Special Thematic Section on "Tracking the Continuous Dynamics of Numerical Processing"

Dissociating Parallel and Serial Processing of Numerical Value

Kassandra R. Lee*[a], Kenith V. Sobel[b], A. Kane York[b], Amrita M. Puri[c]

[a] Schepens Eye Research Institute, Massachusetts Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA. [b] Department of Psychology and Counseling, University of Central Arkansas, Conway, AR, USA. [c] Department of Biology, University of Central Arkansas, Conway, AR, USA.

Abstract

Digits serve as useful tools for studying the interaction between low-level perceptual representations and higher-level semantic information, and also the degree to which processing these stimulus attributes relies on similar or different mechanisms. Following a body of literature that debates the influence of high-level, semantic information on perceptual processing, and work by Van Opstal and colleagues (2011) investigating whether subliminally presented digit arrays affect estimates of numerical averages, here we explored the temporal dynamics of extracting numerical values associated with each digit in an array. Specifically, we examined reaction times (RTs) for estimating the average of digit arrays of varying sizes to determine whether numerical meaning is extracted in parallel, or instead may require serial processing of individual digits. In Experiment 1, participants completed a numerical mean estimation task, along with visual search tasks designed to yield RT patterns across increasing display sizes thought to be characteristic of serial and parallel processes. In Experiment 2, we controlled for brightness cues that could have contributed to performance in Experiment 1. In both experiments, comparing RT patterns for the numerical averaging tasks to those of the search tasks suggested that semantic information from multiple digits may be extracted by a parallel processing mechanism. Unlike in either search task, RTs improved with increasing display size, indicating a potential benefit of larger displays as has been reported for extracting ensemble, or summary statistical representations of lower-level visual information.

Keywords: parallel processing, semantic information, average estimation, ensemble coding, visual search

Letters and digits are symbols with specific meanings; however, they also possess characteristic and incidental low-level perceptual attributes such as shape and size. Here, we use digits to explore the degree to which processing of higher-level semantic information relies on similar or different mechanisms compared to perception of low-level sensory characteristics of environmental stimuli.

The question of how semantic meaning interacts with lower-level stimulus features to influence perception has engendered considerable controversy in recent years. Some evidence points to the ability of semantic information to affect judgments about perceptual attributes of a stimulus (e.g., Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Lupyan, Thompson-Schill, & Swingley, 2010; Puri & Wojciulik, 2008). Furthermore, al-

though Wolfe and Horowitz (2004) proposed that visual search is only influenced by perceptual information, more recent work has shown effects of semantic information on finding particular letters or digits among others (e.g., Lupyan, 2008; Sobel, Puri, & Hogan, 2015), even if the semantic influence is not as strong as guidance by perceptual characteristics such as shape (Godwin, Hout, & Menneer, 2014). Others, however, have argued that visual perception is impenetrable to higher-level cognition (e.g., Firestone & Scholl, 2016), especially in its early stages, deemed "early vision" (Marr, 1982; Pylyshyn, 1999). Recently, the size congruity effect (SCE), which describes the interference between physical and numerical size observed in digit comparison (Henik & Tzelgov, 1982) and digit search tasks (Krause, Bekkering, Pratt, & Lindemann, 2017; Sobel & Puri, 2018; Sobel, Puri, & Faulkenberry, 2016), has led to disagreement about how interactions between semantic and physical attributes occur. One possibility is that they could be combined into a single representation at an early, perceptual stage (Schwarz & Heinze, 1998; Walsh, 2003). Alternatively, the representations could remain separate throughout perceptual processing and only interfere at the decision stage (Faulkenberry, Cruise, Lavro, & Shaki, 2016; Santens & Verguts, 2011). A recent study showed that although the SCE occurs reliably in a variety of contexts and demonstrates that low-level perceptual and higher-level semantic information do interact, it is likely that physical and numerical size interfere at a decision rather than perceptual processing stage (Sobel, Puri, Faulkenberry, & Dague, 2017).

In view of these debates, we wanted to explore the fundamental nature of semantic representations and investigate whether or not processing of arbitrary meanings such as numerical value can occur via similar mechanisms as extraction of lower-level visual information. One way to approach this issue is to ask participants to complete a numerical averaging task, which would require extraction of higher-level, semantic attributes of multiple digits. Van Opstal, de Lange, and Dehaene (2011) addressed whether the semantic meanings of digit stimuli can be incorporated into summary statistical representations, and if so, whether this process relies on mechanisms that support such "ensemble coding" in the purely visual domain. In this context, ensemble coding refers to the rapid extraction of statistical summary information from groups of similar stimuli (e.g., estimating the average size of a set of circles, or in the real world, glancing at a pile of apples and quickly realizing they are, on average, red), which results in an efficient and useful representation of our visually dynamic world (Ariely, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2007; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001).

Ensemble perception has been shown to occur for low-level stimulus features such as size, line orientation, spatial location, and direction of motion (Alvarez & Oliva, 2008; Alvarez & Oliva, 2009; Dakin & Watt, 1997; Oriet & Corbett, 2008; Parkes et al., 2001; Watamaniuk, Sekuler, & Williams, 1989; Williams & Sekuler, 1984), as well as for more complex attributes such as emotional expression of faces, identity, and gaze direction (Haberman & Whitney, 2007; Haberman & Whitney 2009; Haberman & Whitney 2010; Sweeny & Whitney, 2014). Ensemble representations are thought to depend on extraction of information from multiple items in a cluttered scene simultaneously, or in parallel, rather than via a serial process requiring attention to individual items (Alvarez, 2011; Alvarez & Oliva, 2009). Furthermore, accurate estimation of the ensemble does not appear to rely on processing of only a small subset of items as suggested by Myczek and Simons (2008), but instead depends on extraction of at least some information from the entire set (Haberman & Whitney, 2010). Thus, it appears that summary information about perceptual attributes can be gathered from a visual scene implicitly, rapidly, efficiently, and at different levels of processing complexity.

Van Opstal et al. (2011) specifically asked whether the semantic information associated with a "prime" display (a digit array presented briefly prior to presentation of a target digit array) can affect participants' perception of the target digit arrays even when they are not consciously aware of the primes. They reported that subliminally presented digit arrays influenced estimates of the numerical average of target displays, and concluded that participants accurately performed ensemble coding of higher-level, semantic attributes (numerical meaning) in parallel, and without explicit awareness of individual items in the display. These findings are consistent with previous work investigating statistical representation of low-level stimulus attributes (Alvarez & Oliva, 2008; Ariely, 2001; Chong & Treisman, 2005). However, Van Opstal et al. (2011) also observed increases in reaction times (RTs) with larger display sizes (number of items presented in the display) that may nonetheless reflect a serial contribution to estimates of numerical averages. Furthermore, their displays were limited to relatively small numbers of items (3, 4, and 5), and thus participants may have been encouraged to attempt exact averaging despite relatively brief displays (< 800 ms). On the other hand, extracting information from a set of items may proceed in parallel until the number of items exceeds the subitizing range (i.e., four or fewer) and enters the counting range (Railo, Karhu, Mast, Pesonen, & Koivisto, 2016).

The ability to gather semantic information simultaneously from multiple digits would be indicative of parallel processing (Egeth, 1966) rather than serial processing. Serial and parallel cognitive mechanisms have traditionally been dissociated using a variety of approaches, including examining the temporal dynamics of searching for a target item among distractor items. It is generally observed that when a target item possesses a unique feature compared to distractor items and thus "pops out", RTs increase relatively little with additional distractors. This indicates that information about that feature was processed across all items in the display in parallel. In contrast, when targets contain features that are similar to those of surrounding distractors, search is less efficient and RTs increase with the number of distractors, reflecting the possibility that individual items must be selectively attended in a serial fashion until the target is found (Duncan & Humphreys, 1989; Treisman & Gelade, 1980).

