

When more is less: Extraction of summary statistics benefits from larger sets

Nicolas Robitaille

Irina M. Harris

School of Psychology, University of Sydney, Australia, &
International Laboratory for Brain, Music and Sound
Research, Montreal, Quebec, Canada



School of Psychology, University of Sydney, Australia



Despite several processing limitations that have been identified in the visual system, research shows that statistical information about a set of objects could be perceived as accurately as the information about a single object. It has been suggested that extraction of summary statistics represents a different mode of visual processing, which employs a parallel mechanism free of capacity limitations. Here, we demonstrate, using reaction time measures, that increasing the number of stimuli in the set results in faster reaction times and better accuracy for estimating the mean tendency of a set. These results provide clear evidence that extraction of summary statistics relies on a distributed attention mode that operates across the whole display at once and that this process benefits from larger samples across which the summary statistics are calculated.

Keywords: visual cognition, object recognition, detection/discrimination

Citation: Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, 11(12):18, 1–8, <http://www.journalofvision.org/content/11/12/18>, doi:10.1167/11.12.18.

Introduction

Research in visual perception and attention has revealed a limiting bottleneck in our ability to encode more than about four objects for detailed analysis (Cowan, 2001; Luck & Vogel, 1997), and if additional temporal selection pressures are present, this capacity can be as little as one object (see Marois & Ivanoff, 2005, for a review). At the same time, however, other studies have shown that human observers can extract the essence, or gist, of a scene in as little as 100 ms (Potter, 1976; Thorpe, Fize, & Marlot, 1996). This is hard to explain on the basis of serial processing of individual objects in the scene and has prompted some researchers to suggest that we also have an alternative mode of processing visual information, based on the extraction of statistical regularities present in a scene.

The extraction of summary statistics, such as the mean or distribution of a set of similar objects, appears to be a general mechanism that operates on different stimulus attributes, including orientation, size, spatial position of moving objects, size of expanding and contracting objects, and even facial expression and identity (Albrecht & Scholl, 2010; Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2003; Dakin, 1999, 2001; Dakin & Watt, 1997; de Fockert, & Wolfenstein, 2009; Haberman & Whitney, 2007, 2009a; Parkes, Lund, Angelucci, & Solomon, 2001). In a typical experiment (e.g., Ariely, 2001), observers are shown a display containing a group of circles of different sizes and have to decide whether a single circle presented subsequently (the probe) is smaller

or larger than the mean size of the set. Ariely (2001) found that observers could estimate the mean size of the set very precisely and that their accuracy was not impaired by increasing the number of items in the display. In contrast, their ability to discriminate among set members was at chance, indicating that they encoded very little information about the individual items comprising the set.

Chong and Treisman (2005) suggested that extraction of statistical properties occurs automatically and in parallel. Their reasons for this assertion were that participants' accuracy did not suffer when the number of items was increased or when the displays were very brief (50 or 100 ms; Ariely, 2001; Chong & Treisman, 2005; Haberman & Whitney, 2009a). Furthermore, the fact that observers do not seem able to accurately identify individual members of the set suggests that they do not focus attention on the stimuli in a serial fashion. Thus, Chong and Treisman (2005) proposed that, in addition to the focused mode of attentional deployment used when identifying individual items, there is also a distributed mode of attention that operates on the whole display at once, which underpins statistical processing of sets of similar items and extracts the gist of the scene. However, based on the evidence cited above alone, it is difficult to refute serial processing of stimuli with focused attention during statistical processing, for two reasons. First, evaluating similar stimuli on a single dimension (e.g., size) to report a single value may not require full identification of the stimuli and, thus, may take less time per item, allowing sufficient time to process them serially despite brief stimulus exposure. Second, it is possible that observers base their decisions

on a randomly selected subset of the display (one or two items), which may allow slow serial processing of these few items (Myczek & Simons, 2008).

