

The association of color memory and the enumeration of multiple spatially overlapping sets

Sonia Poltoratski

Vanderbilt University, Nashville, TN, USA



Yaoda Xu

Harvard University, Cambridge, MA, USA



Using dot displays, Halberda, Sires, and Feigenson (2006) showed that observers could simultaneously encode the numerosity of two spatially overlapping sets and the superset of all items at a glance. With the brief display and the masking used in Halberda et al., the task required observers to encode the colors of each set in order to select and enumerate all the dots in that set. As such, the observed capacity limit for set enumeration could reflect a limit in visual short-term memory (VSTM) capacity for the set color rather than a limit in set enumeration per se. Here, we largely replicated Halberda et al. and found successful enumeration of approximately two sets (the superset was not probed). We also found that only about two and a half colors could be remembered from the colored dot displays whether or not the enumeration task was performed concurrently with the color VSTM task. Because observers must remember the color of a set prior to enumerating it, the under three-item VSTM capacity for color necessarily dictates that set enumeration capacity in this paradigm could not exceed two sets. Thus, the ability to enumerate multiple spatially overlapping sets is likely limited by VSTM capacity to retain the discriminating feature of these sets. This relationship suggests that the capacity for set enumeration cannot be considered independently from the capacity for the set's defining features.

Introduction

Even without actively counting, we can reliably and effortlessly estimate the number of objects in a visual display (for a review, see Dehaene, Dehaene-Lambertz, & Cohen, 1998; Feigenson, Dehaene, & Spelke, 2004). In fact, the visual approximation of number is often presented as an innate, “core” ability, distinct from symbolic mathematics and largely automatic (Feigenson et al., 2004). As such, enumeration poses a unique challenge to our visual system. It necessitates the

perception of multiple, disconnected objects as a single entity (a “set”). Moreover, unlike other properties by which a set can be defined, like color or shape, enumeration requires the perception of all of the elements in a set: One cannot successfully enumerate by sampling a few objects.

While studies on approximate number judgment have almost always instructed observers to enumerate all the items in a visual display, in daily life we rarely do so. Rather, observers often use selective attention to enumerate a set (or sets) of objects that fulfill certain selection criteria. For example, in an office desk scene one can choose to judge the number of pens in a cup, the number of papers in a stack, or the total number of objects in view.

Halberda et al. (2006) recently examined how sets of objects may be selected and enumerated. They briefly presented displays containing dots of highly distinguishable colors and asked observers to estimate the number of dots in either one specific probe color or the superset of all dots in the display. The color probe was given either before or after the enumeration display, and the total number of color sets varied from one to six. In the probe-before condition, observers could selectively attend and enumerate the specified color set; in fact, performance across this condition was uniformly high, suggesting that extraction of the set from the displays did not noticeably increase in difficulty as distractor sets were added. However, in the probe-after condition, observers had to attend and enumerate all the color sets in order to perform the task accurately. By calculating when performance differed between the probe-before and the probe-after conditions, Halberda et al. found that observers could reliably retain and enumerate approximately three sets at a glance: two colored sets and the superset of all dots in the display. Halberda et al. concluded that the capacity for enumeration of sets is about three.

What determines the capacity limitation of set enumeration? Because Halberda et al. (2006) used brief

Citation: Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, 13(8):6, 1–11, <http://www.journalofvision.org/content/13/8/6>, doi:10.1167/13.8.6.

display presentation duration and masking, it was necessary for observers to first remember a particular color before they could enumerate the number of dots in that color. As such, the observed capacity limitation for set enumeration may be determined by how many colors observers could extract and retain from the dot displays. That is, the visual short-term memory (VSTM) capacity limit for colors, rather than the capacity limit of set enumeration per se, may have determined how many sets can be enumerated successfully from a given display. It is equally possible, however, that VSTM capacity for objects and capacity for set enumeration are dissociable and mediated by different cognitive and neural mechanisms. After all, previous research has shown that VSTM capacity for highly distinguishable colors is between three and four (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Pashler, 1988; Vogel, Woodman, & Luck, 2001). As such, VSTM capacity for the set color may be between three and four, but set enumeration capacity may always be fixed to two sets. In other words, observers may be able to retain the colors of more sets than they can successfully enumerate. Understanding the association between the capacity for set enumeration and VSTM capacity for set colors is thus critical if we wish to know what determines the limitation in set enumeration.