These highly replicated differences in RT patterns across different types of visual search have traditionally been viewed as reflecting a distinction between parallel and serial mechanisms, although whether less efficient search is due to a serial, attentionally demanding process as opposed to factors such as increased noise (and thus reduced discriminability of the target) is a subject of ongoing investigation (e.g., McElree & Carrasco, 1999; Moran, Zehetleitner, Liesefeld, Müller, & Usher, 2016; Townsend, 1990; Verghese, 2001). Nonetheless, for the purpose of examining the effects of additional digits on estimating numerical averages, these well-established RT patterns may provide a useful basis for comparison. Comparing performance across search and ensemble tasks may help us determine whether or not extracting semantic information from multiple digits in an array engages a relatively efficient mechanism akin to that proposed to underlie the relatively flat RT slopes in parallel, or "pop-out" visual search.

In the current study, we used displays that contained varying numbers of digits, and greater numbers of digits than previously tested (Van Opstal et al., 2011), and compared RT patterns for estimating numerical averages with those obtained during visual search for digits. In order to assess the effect of additional items on estimating the mean of digit arrays, we designed the search tasks according to well-established criteria known to yield performance consistent with traditional "serial" (increasing RTs with larger displays) or "parallel" (flat RT slopes) search tasks, and used identical digit arrays across the search and ensemble tasks. Critically, the instructions differed across the search and ensemble tasks. In both search tasks, we instructed participants to search for a

number less-than- or greater-than-five, whereas in the ensemble, or averaging task, the instructions were to estimate whether the average numerical value of the digits in the display was less than or greater than five. An important difference between the two search tasks was that in the serial search, all of the digits were the same color, whereas in the parallel search, the digit that participants were searching for was always red and thus stood out compared to the rest of the display. Our goal was to directly compare RTs for estimation of the numerical average of digit displays with increasing numbers of items to those for a serial search task and a parallel search task (in which the target item pops out due to a unique low-level feature). By doing so, we aimed to determine whether the ability to rapidly extract a set statistic at the semantic level appears to proceed in a serial fashion, such that RTs increase with additional items, or whether instead, estimating a numerical average relies on processing of semantic information from multiple items in parallel.

We predicted that asking participants to estimate the average value of a set of digits would yield one of three types of RT patterns in relation to the search tasks (Figure 1). Increased RTs with increased display size, as in a serial search task, would suggest that generating the estimate is inefficient and may rely on serial processing of individual items. Alternatively, if RTs for the ensemble task show minimal or no increase with increasing display size, as characteristic of traditional parallel search tasks, this would be evidence that estimating numerical averages occurs by extracting semantic meaning from items in parallel. A third possibility is that if generating ensemble representations of numerical value relies on mechanisms similar to that proposed for ensemble perception of lower-level, perceptual stimulus attributes, RTs may actually decrease with increasing display size. This is in accordance with findings that additional items in a sequential digit display contribute to faster RTs and a more accurate ensemble representation (Brezis, Bronfman, & Usher, 2015), and other studies suggesting that additional stimuli provide more information about the ensemble average (Piazza, Sweeny, Wessel, Silver, & Whitney, 2013; Sweeny & Whitney, 2014). We did not have a strong prediction related to overall RTs for the ensemble compared to the search tasks, but for the purpose of illustration in Figure 1, we have indicated them to be similar to the serial search task but slower than the parallel search task, as pop-out searches tend to be highly efficient.
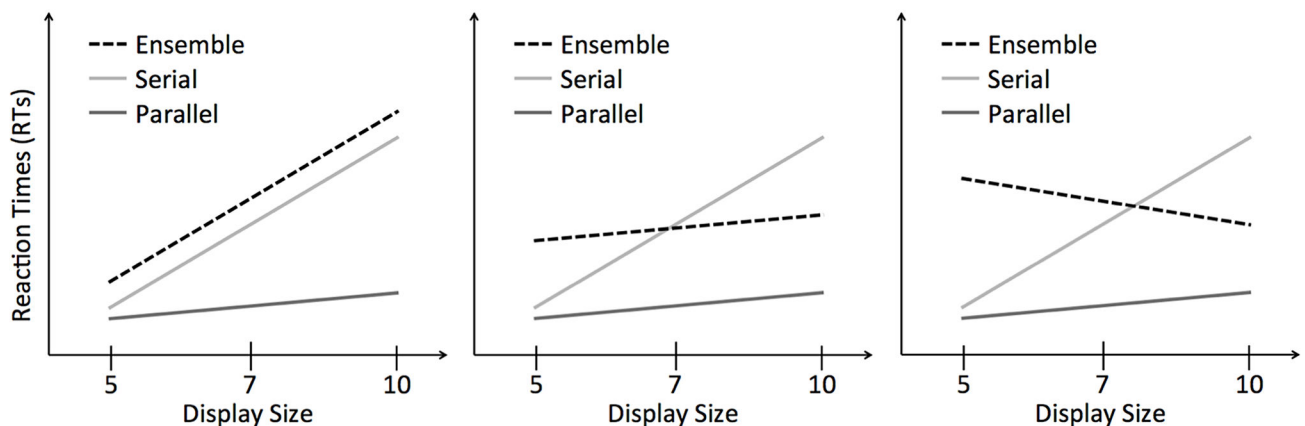

Figure 1. Predictions for Experiment 1.

Note. For serial search (light gray), RTs are expected to increase with increasing display size, and for parallel search (dark gray), RTs typically remain flat or increase only slightly with increasing display size (all panels). The ensemble task (dashed black lines) may give rise to a RT pattern similar to that of either the serial search (left panel) or parallel search (middle panel) tasks, depending on whether numerical meaning is extracted from multiple items serially or in parallel. Alternatively, faster RTs with increasing display size (right panel) would suggest a contribution of ensemble coding mechanisms.

Experiment 1: Parallel vs. Serial Processing of Digit Displays

Method

Participants

For the current study, we predicted a medium effect size (f = .5) based on Cohen's guidelines (Cohen, 1988). A preliminary power analysis was conducted using G*Power 3.1 software (Faul, Erdfelder, Lang, & Buchner, 2007) with a medium effect size of f = .5, an alpha level of .05, and a .95 confidence level. This analysis indicated that for 80% power, a minimum sample size of 10 participants is required for each set of tasks in this study.

Nineteen undergraduate students from a large public university in the Midwestern U.S. volunteered to participate for class credit. The university's Institutional Review Board approved all experimental procedures. Data from three participants were excluded. Two had RTs that were greater than 2.5 standard deviations above the mean across participants (one for the ensemble task and one for the search tasks), and another had low accuracy (< 60%) in the search tasks. Thus, a total of three datasets were excluded across the three tasks to allow for within-subjects comparisons across tasks. The remaining 16 participants included 13 females and three males with normal or corrected-to-normal vision (M_{age} = 20.06, SD = 0.77).

Stimuli and Procedure

All task programs were written in Xojo Basic and conducted on a Mac computer connected to a built-in iMac display with a screen resolution of 1920 x 1080 pixels. Stimulus arrays were presented on the monitor and responses were gathered from the keyboard. Task (serial search, parallel search, ensemble) order was counterbalanced across participants. Stimuli for all tasks consisted of a fixation point surrounded by circular arrays of five, seven, or ten digits ranging from 1-9, but not including 5, presented in white text against a black background (Figure 2). At a viewing distance of about 60 cm, each digit was 0.92° degrees of visual angle wide by 1.8° tall. Digits were arranged on an imaginary circle with a radius of 5.9° centered on a fixation cross spanning 1.0° on each side. In the search tasks, the target digit was located in one of four quadrant locations: upper right, lower right, lower left, or upper left, and was always at least 30° of arc away from vertical to avoid ambiguity with respect to the vertical meridian.
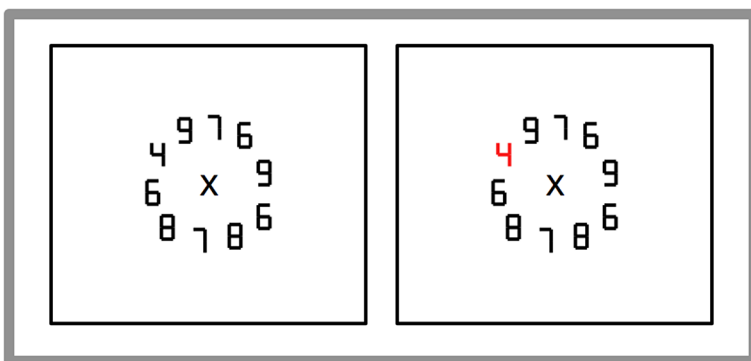

Figure 2. Example digit arrays for serial (left) and parallel search task (right), with the black background and white text color switched for aesthetic purposes.