Myczek and Simons (2008) used simulation data to demonstrate that performance levels on the mean estimation tasks reported in earlier experimental studies could be achieved with a strategy of focusing attention on just a small sample of items in the set and argued against the need to postulate a specialized parallel mechanism for extracting summary statistics. This sampling hypothesis received some support from a study by de Fockert and Marchand (2008), who showed that it is possible to bias the estimated means by directing attention toward the larger or smaller items in the set. However, other researchers have challenged the sampling explanation (Ariely, 2008; Chong, Joo, Emmanouil, & Treisman, 2008). Ariely (2008) argued that the simulations provided by Myczek and Simons did not comfortably capture the whole range of experimental findings without considerable adjustment to the parameters of their model. He also pointed out that estimating the mean on the basis of one or two individuals in the set is inconsistent with the fact that observers seem to have very little information about the individual set members (Ariely, 2001; Haberman & Whitney, 2007, 2009). Chong et al. (2008) provided experimental evidence against some of the sampling strategies proposed by Myczek and Simons, although they acknowledge that it is difficult to devise tests that conclusively demonstrate sampling of more than about 4–5 items, because larger samples quickly approximate the population mean and observer accuracy saturates. Thus, there is still a need to demonstrate performance that *exceeds* what could be accomplished by a sampling strategy, in order to conclusively refute this explanation (Simons & Myczek, 2008).

In the present study, we sought to address the question of whether estimating the summary statistics of object sets is performed with focused attention that is deployed serially from item to item, or in a more global manner, with attention distributed across the whole display at once, by using reaction time measures and varying the set size. This approach is commonly employed in the visual search literature to infer differences between so-called “serial search” and “parallel search” (Treisman & Gelade, 1980). The hallmark of serial search is a systematic and steep increase in reaction times with increasing numbers of stimuli, and it is typically found when searching for feature conjunctions (Treisman & Gelade, 1980). Parallel search, on the other hand, is typically inferred when there is little reaction time cost regardless of the number of stimuli in the display, as occurs with singleton targets that “pop-out” from the distractors. Although a popular distinction, Wolfe (1998) has suggested caution in ascribing these two patterns of performance to serial vs. parallel mechanisms, respectively, and recommended using the terms “efficient” and “inefficient” search instead. As already mentioned above, it is possible for a very efficient

serial mechanism to return performance that appears largely insensitive to the number of stimuli in the display. Likewise, a parallel process that has capacity limitations and has to be applied several times in order to process the whole display could result in systematic increases in reaction time as the set size increases (e.g., Townsend, 1990). However, it should be noted that the notion of “serial” and “parallel” employed in visual search is somewhat different to the constructs we are discussing here. In visual search, the observer is actively looking for a specific target, while at the same time trying to filter out the distractors present in the display. The number and nature of the distractors influences how efficiently the target is “separated” from background information (Duncan & Humphreys, 1989). In contrast, when estimating the statistical properties of a set of items, there are no targets or distractors—all items are relevant to the task. The question of interest is the manner in which these relevant stimuli are processed when the task is to estimate their statistical properties as a whole: is this done through focusing attention successively on individual items (i.e., serially), or is it done in a distributed fashion (i.e., in parallel) across the whole display at once?

We reasoned that if all items in the set are processed with focused attention, then we would expect to see a systematic and steep increase in reaction times with increasing numbers of stimuli, as is usually seen in visual search for feature conjunctions. Alternatively, if attention is only focused on a small subset of the stimuli, this would lead to equivalent reaction times regardless of the number of stimuli in the set, because the observer’s decision would always be based on the same (small) number of stimuli. In distinct contrast to these predictions, we show here that, when estimating the mean size or orientation of a set of items, reaction time is actually *reduced* when the number of stimuli increases, and this reduction in reaction time is accompanied by improved (or unchanged) accuracy. This result cannot be accommodated by a focused attention mechanism deployed to individual items, with or without a sampling strategy. At the same time, we show that searching for a specific member of a set presented under identical conditions results in a steep increase in reaction time, as expected in cases of inefficient visual search (Wolfe, 1998). Taken together, these findings provide direct evidence for two different mechanisms: one involved in identifying specific items in a visual display, which operates serially, and a different one involved in extracting summary statistics from a set of items, which appears to operate automatically and in parallel across the entire visual display.