The present study aims to differentiate between these two possibilities. In Experiment 1, we simultaneously measured the capacity limit of set enumeration and the number of colors observers could retain in memory from the same display. In Experiment 2, using the exact same displays and a similar paradigm, we measured VSTM capacity for color without the enumeration task. In Experiment 3, we compared VSTM capacity for color with and without the enumeration task within the same group of observers.

Experiment 1

In this experiment, we used a slightly modified paradigm of Halberda et al. (2006). In addition to asking observers to enumerate dots of a particular color, we asked observers to enumerate dots of a color that was absent in the display (for which the correct numerosity was zero). This allowed us to simultaneously measure VSTM capacity for color from the same displays.

Methods

Participants

Eight Harvard University undergraduate students (three males) aged between 18 and 25 participated in

this experiment for course credit. All had normal or corrected-to-normal vision acuity and normal color vision. The study was approved by the Harvard University Committee on the Use of Human Subjects in Research.

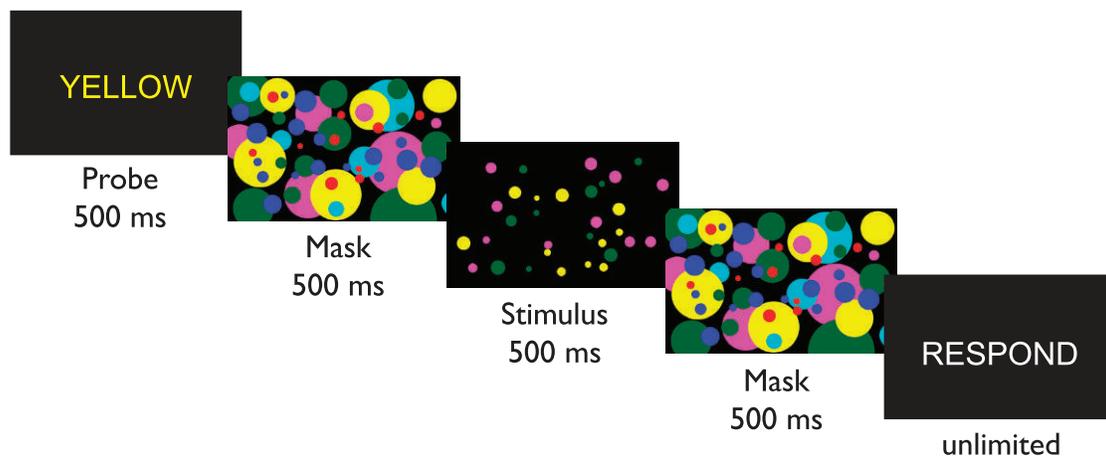
Materials and design

In this experiment, observers viewed displays containing dots whose total number varied from 1 to 35. The dots appeared in one to four sets with each set appearing in a unique color. A mask containing 60-plus partially overlapping dots was also presented before and after each enumeration display. As in Halberda et al. (2006), in the probe-before trials, a color probe was given before the presentation of the dot displays; in the probe-after trials, the color probe appeared after the dot displays were shown. In both types of trials, observers were asked to estimate the numerosity of a probed color set (see Figure 1). Each possible target numerosity (from 1 to 35 for one-set displays, 1 to 34 for two-set displays, 1 to 33 for three-set displays, and 1 to 32 for four-set displays) was probed equally often.