Note. Displays consisted of 5, 7, or 10 digits; participants used the keyboard to indicate on which side of the display the less-than-five or greater-than-five digit appeared. In the serial search task all digits were the same color, whereas in the parallel search task targets were presented in red such that they popped out.

PsychOpen GOLD

In different blocks, participants were instructed to search for the number greater than or less than five, and to report on which side of the display the number was located by pressing the 'z' key to indicate the left side of the display and the '/' key for the right side. Block order within each task was counterbalanced. Trials began with a fixation cross presented for 500 ms, followed by the search display, which remained visible until the participant responded. Participants received feedback only when they responded incorrectly, in the form of a white screen with the word "Incorrect" in the middle presented for 750 ms followed by a fixation cross to begin the next trial. In the serial search task, the target digit was presented in the same white text as the distractors. The parallel search task was designed such that the target would pop out and thus yield RT patterns characteristic of parallel processing mechanisms. Displays and instructions for the parallel search task were identical to those for the serial search task, but the target digit was presented in red. Each task consisted of 12 practice trials and 120 experimental trials, with display sizes randomly interleaved.

The ensemble task consisted of a total of 288 trials that were identical to those in the serial search task, except that participants were instructed to estimate whether the average of the digits in each display was less than or greater than five (Figure 3). Participants responded using the 'z' key to indicate "less-than-five" and the '/' key for "greater-than-five." Averages of the digits within each display were always above or below 5 (never exactly 5); these averages ranged from ±1.2 to ±2.4 relative to 5. Displays across the search and ensemble tasks contained identical digits arrays. Each array contained only one digit that was on the "other side" of five from the average (e.g., the "3" in the display pictured on the left in Figure 3), which served as the target in the search tasks. This constraint, along with ensuring minimal digit repetitions within a display, yielded average differences from 5 for the means of the 5, 7, and 10 item displays that varied within a small range (±1.5, ±1.7, and ±2, respectively).
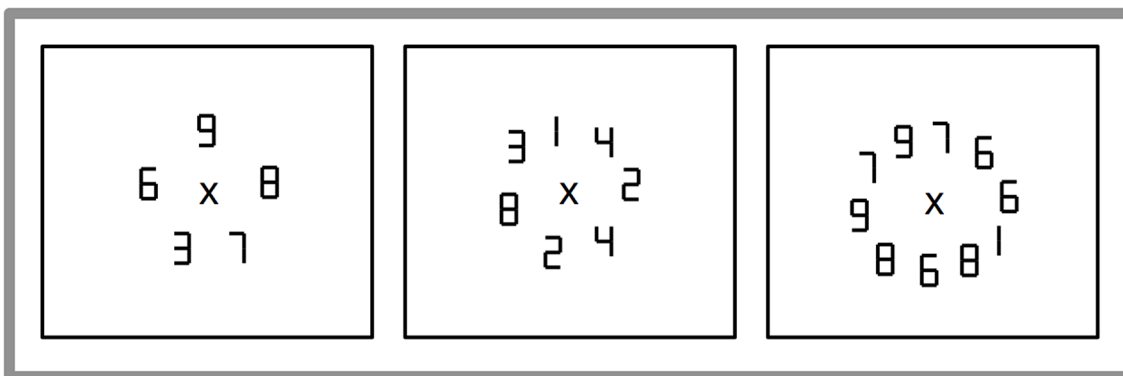

Figure 3. Example digit arrays for ensemble estimation task.

Note. Displays consisted of 5, 7, or 10 digits. Participants were asked to estimate the average numerical value of the digits and use the keyboard to indicate whether the average was less than or greater than five.

Results

For each participant, we excluded all trials with RTs that were greater than the mean plus three standard deviations for that participant, or less than 100 ms; a total of 1.9%, 1.7%, and 1.5% of data points were removed for the ensemble, serial and parallel search tasks, respectively. A 3 (*Task*: Serial Search/Parallel Search/Ensemble) x 2 (*Target Type*: < 5/> 5) x 3 (*Display Size*: 5/7/10 items) repeated measures ANOVA was conducted on RTs. There was a significant main effect of task, $F(2, 30) = 98.33$, $p < .001$, such that RTs for the ensemble task

were longer ($M = 816.81$, $SD = 171.46$) than for the serial ($M = 716.04$, $SD = 117.33$), and parallel ($M = 400.29$, $SD = 33.97$) search tasks, as shown in Figure 4. A significant main effect of target type, $F(1, 15) = 18.55$, $p = .001$ reflected faster RTs in the greater-than-five compared to the less-than-five condition for the ensemble (> 5: $M = 776.82$, $SD = 147.92$; < 5: $M = 856.79$, $SD = 195.01$) and serial tasks (> 5: $M = 686.95$, $SD = 103.77$; < 5: $M = 745.13$, $SD = 130.89$). There was also a main effect of display size, $F(2, 30) = 11.92$, $p < .001$, such that RTs increased with larger displays (5 items: $M = 627.90$, $SD = 105.21$; 7 items: $M = 641.12$, $SD = 102.91$; 10 items: $M = 664.12$, $SD = 114.65$). These main effects were qualified by a significant three-way interaction, $F(4, 60) = 15.02$, $p < .001$. We further examined this interaction by conducting a 2 (Target Type) x 3 (Display Size) repeated measures ANOVA for each task type followed by simple effects analyses and pairwise comparisons as appropriate.
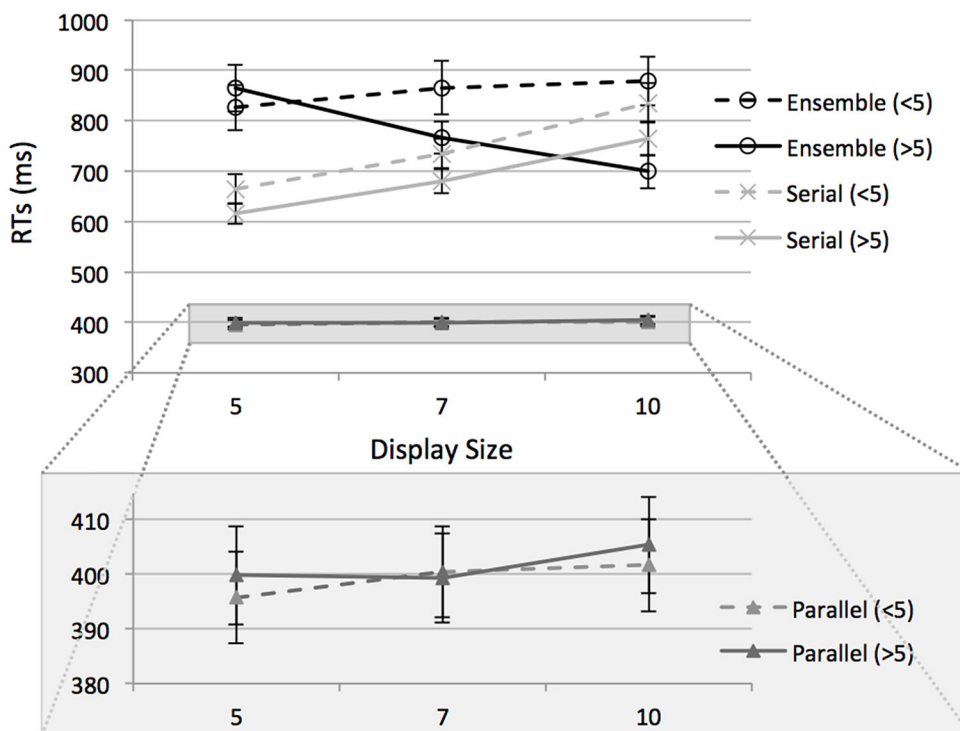

Figure 4. RTs for the serial search (light gray), parallel search (dark gray), and ensemble task (black) in Experiment 1 plotted as a function of display size and separated by target type.