Experiment 1

Experiment 1 used a similar paradigm as Ariely (2001), in which observers estimate the average size of the circles

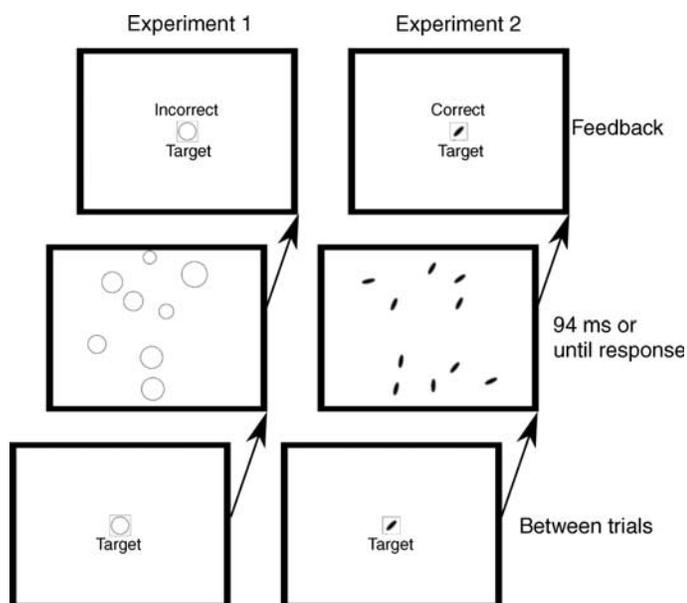


Figure 1. Stimuli and trial sequences used in the experiments (stimuli drawn to scale, text enlarged).

and varied set size. We employed a two-alternative response, which allowed us to measure both accuracy and reaction times to the displays. Observers decided whether the mean size of the set was “larger” or “smaller” than a target circle that appeared before each trial (the same target was used in every trial). Two exposure durations were used (94 ms or until response) to ensure that our findings are not due to the time pressure imposed by brief stimulus exposure.

Methods

Participants

Twenty undergraduate students participated either for course credit or in exchange for payment. All had normal or corrected-to-normal vision.

Stimuli and procedure

Stimuli consisted of black circle outlines (2-pixel width) presented on a white background (see Figure 1). With this

type of stimulus, it is not always clear whether subjects use the diameter or the surface area, when estimating the size of a circle. This is an important consideration because a distribution of circles that are equally spaced in surface area will not be equidistant in diameter. Specifically, since the area is related to the diameter by a power of 2 ($\text{area} = \pi(\text{diameter}/2)^2$), a power factor of 0.5 would have to be applied to an equidistant distribution of diameter values in order to get an equidistant distribution of areas. Chong and Treisman (2003) attempted to determine whether subjects based their estimates on the diameter or the area and concluded that the metric that seemed to be used was somewhere in between what would be predicted by averaging the diameters or the areas. It was unclear if this was due to a mixture of subjects that used one or the other strategy or if each subject used a mix of both metrics. Chong and Treisman therefore adjusted the size of their circle stimuli by a power of 0.76. This specific exponent value accords well with Teghtsoonian’s (1965) finding that the apparent size of circles grows somewhat more slowly than the actual area of the stimuli, in a manner that was best fit by an exponential factor of 0.76. Therefore, in the present experiment, we followed Chong and Treisman and adjusted the diameters of our distribution of circles by a factor of 0.76.

Forty-nine circle sizes were used in the experiment. The circle at step 25, which represents the middle of the range, was selected as the target to be used on all trials. Thus, there were 24 circles with a smaller diameter than the target and 24 circles with a larger diameter, and their diameters increased in equal steps. A power function of 0.76 was applied to the stimuli before drawing them on the screen. The formula used to calculate the actual size in pixels for each diameter step through the range was $\text{diameter in pixels} = ((\text{step} * 10) + 110)^{0.76}$. These translate to a diameter range of 38 to 129 pixels or 1.27° to 4.36° on the screen. The addition of 110 to the size, before the power function was applied, ensured that even the smallest circles were visible on the screen. Step 25 was used as the target, yielding a target size of 2.93° of visual angle (Figure 2).