Six highly distinguishable colors were used: magenta [RGB: 255 0 255], yellow [255 255 0], green [0 100 0], cyan [0 255 255], blue [0 0 255], and red [255 0 0]. Colors for each display were randomly selected among these six. Following Halberda et al. (2006), we generated paired displays that were matched for the total number of color sets present and the numerosity in each set with one of these paired displays appearing in the probe-before and the other in the probe-after condition. The paired displays were not identical in that the color of each set and the dot locations were randomly chosen and not matched between the paired displays. To probe VSTM for color, on half of the trials the probed color was absent in the display. A total of 536 displays were pregenerated with 128 to 140 displays for each total number of sets (because only up to 34 dots with two sets, 33 with three sets, and 32 with four sets could be probed).

Displays subtended $38.4^\circ \times 21.6^\circ$ at a 76 cm viewing distance. As in Halberda et al. (2006), the total paint area of all dots in each set was fixed (at 26,214 pixels) with individual dot radii varying from 50% to 150% of the mean dot radius for that color set. This ensured that numerosity could not be inferred from estimating either the total paint area of all dots in a set or the radius of a single dot in any set. Following Halberda et al., we also made sure that in half of all the displays the probed set did not contain the largest numerosity in the display, discouraging observers from simply attending to the largest set on each trial. During the experiment, trials with varying numbers of colored sets and the two probe conditions were randomly intermixed.

Probe Before



Probe After

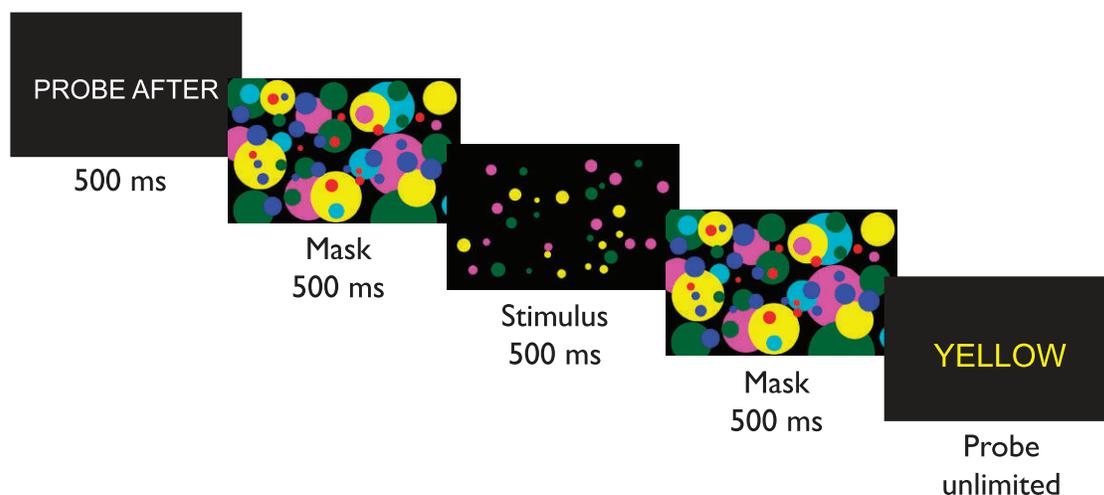


Figure 1. Examples of trials used in Experiment 1. Observers were shown briefly presented displays containing dots of highly distinguishable colors and were asked to estimate the number of dots in the probe color. The color probe was given either before or after the presentation of the enumeration display, which contained between one and four different color sets. To measure both the number of sets and the number of colors that observers could remember simultaneously from the same visual display, in half of the trials we also asked observers to enumerate dots of a color that was not present in the display (for which the correct numerosity was zero).

Apparatus

Stimulus generation and response recording were done using Matlab Psychophysics Toolbox (Brainard, 1997) on an Apple iMac computer running a 2.8 GHz processor. The stimuli were displayed on a 24-inch LCD monitor. Responses were collected on an Apple keyboard with a numeric keypad.