Note. Dashed lines indicate < 5 and solid lines indicate > 5. Error bars represent standard error of the mean. RTs increased substantially with display size in the serial search task, but only slightly for parallel search (magnified in subfigure). RTs for the ensemble task decreased with larger displays in the greater-than-five condition (black solid line).

For the serial search task, a significant main effect of display size, $F(2, 30) = 60.52$, $p < .001$ reflected that as display size increased, RTs also increased (5 items: $M = 640.37$, $SD = 98.07$; 7 items: $M = 707.51$, $SD = 106.89$; 10 items: $M = 800.25$, $SD = 147.04$). A significant main effect of target type, $F(1, 15) = 8.65$, $p = .01$ was due to faster RTs for greater-than-five targets ($M = 686.95$, $SD = 103.77$) compared to less-than-five targets ($M = 745.13$, $SD = 130.89$). There was no interaction between these variables.

A 2 (*Task*: Serial Search/Parallel Search) x 3 (*Display Size*: 5/7/10 items) ANOVA compared the increases in RTs across the two search tasks. There was a main effect of task, $F(1, 31) = 266.96$, $p < .001$, such that RTs for

the serial search (*M* = 716.04, *SD* = 119.97) were slower than for the parallel search (*M* = 400.29, *SD* = 33.47). There was also a main effect of display size, $F(2, 62) = 100.47$, $p < .001$, due to an overall increase in RTs with increasing display size. A significant interaction between task and display size, $F(2, 62) = 81.30$, $p < .001$ was followed up with pairwise comparisons to determine the relative effect of display size on RTs across the different search tasks.

For the serial search, there were substantial and significant differences in RTs between all display sizes (5 items: *M* = 640.37, *SD* = 98.07; 7 items: *M* = 707.51, *SD* = 106.89; 10 items: *M* = 800.25, *SD* = 147.04, all *ps* < .001), whereas the parallel search yielded an increase of less than 6 ms overall, with a significant difference only between the 5- (*M* = 397.68, *SD* = 34.07) and 10-item (*M* = 403.39, *SD* = 33.92) displays, $t(31) = -2.79$, $p = .009$.

For the ensemble task, there was a main effect of target type, $F(1, 15) = 12.57$, $p = .003$, due to faster RTs for the greater-than-five condition (*M* = 776.82, *SD* = 147.92) compared to the less-than-five condition (*M* = 856.79, *SD* = 195.01), just as observed for serial search (Figure 5). A significant main effect of display size, $F(2, 30) = 5.52$, $p = .009$, reflected a pattern different than that in either of the search tasks such that larger display sizes yielded faster RTs (5 items: *M* = 845.65, *SD* = 183.01; 7 items: *M* = 816.05, *SD* = 168.90; 10 items: *M* = 788.71, *SD* = 162.48). Because we also observed a significant interaction between target type and display size, $F(2, 30) = 20.37$, $p < .001$, we conducted further analyses to determine the simple effect of display size for each target type and found that RTs decreased with increasing display size for the greater-than-five condition, $F(1, 15) = 35.86$, $p < .001$ (5 items: *M* = 864.81, *SD* = 185.21; 7 items: *M* = 766.41, *SD* = 126.03; 10 items: *M* = 699.25, *SD* = 132.52; all *ps* < .01 for pairwise comparisons). The slight increases in RTs with increasing display size for the less-than-five condition were not significant, $F(2,30) = 2.29$, $p = .119$.
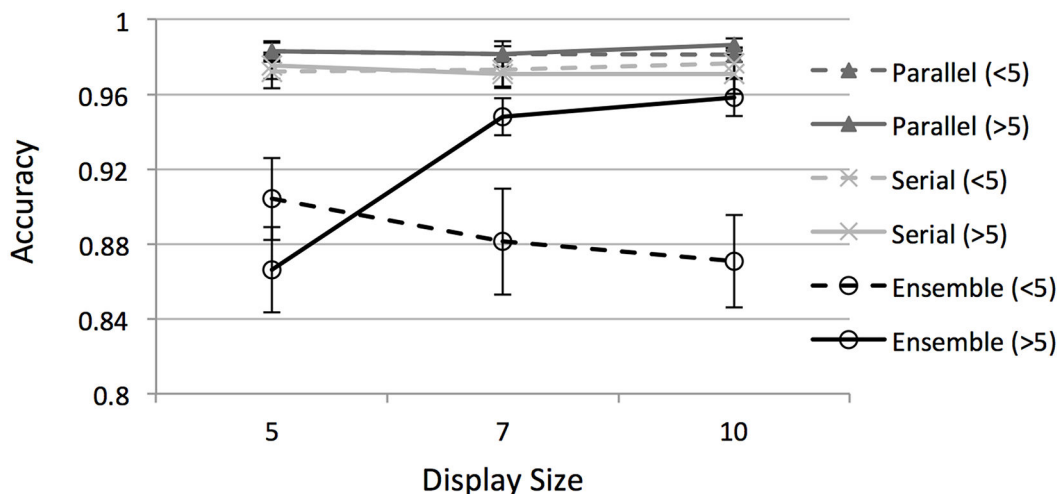

Figure 5. Accuracy for the serial search (light gray), parallel search (dark gray), and ensemble (black) tasks in Experiment 1 plotted as a function of display size and separated by target type (dashed lines indicate < 5 and solid lines indicate > 5).

Note. Error bars represent standard error of the mean. The only significant effect of display size on accuracy reflected increased accuracy with larger displays in the greater-than-five condition (black solid line), $F(2, 30) = 19.77$, $p < .001$.

In order to determine whether any observed effects on RT were due to speed/accuracy trade-offs, accuracy was also submitted to a 3 (*Task*: Ensemble/Serial Search/Parallel Search) x 2 (*Target Type*: < 5/> 5) x 3 (*Dis-*

play Size: 5/7/10 items) repeated measures ANOVA. There was a significant main effect of task, $F(2, 30)$ = 17.92, $p < .001$, such that accuracy was lower for the ensemble ($M = .90$, $SD = .08$) compared to the serial ($M = 0.97$, $SD = .03$) and parallel ($M = .98$, $SD = .02$) search tasks, consistent with the differences observed in RTs. A main effect of target type, $F(1, 15) = 6.91$, $p = .019$, was due to higher accuracy in the greater-than-five condition for the ensemble task ($M = .92$, $SD = 0.06$) compared to the less-than-five condition ($M = .89$, $SD = 0.10$). There was also a main effect of display size due to changes in accuracy with changing display sizes detailed below, $F(2, 30) = 3.93$, $p = .031$.

A significant task x target type x display size interaction, $F(4, 60) = 7.24$, $p < .001$, was further examined with a 2 (Target Type) x 3 (Display Size) ANOVA for each task type, which revealed a significant main effect of display size for the ensemble task, $F(2, 30) = 8.95$, $p = .001$, due to increased accuracy with more items in the display (5 items: $M = 0.89$, $SD = 0.09$; 7 items: $M = 0.91$, $SD = 0.08$; 10 items: $M = 0.91$, $SD = 0.07$). A significant main effect of target type, $F(1, 15) = 7.11$, $p = .018$, reflected higher accuracy for greater-than-five ($M = 0.92$, $SD = 0.06$) compared to the less-than-five displays ($M = 0.88$, $SD = 0.10$). A significant interaction between target type and display size, $F(2, 30) = 8.39$, $p = .001$, reflected that the main effect of display size described above was driven by increasing accuracy with larger displays in the greater-than-five condition, $F(2, 30) = 19.77$, $p < .001$. Pairwise comparisons revealed significantly lower accuracy for 5-item displays ($M = 0.87$, $SD = 0.09$) compared to 7- ($M = 0.95$, $SD = 0.04$) and 10-item ($M = 0.96$, $SD = 0.04$) displays in the greater-than-five condition (both ps $< .005$), but no significant difference between 7- and 10-item displays. There was no effect of display size in the less-than-five condition. All of these effects on accuracy are consistent with the RT effects, thus ruling out speed/accuracy trade-offs. There were no main effects of display size or target type on accuracy for either search task.