The displays consisted of sets of 2, 4, 6, 8, 10, or 12 circles. For each trial, a set of stimuli was selected to be

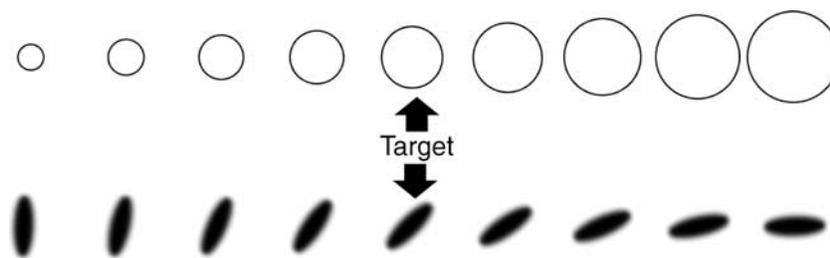


Figure 2. A range of the stimuli used in Experiment 1 (circles) and Experiment 2 (bars). The circles depicted here were rendered perceptually equidistant from their neighbors by applying a power function of 0.76 to their absolute diameters (Teghtsoonian, 1965; see text). Each bar is equally distant in orientation from its neighbors. Forty-nine possible stimuli were used in each experiment, including the targets, evenly distributed between the extrema (illustrated here).

displayed on the screen as follows. First, the average size of the sample was determined according to a staircase procedure (see next paragraph). Random samples were then generated from all possible stimuli, except the target, and without repeating any of the same stimuli in a set, until a set with the appropriate mean was found. All the calculations of the average size of a display and their distance from the target were based on the arithmetic mean of the diameters as defined by their step identity (i.e., the untransformed size). Once this was done, the transformation was applied to that step value to obtain the actual size in pixels for drawing the stimulus on the screen. The screen position of each circle was chosen at random, with the constraints that each stimulus would be entirely present on the display, would not overlap with another circle on the screen, and would not be located at the center of the screen (where the target circle was displayed between the trials).

The average size of the displayed sample was always smaller or larger than the target, which meant that there was always an unambiguously correct answer on each trial on which to base participant feedback. The distance between the average of the set of stimuli and the target circle was set initially to 32% of the range (i.e., a smaller-than-the-target set had a mean of step 9, translating to 1.87° , and a larger-than-the-target set had a mean of step 41, translating to 3.86°). This distance was rapidly decreased to 20% of the range during the practice, ensuring that the answer to the initial trials was quite obvious. A staircase procedure modified this distance (± 1 step) after each block of 12 trials during the task to keep accuracy within the range of 80–85%. The distance used was always the same for each set size, so any differences in the accuracy across set size cannot be explained by the use of the staircase.

Each trial started when the participant pressed and released the two response keys simultaneously. The target was presented first and was followed by a blank screen for 200 ms and then by the stimulus set. The set remained on the screen for 94 ms or until response (in separate blocks of trials). The participant used two keys to respond “larger” vs. “smaller” than the target (counterbalanced across participants). Feedback (“Correct” vs. “Incorrect”) was given on every trial and the target was presented again in preparation for the next trial. Each participant performed two blocks of trials in counterbalanced order: brief (94 ms) and long (until response) display durations. Separate staircases were used for each block. Each block consisted of 264 trials (22 repetitions of the 12 combination of stimuli and response), preceded by 44 practice trials.

The experiment was run using Matlab and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and was displayed using a Dell 19-in. CRT monitor refreshing at 85 Hz.

