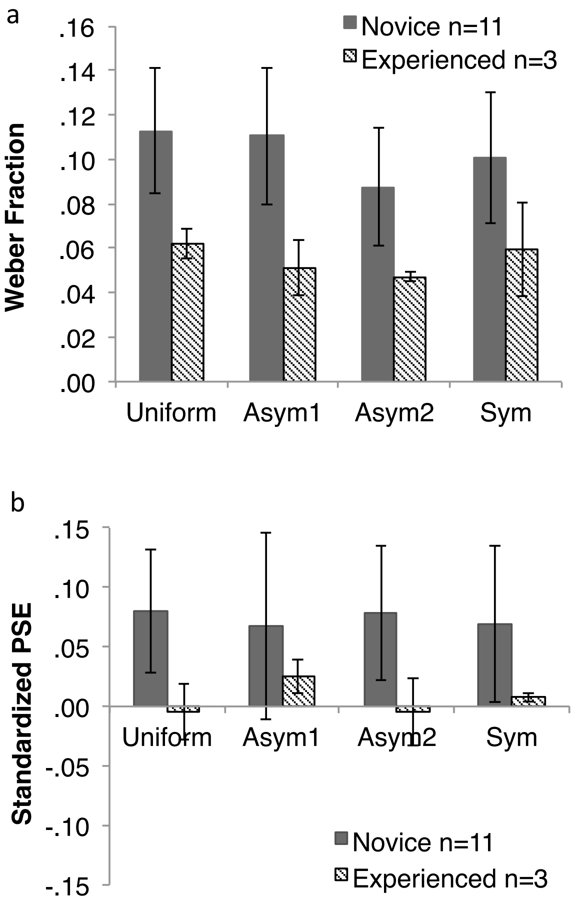

Figure 4: a Mean Weber fraction for each distribution condition. Error bars represent standard deviations. b. Mean PSEs for each distribution condition. Error bars represent standard deviation.

Second, the estimated average values of the items in a set were overestimated relative to the actual values under all item distributions. Third, the accuracy and precision of the experienced observers was unaffected by distribution type when carrying out the averaging task.

The first set of results implies that observers may not use all of the available information equally when assessing the average value of the items in a set, which could be because of two alternative processes. One possibility is that a limited number of elements could be being used to calculate average values. Another possibility is that all of the information on size in a display may not be weighted equally when judging averages. Instead, some of the items could be being weighted more than others. The results partially support those of Marchant et al. (2013), who indicated that the difference in the component of the item sizes affects the performance of perceiving the average.

The finding that overestimation of the average size relative to the actual size under all distributions is intriguing. Bias in the judgment of the average has rarely been investigated, and no consistent pattern of bias has yet been observed. There are two possible causes for the bias. One is the observer's perceptual saliency of large stimuli; the observer may process the large items more attentively than the small ones, and thus, it is possible that observers utilize the information of larger items relatively more frequently in computing the average size. Another possibility is that observers may automatically give more weight to the larger items than to the smaller ones, irrespective of their saliency.

The fact that experienced observers were unaffected by distribution type suggests that they may process all the items in a set equally when elucidating their average value. These results support the idea that people can extract the average size of a set of items without relying on focused attention to individual items in the set. Taken together with the results for novice observers, it is implied that how we represent the average value may not be accounted for just from a single process but from a variety of processes that depend on individual cognitive characteristics. The implication is somewhat consistent with the findings of Haberman, Brady, and Alvarez (2015). They pointed out that the mechanism of representation of summary statistics may involve various levels of processes, and individual differences may reveal those levels. As our data shows, with a wide variety of item distributions, it is important to explore the basis of those differences.

In conclusion, our results demonstrate that average judgement precision differs with distribution and the size of the average value was overestimated by the observers under all conditions. However, the results for experienced observers showed that the performance of their judgement of the average was unaffected by distribution type. Our findings imply that the process of assessing the average value may not be explained by a single definitive mechanism but is rather mediated by a mixture of multiple cognitive processes.

Acknowledgments

This research was partially supported by a Grant-in-Aid for Scientific Research (C) (40571112) and Scientific Research(B) (15H03462) provided by the Japan Society for the Promotion of Science to the authors.

References

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

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: computing summary statistics from multimodal stimuli. *Atten Percept Psychophys, 74*, 810-815. doi: 10.3758/s13414-012-0293-0

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhance visual cognition. *Trends in Cognitive Science, 15,* 122-131.

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. *Psychological Science, 12*, 157-162.

Attarha, M., Moore, C. M., & Vecera, S. P. (2014). Summary Statistics of Size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensample. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1440-1449.

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 Psychologia, 138*, 289-301.

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

Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Percept Psychophys, 70*, 789-794.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718-734.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Atten Percept Psychophys, 72*, 1825-1838. doi: 10.3758/APP.72.7.1825

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.

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in Ensemble perception Reveal Multiple, Independent Levels of Ensemble Representation, *Journal of Experimental Psychology: General*, *144*, 432-446.

Hubert-Wallander, B., Boynton, G. M. (2015). Not all summary statistics are equal: Evidence from extracting summaries across time, *Journal of Vision*, 15:5, doi: 10.1167/15.4.5

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.

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nat Neurosci, 16*, 493-498. doi: 10.1038/nn.3347nn.3347 [pii]

Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to summary statistics 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.

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

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*, 1389-1397.

Robitaille, N., & Harris, I. M. (2011). When more is less: extraction of summary statistics benefits from larger sets. *Journal of Vision*, 11(12). doi: 10.1167/11.12.18 S0001-6918(11)00142-9 [pii]

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. *Journal of Vision*, 11, 13.

Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Atten Percept Psychophys, 72*, 1839-1853.doi:10.3758/app.72.7.1839

Tokita, M., Ueda, S., & Ishiguchi, A. (2016). Evidence for a global sampling process in extraction of summary statistics of item sizes in a set. *Frontiers in Psychology, 7*. doi: 10.3389/fpsyg.2016.00711