Discussion

As expected, in Experiment 1, increasing the number of items in the search task displays resulted in longer RTs for finding the digit that was greater than five or less than five, and much more so for the serial compared to parallel task in which the target was always red. This pattern indicates that in our serial search task, participants required additional processing time for additional items. Critically, when participants were asked to estimate the average of digits in the display (ensemble task), RTs decreased (for the > 5 displays) or stayed the same (for the < 5 displays), suggesting that individual items do not contribute serially to mean estimates; if anything, additional items may increase the efficiency of generating the set statistic as has been shown for high-level visual attributes such as gaze direction (Sweeny & Whitney, 2014). Moreover, accuracy improved with increased display size in the greater-than-five condition, ruling out the possibility that faster RTs for larger displays was due to a speed-accuracy tradeoff, and further supporting the notion that larger sets allow more efficient extraction of ensemble representations at the semantic level.

These results are consistent with the idea that a parallel processing mechanism underlies rapid estimation of the average of digit sets. However, because the digits we chose for the displays resulted in averages with mean differences from 5 that increased slightly across display sizes (from ±1.5 in the 5 item displays to ±2 in the 10 item displays), we wanted to determine whether this increase in distance from 5 contributed to the pattern of performance across display sizes. We therefore analyzed the subset of trials (~25 trials per display size per participant) for which the average distance from 5 was the same across display sizes (±1.8), and found the same pattern of results as in the main analysis (faster RTs for displays with more items, $F(2, 30) = 4.18$,

p < .05, with no speed/accuracy trade-off). This additional analysis confirmed that the observed improvement in performance with larger displays was not due to the small difference in the means of displays across conditions.

Although the fact that RTs did not increase with larger display sizes in the ensemble task is consistent with parallel processing of numerical value during average estimation, the differences in RT slopes seen for the greater-than-five and less-than-five conditions in Experiment 1 were surprising. We considered the possibility that because displays with numerical means that were greater than five tended to include digits with a greater number of line segments and thus may have appeared brighter, participants could have taken advantage of this low-level visual information in performing the ensemble task. In addition to serving as a cue to whether the average was greater-than or less-than-five, this brightness difference could have resulted in a relative benefit (faster RTs) for larger displays particularly in the brighter, greater-than-five condition, and thus may explain why the decrease in RTs with larger displays occurred only for that condition. Therefore, we conducted a second experiment in which we used a restricted set of digits (2, 3 and 7, 8). This approach eliminated the systematic relationship between the correct response and the brightness of the displays present in Experiment 1, as displays composed of 2s and 3s contain the same number of line segments, on average, as those with 7s and 8s. In Experiment 2, participants performed either a serial search task in which they searched for a number less than (2 or 3) or greater than five (7 or 8), or an ensemble task in which they indicated whether the average of the digits was less than or greater than five.

Experiment 2: Ruling out Brightness as a Cue

Method

Participants

Thirty-two undergraduate students from a medium-sized public university in the Mid-South U.S. volunteered to participate for class credit. The university's Institutional Review Board approved all experimental procedures. Task was a between group factor; 16 participants completed the search task and 16 completed the ensemble task. Data from one participant was excluded due to RTs that were greater than 2.5 standard deviations above the mean in the ensemble task. Of the remaining 31 participants, 14 females and two males (M_{age} = 21.5, *SD* = 2.34) completed the search task, and 10 females and five males (M_{age} = 21.5, *SD* = .99) completed the ensemble task.

Stimuli and Procedure

The procedure for Experiment 2, which was conducted on a MacBook computer connected to a CRT monitor with a screen resolution of 1024 x 768 pixels, was the same as for Experiment 1, except that task type (serial search/ensemble task) was a between group factor. In addition, we did not include a parallel search block because the RT pattern from Experiment 1 clearly replicated numerous previous studies showing small or no effects of display size on RT or accuracy with pop-out targets such as the red digit among white digits. The stimuli used for Experiment 2 were displayed in the same manner as in Experiment 1, but to eliminate the systematic relationship between correct response and brightness of the displays, digits were limited to 2, 3 and 7, 8 (Figure 6). In our "digital clock" font, the 2 and 3 each contain five line segments, the 7 contains three line segments, and the 8 contains seven line segments. As a result, the set of distractors (in the serial search task) or numbers

on the "wrong" side of 5 (in the ensemble task) yielded the same overall brightness in all displays. An additional benefit of using 2, 3 and 7, 8 is that the average numerical distance from 5 is the same for each pair. Means of each display ranged from ±1.4 to ±2.1 relative to 5, with average differences from 5 for the 5, 7, and 10 item displays that varied within a similar range as in Experiment 1 (±1.5, ±1.8, and ±2, respectively).
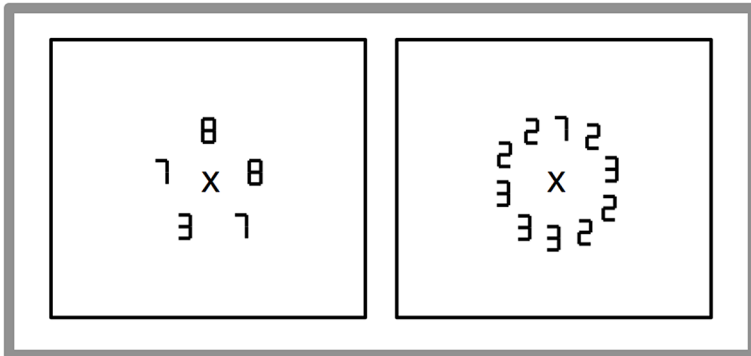

Figure 6. Example digit arrays for serial search and ensemble estimation task in Experiment 2.

Note. Displays consisted of 5, 7, or 10 items with digits consisting of only 2, 3 and 7, 8. Participants were asked to estimate the average numerical value of the digits as less than or greater than five for the ensemble task, and to report the location (left or right side of the screen) of the less-than-five or greater-than-five digit for the search task.

Results

For each participant, we excluded all trials with RTs that were either three standard deviations longer than the mean for that participant or less than 100 ms; a total of 2.0% of data points were removed from each task. RTs were submitted to a 2 (*Task*: Serial Search/Ensemble) x 2 (*Target Type*: < 5/> 5) x 3 (*Display Size*: 5/7/10 items) repeated measures ANOVA with task as a between-subjects factor. There was a marginal main effect of display size, $F(2, 58) = 3.14$, $p = .05$, with larger display sizes yielding longer RTs (5 items: $M = 684.26$, $SD = 156.14$; 7 items: $M = 682.53$, $SD = 127.87$; 10 items: $M = 705.27$, $SD = 126.07$), and a significant interaction between task and display size, $F(2, 58) = 95.96$, $p < .001$. We examined this two-way interaction by conducting a 2 (Target Type) x 3 (Display Size) repeated measures ANOVA for each task type.

For the search task, a significant effect of display size was observed, $F(2, 30) = 187.73$, $p < .001$, such that RTs increased with increasing display size (5 items: $M = 591.75$, $SD = 81.83$; 7 items: $M = 643.77$, $SD = 87.33$; 10 items: $M = 747$, $SD = 103.03$). In contrast, as can be seen in Figure 7, a significant main effect of display size for the ensemble task, $F(2, 28) = 19.51$, $p < .001$ was due to larger display sizes yielding faster RTs (5: $M = 776.77$, $SD = 206.94$; 7: $M = 722.14$, $SD = 162.81$; 10: $M = 662.63$, $SD = 144.81$). There was no main effect of target type, $F(1, 15) = 3.49$, $p = .082$, and no interaction between target type and display size for either task.