Results and discussion

Reaction time (RT) analyses were performed on correct trials only. This is a conservative approach that ensures

observers were performing the task adequately on those trials. However, the same pattern of results was obtained when all trials (including the incorrect ones) were included in the analysis. Outliers were removed prior to analysis using a “modified recursive” method as follows. For each condition and subject, the average RT and the standard deviation were calculated. The observation furthest from the mean was then temporarily excluded and the standard deviation and mean were recalculated. If the new mean was lower than the mean minus $C \times$ the standard deviation or greater than the mean plus $C \times$ the standard deviation, the outlier value was excluded permanently, and the process was repeated until no observation was rejected. The criterion, C , was 3.0 or greater and was adjusted as a function of sample size, as explained in Van Selst and Jolicoeur (1994). The retained RTs were analyzed with a 2 (exposure duration: 94 ms or until response) \times 6 (set size: 2, 4, 6, 8, 10, 12) repeated-measures ANOVA. RT was significantly different across set size, $F(5,95) = 14.31$, $MSE = 4388$, $p < 0.0001$ (see Figure 3). There was also a main effect of exposure duration, $F(1,19) = 55.46$, $MSE = 4388$, $p < 0.0001$, with RTs being on average 142 ms longer in the long than in the short exposure condition. However, the modulation by set size was not different for the short and the long exposure, $F < 1$. The slope of the regression line relating RT and set size (collapsed across brief and long exposure) indicated a reduction of 9.5 ms per item, which was significantly different from zero, $t(19) = -4.6$, $p < 0.0002$.

Accuracy was also modulated by set size, $F(5,95) = 13.65$, $MSE = 0.004$, $p < 0.0001$. This modulation was not different across brief and long exposure, $F(5,95) = 2.01$, $MSE = 0.0023$, $p > 0.08$, although there was overall better accuracy (by 1.5%) for long exposure compared to short exposure, $F(1,19) = 10.7$, $MSE = 0.0014$, $p < 0.004$. The

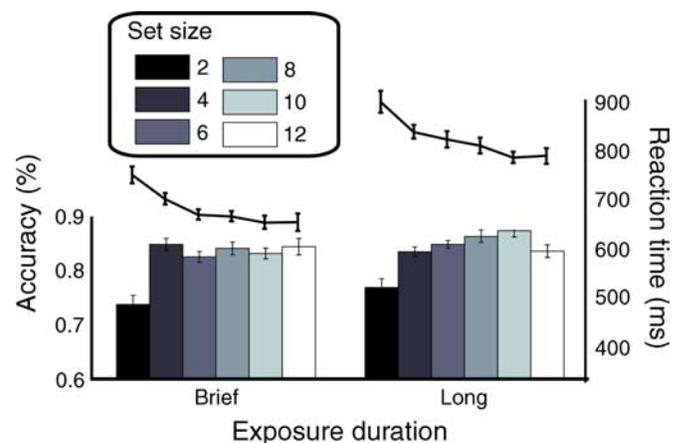


Figure 3. Accuracy (bars) and reaction time (solid line) for Experiment 1, plotted as a function of set size and exposure duration. Error bars represent standard error of the mean, with the intersubject variability removed.

slightly lower accuracy for the brief exposure resulted in the staircase adjusting the average set size to be slightly more distant from the target (i.e., easier discrimination) in this condition compared to the long exposure condition, $F(1,19) = 4.75$, $MSE = 1.03158$, $p < 0.042$. The slope of the regression line relating accuracy and set size (collapsed across brief and long exposure) showed an increase in accuracy of 0.6% per item, which was significantly different from zero, $t(19) = 4.3$, $p < 0.0004$.

These results are inconsistent both with a focused attention mechanism deployed sequentially across all the individual items, which would be expected to produce longer RTs as set size increases, and with a sampling of a small subset of stimuli, which presumably would result in constant RTs regardless of set size. Thus, the pattern of results obtained here seems to be best accommodated by a distributed attention mechanism that operates in a global manner across the whole set at once, although it is not entirely clear why this would lead to a reduction in RT for larger sets. We return to this point in the [General discussion](#) section, after presenting the results of [Experiment 2](#).

Experiment 2

In order to definitively conclude that statistical representation is a different mode of perception to the focused attention typically employed for processing individual objects (and not just a result of using stimuli that generally engender parallel processing), it is important to show that the stimuli used in calculating the mean statistic are processed in a serial fashion if the task required it. Therefore, in [Experiment 2](#) we used both a mean estimation task and a visual search for a specific target item present in the stimulus displays. To further generalize the results of [Experiment 1](#), here we used oriented bars instead of circles and required observers to estimate the mean orientation of the bars in the display.