Procedure

Observers were seated in a dimly lit, quiet room and were given detailed verbal and visual instructions and practice. Trial sequence and timing are illustrated in Figure 1. Observers typed in their responses on the numeric keypad. They were instructed to respond “0” when the probe color was absent from the display. To

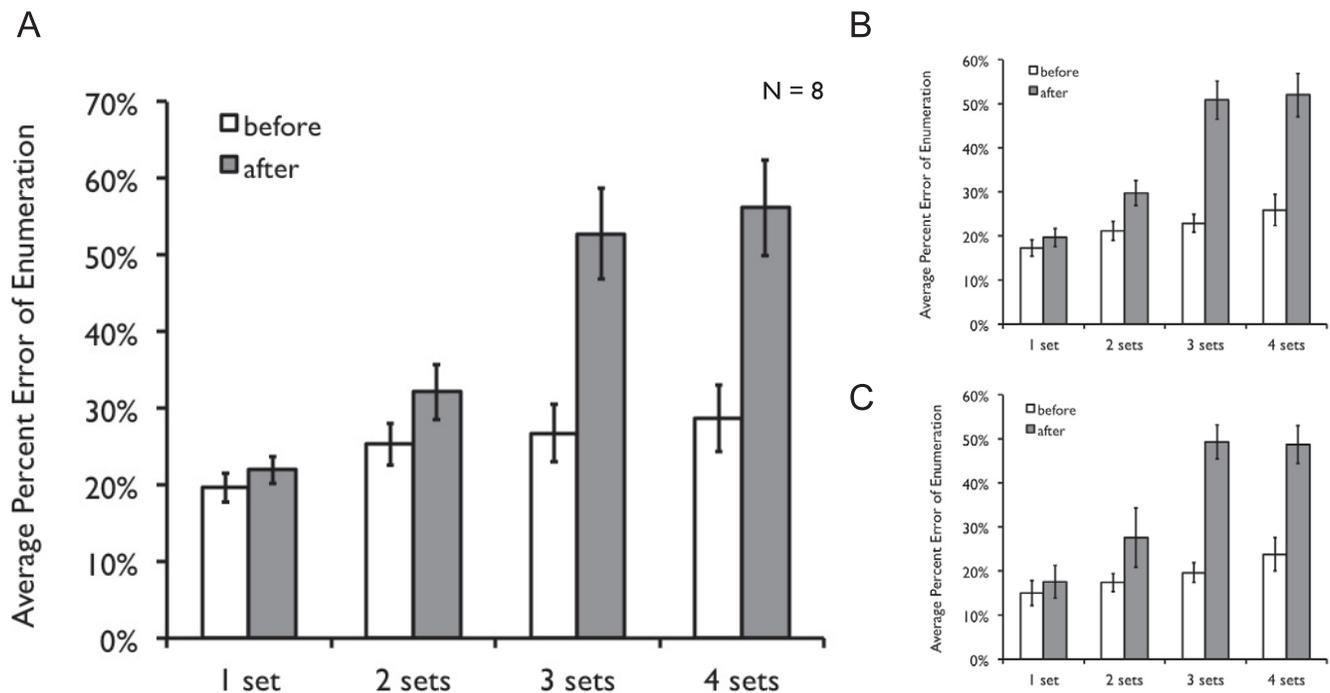


Figure 2. Experiment 1 enumeration results. Enumeration percentage errors were calculated by dividing the absolute difference between correct and reported numerosities by the correct numerosity. Enumeration percentage errors from all trials were plotted in (A). Replicating Halberda et al. (2006), we found that enumeration errors increased greatly in the probe-after condition in which the displays contained three or more sets, indicating that the capacity for enumeration was limited to two sets. We also found a small but significant probe-before advantage on two-set trials, suggesting that attentional selection of the task-relevant set beforehand could facilitate performance even when all the sets could be encoded successfully. Enumeration percentage errors were also calculated for trials in which the biggest set was probed (B) and in which it was not probed (C). There was no difference between (B) and (C), and both were similar to (A).

minimize the use of this response due to uncertainty, we emphasized that the “0” response should be used only when a color set was absent from the display. While observers were allowed an unlimited amount of response time, they were instructed to move through the trials quickly. The entire experiment lasted approximately 55 minutes.