A 2 (*Task*: Serial Search/Ensemble) x 2 (*Target Type*: < 5/> 5) x 3 (*Display Size*: 5/7/10 items) repeated measures ANOVA with task as a between-subjects factor was also conducted on accuracy. Accuracy was significantly higher for the search ($M = 0.97$, $SD = .02$) compared to the ensemble ($M = .92$, $SD = .06$) task, $F(1, 29) = 15.21$, $p = .001$. This main effect was qualified by a significant task x display size interaction, $F(2, 58) = 4.27$, $p = .019$; follow-up 2 (Task) x 3 (Display Size) repeated measures ANOVAs for each task type revealed no main effects or interactions for the search task. As shown in Figure 8, there was a significant main effect of display size for the ensemble task, $F(2, 28) = 4.06$, $p = .028$ due to increasing accuracy with increasing display size.

Pairwise comparisons revealed significantly higher accuracy for the 10-item display (*M* = .93, *SD* = .04) com-pared to the 5-item display (*M* = .91, *SD* = .07), *t*(14)= -2.30, *p* = .04, but no significant difference between the 5- and 7-item displays or 7- and 10-item displays. There was no interaction between display size and target type, *F*(2,28) = 2.843, *p* = .075.
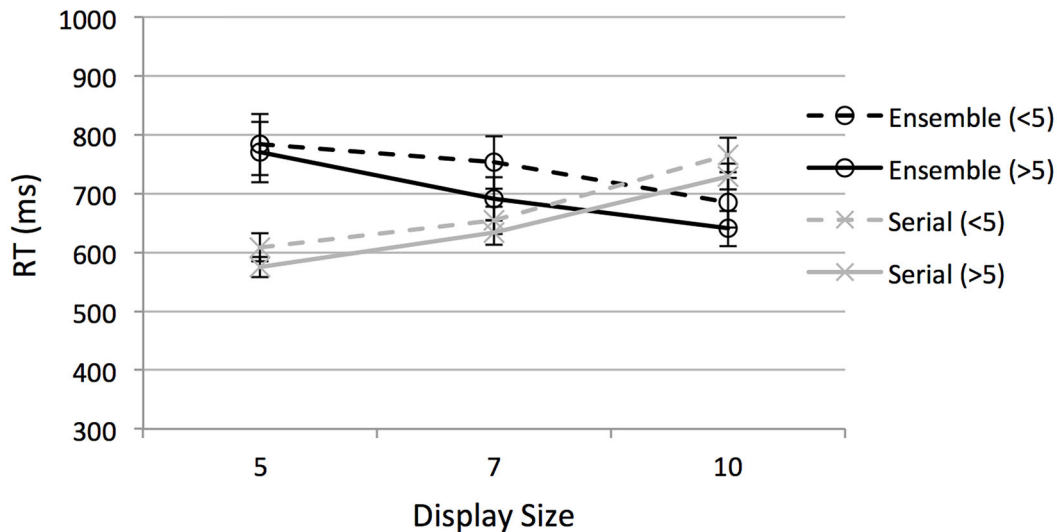

Figure 7. RTs for the serial search (gray) and ensemble (black) tasks in Experiment 2 plotted as a function of display size and separated by target type.

Note. Dashed lines indicate < 5 and solid lines indicate > 5. Error bars represent standard error of the mean. For serial search, RTs significantly increased with increasing display size, whereas for the ensemble task, RTs significantly decreased with increasing display size.
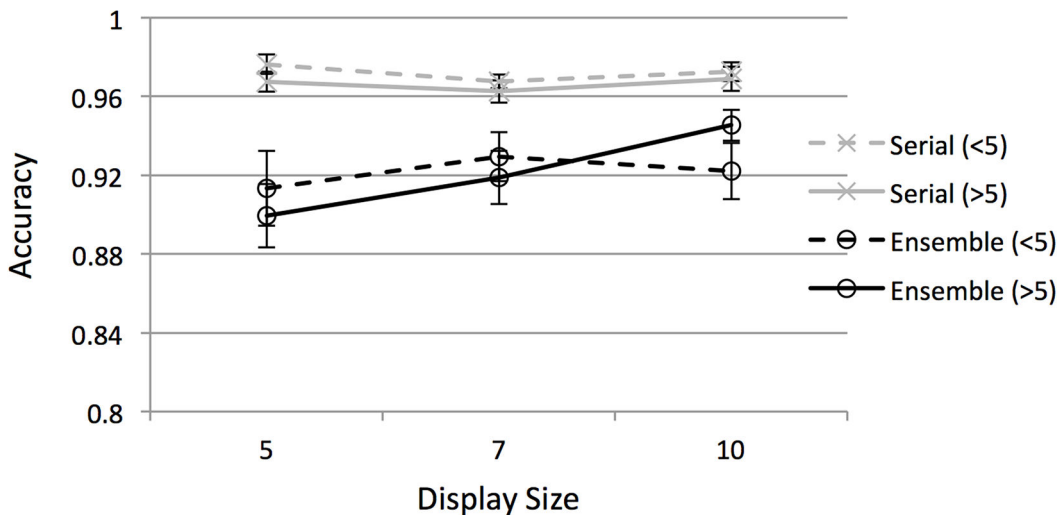

Figure 8. Accuracy for the serial search (gray) and ensemble (black) tasks in Experiment 2 plotted as a function of display size and separated by target type.

Note. Dashed lines indicate < 5 and solid lines indicate > 5. Error bars represent standard error of the mean. Overall, accuracy was significantly higher for serial search compared to the ensemble task. Accuracy was not affected by display size for search, whereas larger display sizes led to greater accuracy for the ensemble task.

Discussion

In Experiment 2, we controlled for the possibility that in Experiment 1, the reduction in RTs in the greater-than-five condition, and indeed the overall lack of longer RTs with additional items in the ensemble task, was due to participants relying on a low-level cue such as brightness rather than semantic information in the displays to perform the averaging task. Using a digit set that allowed us to control for brightness differences, we found that as in Experiment 1, RTs increased with increasing display size for the search task. Just as in the "serial search" task in Experiment 1, the target did not contain a unique feature and therefore search through the items proceeded in a serial fashion, requiring additional processing time for larger displays (Treisman & Gelade, 1980).

Most important, however, is that even after ensuring that brightness could not be used as a reliable indicator of whether the average value of the display was greater than or less than five, we again found that in the ensemble task, RTs did not increase, but instead decreased with increasing display size for both the less-than-five and greater-than-five conditions. Furthermore, accuracy of mean estimates increased with additional items. This improvement in performance with increasing display size is consistent with our findings in Experiment 1 for displays with averages greater than five, but in this case, occurred for both types of displays, and cannot be due to brightness differences between conditions. These results suggest not only that additional digits in the display do not require additional processing time, but also that they may increase the efficiency of ensemble perception at the semantic level (Brezis et al., 2015; Piazza et al., 2013; Sweeny & Whitney, 2014).

General Discussion

The current research investigated the mechanisms involved in extracting semantic information from groups of digits by examining the dynamics of estimating numerical averages of digit ensembles that varied in size. In separate tasks, we presented participants with identical digit displays and asked them to either search for a target digit that is less than or greater than five, or report whether the numerical mean of the digits is less than or greater than five. By equating the search and numerical averaging (ensemble) tasks but providing different instructions, we were able to compare RT patterns for the search and ensemble tasks to each other, and also to well-established RT patterns reported in the visual search literature. These comparisons allowed us to make inferences about the mechanisms underlying ensemble perception of numerical value based on the effect of additional items on RTs. In Experiment 1, along with the ensemble task, we included two versions of the search task. In the serial search task, finding the target took significantly longer with increased display size, as the target did not possess a unique salient feature. The classical view of visual search behavior interprets such RT increases with additional items as characteristic of a serial process in which information associated with each item is processed individually. In the parallel search task, the target was presented in a salient color, and so it was not surprising that additional items resulted in only a miniscule increase in RTs (1-2 ms per item). The key result, however, is that RTs in the ensemble task, unlike in either of the search tasks, reduced overall as display size increased.