Participants saw sets of 2, 4, 6, 8, or 10 tilted bars and performed two tasks: (1) an estimation of the mean orientation of the set and (2) a visual search of a prespecified target. Both tasks employed a two-alternative response that allowed us to measure both accuracy and RT to the displays. For the summary statistics task, the two alternative responses were “more horizontal” and “more vertical” than a target bar that appeared before each trial. For the visual search task, the two responses were “present” and “absent”, referring to a target similarly presented before each trial. In both cases, the target was a perfect diagonal (45° orientation). Thus, the trial structure and response requirements were kept constant across the two tasks, allowing us to compare the pattern of RT obtained in the summary statistics task to that obtained in the visual search task, which is known to entrain serial processing of stimuli.

Methods

Participants

Twenty-four undergraduate psychology students participated in the experiment for partial course credit. All had normal or corrected-to-normal vision. Small groups of up to six participants were run at a time.

Stimuli and procedure

The apparatus was the same as in [Experiment 1](#). The stimuli were elongated oval shapes (3.23° height and 1.46° width), blurred to avoid aliasing effects (we refer to them as “bars” for simplicity), and were presented in a variety of orientations in random locations within a central 32° × 32° square area (see [Figure 1](#)). Bar orientations varied from 0° (vertical) to 90° (horizontal), in 49 equidistant steps, with all stimuli in between pointing to somewhere in the upper right quadrant (see [Figure 2](#)).

For the summary statistics task, the mean orientation of the stimulus set had a maximum distance from the target of 10.8° (so 34.2° for a “more vertical” set and 55.8° for a “more horizontal” set). The distance between the mean and the 45° target was reevaluated after 10 trials by a staircase procedure that increased the distance by 1.8° when the accuracy was inferior to 80% (to a maximum of 10.8°) and reduced it by 1.8° when the accuracy was superior to 90%. For the visual search task, the difficulty was similarly titrated by varying the minimal angular distance between the target and any of the other stimuli. This distance was adjusted throughout the experiment, as for the summary statistics task but with an initial value of 18° and a maximum of 27°.

Participants performed the summary statistics and visual search tasks in counterbalanced order. Each task comprised 240 trials (5 set sizes × 2 responses × 24 repetitions), preceded by 20 practice trials. Each trial started when the participant pressed and released the two response keys simultaneously. The target (a bar oriented at 45°) was presented first and was followed by a blank screen for 200 ms and then by the stimulus set. The set remained on the screen for 94 ms for the summary statistics task and until response for the search task. The participant used two keys to make a response (“more vertical” vs. “more horizontal” than the target for the summary statistics task; “target present” vs. “target absent” for the visual search task). The key binding was counterbalanced across participants. Once the response was given, a feedback screen informed the participants of the correctness of their response and the target was presented again in preparation for the next trial.

Results and discussion

Accuracy and RT for correct responses on the summary statistics and visual search tasks are presented in [Figure 4](#),

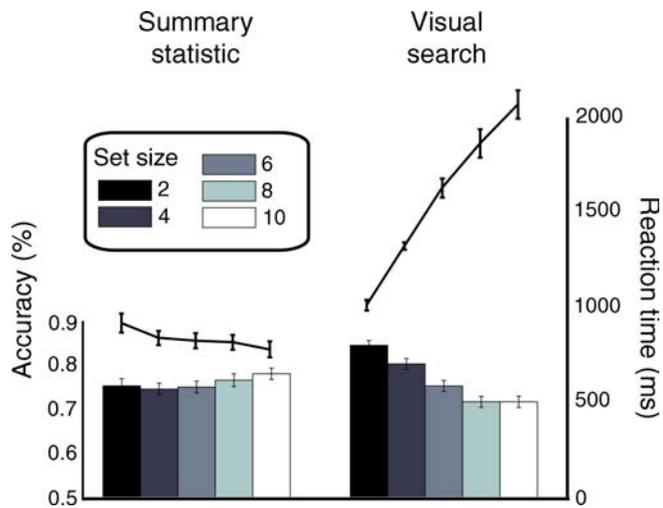


Figure 4. Accuracy (bars) and reaction time (solid line) for Experiment 2, plotted as a function of set size and task. Error bars represent standard error of the mean, with the intersubject variability removed.

plotted as a function of set size. RT outliers were removed using the same method as in Experiment 1. The data for each task were analyzed with separate ANOVAs using set size as a within-subject factor. Regression analyses were also performed to evaluate the slope of the regression line relating set size and level of performance (RT or accuracy).