Results and discussion

Capacity for enumerating sets

As in Halberda et al. (2006), to account for the fact that both mean and standard deviation of estimates increase linearly with the correct numerosity (Cordes, Gelman, Gallistel, & Whalen, 2001), we calculated the average percentage of enumeration error for a given display using the formula

$$\frac{|(\text{correct numerosity}) - (\text{measured numerosity})|}{(\text{correct numerosity})}$$

In Halberda et al. (2006), observers could never enter a “0” response and thus had to guess when they did not see the probed color. To match responses from our

study as closely as possible to those from Halberda et al., for trials in which observers reported “0” when the probed color set was present we substituted the averaged guessing response from trials in which the probed color was absent rather than entering zero as the measured numerosity. In other words, we used responses from the false-alarm trials to estimate what would have been guessing responses from the missed trials. This error correction was calculated separately for each observer and for the probe-before and probe-after conditions of each total number of sets. The percentage error was then calculated from this error correction and the correct numerosity of that trial. Overall, 0.6% of probe-before trials and 5.9% of probe-after trials were corrected in this way. The resulting corrected percentage error was, on average, 79.9% ($SD = 42.1\%$). With this error correction, our overall average percentage enumeration errors for three and four sets were around 50% (see Figure 2), very close to the corresponding errors reported by Halberda et al. In comparison, if we had used the response of “0” to calculate percentage error on those trials, the calculation would have yielded a uniform 100% error rate, and the overall error rate would be inflated.

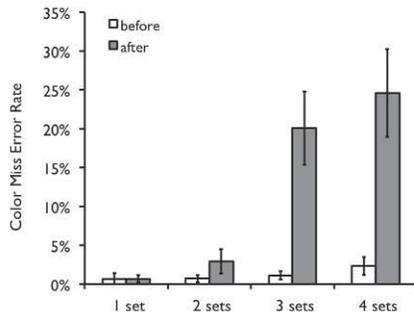


Figure 3. Average proportion of “miss” trials in Experiment 1. A “miss” was defined as a trial in which the probed color was present in the dot display but observers reported a numerosity of zero. Miss rates increased sharply in displays comprised of more than two subsets.

The average percentage enumeration error showed significant main effects of probe condition, set size, and an interaction between the two (multivariate general linear model: $F_s > 15.26$, $p_s < 0.002$). Within probe-before trials, there was no effect of set size, $F(3, 28) = 2.11$, $p = 0.12$, replicating the findings of Halberda et al. (2006). Within probe-after trials, except between three and four sets, which did not differ from each other, $t(7) = .57$, $p = 0.59$, the overall set-size effect and the differences between each adjacent set size were all significant ($t_s > 3.34$, $p_s < 0.02$). Probe-before and probe-after conditions did not differ for one set, $t(7) = 1.11$, $p = 0.31$; had a small but significant difference for two sets, $t(7) = 2.85$, $p < 0.03$; and had a large difference for three and four sets (both $t_s > 4.91$, $p_s < 0.002$).¹

Thus, when observers knew exactly which color to attend, they could select and enumerate that set regardless of the total number of sets present in the display. However, when the probe was given after the display, error rates increased with each added set and plateaued when three or more sets were present. These results largely replicated those of Halberda et al. (2006), showing that the numerosity of approximately two sets from a briefly presented display could be encoded.

We obtained a small but significant probe-before advantage on two-set trials. Halberda et al. (2006) also reported a hint of this effect although not significant (see figure 2c of Halberda et al., 2006). This suggests that attentional selection of the task-relevant set beforehand can facilitate performance even when all the sets can be encoded successfully.

VSTM capacity for set color

Figure 3 shows the proportion of trials in which observers “missed” the set, reporting a numerosity of zero when the probe color was actually present in the

display. It is evident that misses occurred much more frequently in three- and four-set displays than in displays containing fewer sets; the probe-after error rate is not significantly different between one- and two-set displays, $t(7) = 1.82$, $p = 0.11$; jumps between two and three sets, $t(7) = 3.75$, $p < 0.008$; and plateaus between three- and four-set displays, $t(7) = 1.41$, $p = 0.20$. These data suggest that as in Halberda et al. (2006), substantially more guessing occurred when a display contained more than two sets. As such, a large part of the enumeration error increase beyond two sets in Halberda et al. likely came from observers’ failure to encode the color of the probed color set and their subsequent guessing of numerosity in those trials.