Because we were interested in the contribution of semantic as opposed to lower-level visual information to perception of the average of digit displays, in Experiment 2 we addressed the concern that in Experiment 1, displays with greater-than-five averages tended to have more line segments, and therefore greater overall brightness than those with less-than-five averages. This brightness difference could have served as a low-level, non-

semantic cue to the correct response, and also potentially contributed to the difference seen between the greater-than-five (decrease in RTs with increasing display size) and less-than-five (no change in RTs across display sizes) conditions. Thus, we conducted Experiment 2 with a limited stimulus set in order to control for systematic brightness differences across display sizes. Experiment 2 again demonstrated reduced RTs with increased display size while controlling for any possible contribution of brightness differences to performance on the ensemble task. Thus, a relationship between brightness and set average cannot explain the lack of increased RTs with additional set items, and the fact that RTs decreased with larger sets in Experiment 2 is not attributable to the relatively greater salience of displays with more items in the greater-than-five condition as it could have been in Experiment 1.

This overall decrease in RTs with increasing display size stands in contrast to the small increase or lack of increase often seen in pop-out, or parallel search tasks (Egeth, 1966; Treisman & Gelade, 1980), but is consistent with reports of performance benefits with the addition of more items during ensemble perception of gaze, a relatively high-level attribute (Sweeny & Whitney, 2014). This pattern of shorter RTs with increasing display size may also relate to other findings that additional items in sequentially presented digit displays contribute to faster RTs (Brezis et al., 2015). It could be that the numerical average estimation task engages an ensemble coding mechanism and thus the additional information provided in larger displays increases the efficiency of the response. Brezis et al. (2015) suggest that different mechanisms are engaged for digit averaging tasks depending on display size; when participants were asked to estimate the average of series of digits presented rapidly one after another, RTs were shorter for larger displays of 16 compared to smaller displays of 8 digits, and also for 8 compared to 4 digits. Because ensemble coding allows observers to generate an efficient representation of a set by extracting information from multiple items in parallel, larger sets may provide more information on which to base the average and thus improve the efficiency of ensemble perception. Our finding is in line with the suggestion by Alvarez (2011) that displays with more items result in a higher signal-to-noise ratio (SNR), and reports that auditory stimuli with higher SNR are associated with faster RTs (Lentz, He, & Townsend, 2014).

Our data suggest that it is not necessary for participants to attend to each item individually in a display in order to extract semantic meaning from the display as a whole, as the ensemble task generated an RT pattern more similar to the parallel than the serial search, for which RTs steeply increased. This finding speaks to the temporal dynamics of extracting semantic information from multiple digits, as it demonstrates that there is no RT cost of adding additional items to a display. Rather, extracting ensemble information at the semantic level appears to occur in parallel across items within a display, and may even benefit from additional items.

To address the overall differences in RTs and accuracy across the three tasks in Experiment 1, RTs were slower and accuracy was lower for the ensemble task compared to both of the search tasks. This difference between the ensemble and search tasks can potentially be explained by differences in task demands. In the parallel search task, participants' attention was likely immediately drawn to the pop-out digit, generating the fastest RTs. In the serial search task, participants responded immediately after locating a digit less than five or greater than five, whereas for the ensemble task, participants were asked to estimate the average, a process that may involve inherently greater uncertainty. Note that whereas RTs for both the serial search and the parallel search increased across display sizes, in the parallel search task, RTs increased by less than 6 ms from the smallest to largest display size, compared to a 160 ms increase for the serial search task. This pattern was predicted based on Egeth (1966), who found small increases in RTs as display sizes increased even in parallel tasks. Our

data also confirm slower RTs for the serial search compared to the parallel search. This is expected because the salient target in the parallel search captures participants' attention, whereas according to the classical view, in the serial search participants would have to scan each item in the display until finding the target (Duncan & Humphreys, 1989; Treisman & Gelade, 1980). Overall, RT patterns for the ensemble task did not increase with increased display size as they did in serial search, but in fact stayed the same or decreased, suggesting that participants did not need to examine each digit individually in order to incorporate it into a representation of the average. Rather, participants were able to extract information from items across the entire display in parallel.

As discussed above, in Experiment 1, the reduction in RTs for larger displays in the ensemble task was driven by the greater-than-five condition; RTs in the less-than-five condition remained consistent across display sizes. Furthermore, in Experiment 1, accuracy increased with display size for only the greater-than-five condition. In addition to possible contribution of the brightness difference between conditions, it may be that participants showed a sample size bias such that they were more likely to respond that the mean is greater than five when there were more items in the display. Evidence for sample size biases whereby larger set sizes yield larger average estimations has been demonstrated in a task where participants were asked to provide numerical estimates of the mean of set of digits (Smith & Price, 2010). In our experiment, such a bias would be expected to also result in proportionately lower accuracy with increasing display size in the less-than-five condition; in Experiment 1, there was a non-significant trend in this direction, however in Experiment 2, the pattern of increased accuracy across display sizes did not differ based on whether the average was greater or less than five. In Experiment 1, RTs were faster overall for displays with greater-than-five averages in the ensemble task, as well as for the trials with greater-than-five targets in the serial search task. For the ensemble task, this could have been related to a bias as discussed above because it was seen only for larger sets (7 and 10 items), or because displays with more items in the greater-than-five condition might be more likely to engage a brightness strategy (arrays of digits whose mean is greater than five contained more line segments on average across the whole digit set, so brightness would increase more with larger sets for greater-than-five compared to less-than-five displays). Average brightness differences between target types could also potentially explain faster responses to the greater-than-five target observed in the serial search task in Experiment 1; alternatively, this pattern may reflect an advantage for overweighting of digits with larger values, as has been reported in previous work (Sobel et al., 2016; Van Opstal et al., 2011). However, in Experiment 2, there was no effect of target or display type (> 5 or < 5) on RTs or accuracy for either the search or ensemble task.

Although our second experiment successfully controlled for increased brightness with increasing display size, we considered whether participants could have used other low-level cues to perform the task. For example, in Experiment 1, the digit in each display that differed from the rest by virtue of being the only one greater than or less than five also possessed a unique shape. However, in the serial search task in Experiment 1, the target does not appear to have popped out the way the red targets did in the parallel search task, because we saw much more substantial RT costs for additional items in the serial compared to parallel search tasks. Thus, it is unlikely that the unique shape of the one item on the "other side" of five determined performance in either the search or ensemble tasks. It is also possible that in Experiment 2, the limited digit set (2, 3 and 7, 8) could have encouraged perceptual grouping of the more prevalent numbers (because only two shapes were represented) and thus facilitated arrival at a mean estimate. On the other hand, such perceptual grouping should have also allowed the lone digit on the other side of five to pop out in the search task; in contrast, we saw that additional items gave rise to longer RTs just as in Experiment 1, as is commonly observed in search tasks in which attention to individual items is thought to be required (Duncan & Humphreys, 1989; Treisman & Gelade, 1980).

Also related to the shape of the stimuli, there is a chance that in Experiment 2, the similarity between the shapes of the digits 2 and 3 led to greater homogeneity in displays with less-than-five averages compared to those with greater-than-five averages, which primarily contained the less similar (to one another) 7s and 8s. Here, as above, it is useful to examine the results of the search task using identical displays to explore the possibility that a difference in homogeneity, which would be more pronounced for larger displays, can explain the reduction in RTs with more items in the ensemble task. Specifically, if the decrease in RTs with larger displays in the averaging task was driven by a variance difference between the two conditions, we would expect to see an effect of this difference in our search tasks such that searching for a digit greater than five among 2s and 3s would be more efficient (produce smaller RT increases with larger displays) than searching for digit less than five among 7s and 8s, as the target digit would pop out to a greater extent among the more homogeneous array of 2s and 3s. Such a prediction is in accordance with the well-established result that searching for a target among distractors that are more similar to one another leads to shallower RT slopes (Duncan & Humphreys, 1989). However, our data show no main effect of or interaction with target type on RTs, indicating no difference in the efficiency of search between the greater-than-five and less-than-five targets in Experiment 2, and suggesting that participants were not sensitive to the difference in variance between conditions. Because these variance differences did not seem to affect performance in the search task, they are also unlikely to explain the results of our averaging task. Nonetheless, there is a possibility that the ensemble task engaged mechanisms that are relatively sensitive to variance in the display (Haberman, Lee, & Whitney, 2015), which may reflect a limitation of our study that should be addressed by future work.