Summary statistics task

Set size had no effect on accuracy, $F(4,92) = 0.4$, $MSE = 0.0359$, $p > 0.8$, but modulated RT in this task, $F(4,92) = 8.5$, $MSE = 7295$, $p < 0.0001$. As shown in Figure 4a, RT generally decreased as set size increased. The regression analysis relating the number of stimuli and RT revealed an average slope of -14.9 ms/item, which was significantly different from zero, $t(23) = -3.96$, $p < 0.001$, replicating the findings of Experiment 1.

Visual search task

Set size had a significant effect on RT, $F(4,92) = 97.4$, $MSE = 43,441.24$, $p < 0.0001$, and lead to a search slope of 132 ms/item, which is significantly different from zero, $t(23) = 10.91$, $p < 0.0001$. Accuracy was also modulated by set size, $F(4,92) = 29.3$, $MSE = 0.00252$, $p < 0.0001$. The regression analysis relating the number of stimuli and the accuracy level revealed an average slope of -1.7% /item; this was significantly different from a slope of zero, $t(23) = -10.84$, $p < 0.0001$. As expected, this task showed the steep search slope that is typically associated with serial processing of the stimuli (Treisman & Gelade, 1980). The long time required to perform the search task is in accordance with the difficulty observers experience when identifying individual members of a set (Ariely,

2001; Haberman & Whitney, 2007, 2009a) and stands in marked contrast to the rapid estimation of the mean of the set. The results of this experiment clearly show that stimuli that require slow serial search for individual identification are processed rapidly and with the hallmarks of a parallel mechanism when forming a summary statistical representation.

General discussion

The results of these experiments are clear. Across different stimulus dimensions (size, orientation), increasing the size of a stimulus set resulted in a decrease in the time required to estimate the mean size or orientation of the set. The decrease in RT was accompanied by an increase in accuracy for larger set sizes, confirming that extraction of summary statistics generally improves for larger sets. The pattern of performance on these mean estimation tasks is very different to the steep increase in RT with increasing set size that is the hallmark of serial processing in typical visual search tasks and that was demonstrated for processing individual items in Experiment 2.

These findings are clearly inconsistent with the involvement of a serial, focused attention process in the representation of summary statistics. Furthermore, the negative RT slope identified in our data also contradicts models that postulate that observers base their estimations on a subset of the visual information (Myczek & Simons, 2008). Such models would predict no modulation of RT across set size, as the task would always be executed based on only one or two stimuli. Instead, our results show that adding more stimuli does have an impact on the observer's performance, indicating that these additional stimuli are not simply ignored. As discussed in the Introduction section, it is generally quite difficult to devise tasks that conclusively demonstrate sampling of more than 4–5 items when using accuracy measures, because larger samples quickly approximate the population mean and accuracy saturates. Through the use of reaction times, this study provides conclusive evidence that performance can exceed what would be accomplished by a sampling strategy. We believe that these results could only be accommodated by a mechanism whose operation is automatically distributed across the whole visual display.

A somewhat unexpected result of this study was that reaction times sped up as the number of stimuli in the set increased. In general, a parallel distributed process is predicted to be insensitive to the number of stimuli present (although see Townsend, 1990). Instead, the reduction in reaction times seen here suggests that the estimation of summary statistics becomes easier the more stimuli there are. A number of previous studies investigating ensemble coding have reported similar trends in their data, whereby performance (measured in terms of