Thus, although our data may not demonstrate unequivocally that the same mechanism is engaged for estimating the numerical average of digit displays as for extracting set statistics based on lower-level visual characteristics, our results indicate that examining individual digits serially is not required for making a decision about average value. The notion that semantic information can be efficiently extracted from multiple digits is consistent with the findings of Van Opstal and colleagues (2011) in which participants were sensitive to the average value of subliminally presented digit displays. Results from the current study align with others aimed at understanding the temporal dynamics of ensemble perception of higher-level stimulus attributes (Haberman & Whitney, 2007; Marchant & de Fockert, 2009; Van Opstal et al., 2011), and suggest that we are able to gather useful semantic information and abstract meaning from groups of stimuli rapidly and without attention to individual items as has been shown for more basic stimulus attributes (Alvarez & Oliva, 2008; Ariely, 2001; Parkes et al., 2001). Moreover, our findings extend beyond those of Van Opstal and colleagues (2011) because we used larger digit arrays such as to exceed the subitizing range and also to reduce the likelihood that participants would attempt an exact averaging technique. We additionally utilized larger differences in the number of items in the displays, and controlled for low-level brightness information to more strongly support the conclusion that observers likely use a parallel processing mechanism to quickly estimate the average from a group of digits.

Further research is needed to be able to determine whether such averaging tasks are completed using an ensemble coding mechanism as it has been defined thus far. It may be that another type of averaging mechanism is engaged for digits because they have a true average and because they contain both semantic and perceptual information with an arbitrary relationship. It has been demonstrated that semantic information associated with digits influences visual search independently of perceptual factors (Sobel et al., 2015); however, processing of digits may be affected by the relationship between perceptual and semantic information differently depending on the task. In studies using digit comparison tasks, there is a clear interaction between digits' physical and numerical size in the form of the size congruity effect, in which RTs are faster when physical and numerical size

is congruent compared to incongruent (Besner & Coltheart, 1979; Henik & Tzelgov, 1982). However, in a recent visual search study, participants' performance was driven primarily by the physical size of digits (perceptual) and not by their value (semantic meaning), indicating separate mental processes for each (Sobel et al., 2017). There is still debate about which level of processing may join these two types of information if they truly do occupy distinct mental representations that interact at later decision-making stages (Faulkenberry et al., 2016; Santens & Verguts, 2011; Sobel et al., 2017). It may be that although both perceptual and semantic information can be extracted from multiple items in parallel, they remain in separate channels until their interaction is facilitated by selective attention to individual items, as has been proposed for integration of separate perceptual features (Treisman & Gelade, 1980) as well as for the interference observed in the SCE (Sobel et al., 2016).

Funding

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*, 122-131. https://doi.org/1https://doi.org/10.1016/j.tics.2011.01.003doi:10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392-398. doi:10.1111/j.1467-9280.2008.02098.x

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 7345-7350. doi:10.1073/pnas.0808981106

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

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia, 17*, 467-472. doi:10.1016/0028-3932(79)90053-8

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

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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

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

Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433-458. doi:10.1037/0033-295X.96.3.433

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics, 1*, 245-252. doi:10.3758/BF03207389

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect. *Acta Psychologica, 163*, 114-123. doi:10.1016/j.actpsy.2015.11.010

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for top-down effects. *Behavioral and Brain Sciences, 39*, Article e229. doi:10.1017/S0140525X15000965

Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review, 21*, 689-695. doi:10.3758/s13423-013-0547-4

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

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*, R751-R753. doi:10.1016/j.cub.2007.06.039

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

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

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*, 1367-1368. doi:10.1038/nn1794

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*, 389-395. doi:10.3758/BF03202431

Krause, F., Bekkering, H., Pratt, J., & Lindemann, O. (2017). Interaction between numbers and size during visual search. *Psychological Research, 81*(3), 664-677. doi:10.1007/s00426-016-0771-4

Lentz, J. J., He, Y., & Townsend, J. T. (2014). A new perspective on binaural integration using response time methodology: super capacity revealed in conditions of binaural masking release. *Frontiers in Human Neuroscience, 8*, Article 641. doi:10.3389/fnhum.2014.00641

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*, 566-577. doi:10.1016/j.cognition.2008.03.009

Lupyan, G., Thompson-Schill, S. L., & Swingley, D. (2010). Conceptual penetration of visual processing. *Psychological Science, 21*, 682-691. doi:10.1177/0956797610366099

Marchant, A. P., & de Fockert, J. W. (2009). Priming by the mean representation of a set. *Quarterly Journal of Experimental Psychology, 62*, 1889-1895. doi:10.1080/17470210902871045

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information* (1st ed.). New York, NY, USA: W.H. Freeman and Company.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1517-1539. doi:10.1037/0096-1523.25.6.1517

Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2016). Serial vs. parallel models of attention in visual search: Accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review, 23*(5), 1300-1315. doi:10.3758/s13423-015-0978-1

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

Oriet, C., & Corbett, J. (2008). Evidence for rapid extraction of average size in RSVP displays of circles [Abstract]. *Journal of Vision, 8*(6), Article 13. doi:10.1167/8.6.13

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

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. doi:10.1177/0956797612473759

Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research, 48*, 589-597. doi:10.1016/j.visres.2007.11.017

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341-365. doi:10.1017/S0140525X99002022

Railo, H., Karhu, V. M., Mast, J., Pesonen, H., & Koivisto, M. (2016). Rapid and accurate processing of multiple objects in briefly presented scenes. *Journal of Vision, 16*, Article 8. doi:10.1167/16.3.8

Santens, S., & Verguts, T. (2011). The size congruity effect: Is bigger always more? *Cognition, 118*, 94-110. doi:10.1016/j.cognition.2010.10.014

Schwarz, W., & Heinze, H. J. (1998). On the interaction of numerical and size information in digit comparison: A behavioral and event-related potential study. *Neuropsychologia, 36*, 1167-1179. doi:10.1016/S0028-3932(98)00001-3

PsychOpen GOLD

Smith, A. R., & Price, P. C. (2010). Sample size bias in the estimation of means. *Psychonomic Bulletin & Review, 17*(4), 499-503. doi:10.3758/PBR.17.4.499

Sobel, K. V., & Puri, A. M. (2018). Size congruity influences visual search via the target template. *Acta Psychologica, 183*, 66-74. doi:10.1016/j.actpsy.2018.01.003

Sobel, K. V., Puri, A. M., & Faulkenberry, T. J. (2016). Bottom-up and top-down attentional contributions to the size congruity effect. *Attention, Perception & Psychophysics, 78*(5), 1324-1336. doi:10.3758/s13414-016-1098-3

Sobel, K. V., Puri, A. M., Faulkenberry, T. J., & Dague, T. D. (2017). Visual search for conjunctions of physical and numerical size shows that they are processed independently. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 444-453. doi:10.1037/xhp0000323

Sobel, K. V., Puri, A. M., & Hogan, J. (2015). Target grouping in visual search for multiple digits. *Attention, Perception & Psychophysics, 77*, 67-77. doi:10.3758/s13414-014-0761-9

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

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46-54. doi:10.1111/j.1467-9280.1990.tb00067.x

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136. doi:10.1016/0010-0285(80)90005-5

Van Opstal, F., de Lange, F. P., & Dehaene, S. (2011). Rapid parallel semantic processing of numbers without awareness. *Cognition, 120*, 136-147. doi:10.1016/j.cognition.2011.03.005

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31*(4), 523-535. doi:10.1016/S0896-6273(01)00392-0

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space, and quantity. *Trends in Cognitive Sciences, 7*, 483-488. doi:10.1016/j.tics.2003.09.002

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research, 29*, 47-59. doi:10.1016/0042-6989(89)90173-9

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

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews: Neuroscience, 5*, 495-501. doi:10.1038/nrn1411