discrimination thresholds or error on a mean adjustment task) tends to improve with larger sets (e.g., Haberman & Whitney, 2009a, 2009b). In these previous studies, there was only a trend in this direction, but this may be due to the fact that the accuracy measures used were less sensitive than our reaction time measure. There are several possible reasons for this improvement in mean estimation for larger sets. One possibility is that this finding represents a form of redundancy gain, which is typically reflected in a reduction in reaction time when two identical signals are presented to the subject (e.g., Miller, Beutinger, & Ulrich, 2009). This could particularly be the case in Haberman and Whitney's studies, because their displays are usually made up of replications of a basic set of 4 items (i.e., a set size of 16 is made up of 4 copies of 4 distinct stimuli), though it is perhaps a less likely explanation for our results, given that all members of our sets were unique. Another possibility is that, as set size increases, the mean of the set more closely approximates the population mean. As a result, it may become easier to discount outlier values, which could speed up the subject's decision. This notion would be consistent with recent findings that human observers are quite good at discounting outliers, particularly with larger sets of stimuli, achieving better performance than a model that perfectly averages all the samples present (Haberman & Whitney, 2010). These possibilities could be tested in future studies by systematically manipulating the properties of the distributions used.

In conclusion, this study provides clear evidence that the estimation of summary statistics relies on a processing mode that is distributed across the entire visual display and that benefits from larger samples across which the summary statistics are calculated. The new methodology we introduce here, the use of reaction time in a mean estimation task with a predetermined target, is a promising way to study the dynamics of statistical processing in more detail. Further studies exploring the limits of this statistical processing mode of vision will help bridge the gap between the limited capacity of our central cognitive systems and the seemingly highly adequate and effort-free processing of the vast amount of visual information we are exposed to in our everyday life.

Acknowledgments

This work was supported by an ARC Research Grant. We would like to thank Justin Harris for comments on the manuscript.

Commercial relationships: none.

Corresponding author: Nicolas Robitaille.

Email: nicolas.robitaille@umontreal.ca.

Address: International Laboratory for Brain, Music and Sound Research, Montreal, Quebec H3C 3J7, Canada.

References

- Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*, 560–567. [PubMed]
- Alvarez, G., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392–398. [PubMed]
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162. [PubMed]
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics, 70*, 1325–1326. [PubMed]
- Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436. [PubMed]
- Chong, S., Joo, S., Emmanouil, T., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics, 70*, 1327–1334. [PubMed]
- Chong, S., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404. [PubMed]
- Chong, S., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*, 891–900. [PubMed]
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*, 87–185. [PubMed]
- Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vision, 12*, 1–30. [PubMed]
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A, 18*, 1016–1026. [PubMed]
- Dakin, S. C., & Watt, R. (1997). The computation of orientation statistics from visual texture. *Vision Research, 37*, 3181–3192. [PubMed]
- de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70*, 789–794. [PubMed]
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology, 62*, 1716–1722. [PubMed]
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433–458. [PubMed]

- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, 751–753. [PubMed]
- Haberman, J., & Whitney, D. (2009a). Averaging facial expression over time. *Journal of Vision*, *9*(11):1, 1–13, <http://www.journalofvision.org/content/9/11/1>, doi:10.1167/9.11.1. [PubMed] [Article]
- Haberman, J., & Whitney, D. (2009b). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718–734. [PubMed]
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception & Psychophysics*, *72*, 1825–1838. [PubMed]
- Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. [PubMed]
- Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*, 415–415. [PubMed]
- Miller, J., Beutinger, D., Ulrich, R. (2009). Visuospatial attention and redundancy gain. *Psychological Research*, *73*, 254–262. [PubMed]
- Myczek, K., & Simons, D. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*, 772–788. [PubMed]
- Parkes, L., Lund, J., Angelucci, A., & Solomon, J. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744. [PubMed]
- Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [PubMed]
- Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Cognition*, *2*, 509–522. [PubMed]
- Simons, D., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, *70*, 1335–1336. [Article]
- Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, *78*, 392–402. [PubMed]
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522. [PubMed]
- Townsend, J. T. (1990). Serial and parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46–54. [Article]
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. [PubMed]
- Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology: Section A*, *47*, 631–650. [Article]
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Hove, UK: Psychology Press. [Chapter]