To quantify VSTM capacity for color, we calculated Cowan’s (2001) K according to the formula: $(\text{hit rate} - \text{false alarm rate}) \times \text{set size}$. Hits were defined as color-present trials on which numerosity was correctly reported as greater than zero, and false alarms were color-absent trials on which numerosity was incorrectly reported as greater than zero. Only probe-after trials were included in this analysis. K values increased with total set size and plateaued at a set size of two (Figure 4). Specifically, Cowan’s K capacities at one to four sets were .92 (with SD of .05), 1.79 (.12), 2.08 (.44), and 2.19 (.82). Pair-wise comparisons revealed no difference in K s between two- and three-set displays or between three- and four-set displays ($t_s < 1.9$, $p_s > 0.1$). Thus, even for these highly discriminable colors, maximum VSTM capacity was about two colors, much lower than the three- or four-color capacity typically reported in previous VSTM studies (e.g., Alvarez & Cavanagh, 2004; Luck & Vogel, 1997).

Encoding bias

Following Halberda et al. (2006), in our experiment the total number of dots appearing in a given display ranged from 1 to 35, and each number was probed equally often. This necessarily meant that in half of the trials the probed set had to be the largest in the display. Although we made sure the probed set was not the largest in the other half of the trials, observers could nevertheless adopt the strategy of always encoding the largest set in the probe-after trials to improve task performance. To examine this possibility, we split the data into two groups according to whether the probed set was the largest or not. Across analyses, we found virtually identical patterns of results in the two groups. In the numerosity measures, there was no effect of the largest set probed nor did this interact with probe timing or set size (all $F_s > 2.83$, $p_s > 0.13$; see Figure 2B and C). In K capacity calculations, there was no difference at any set size (all $p_s > 0.2$; see Figure 4B and C). Our results were thus not inflated by observers

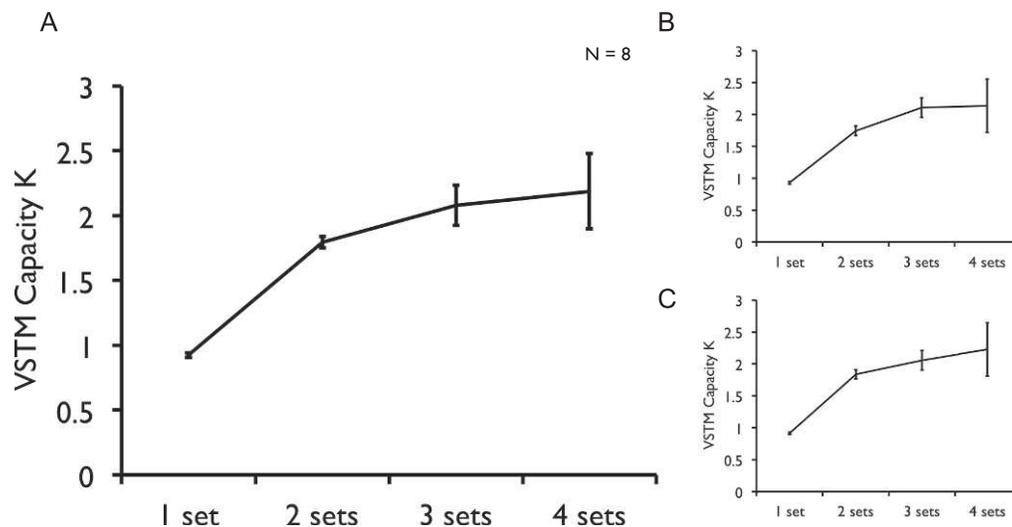


Figure 4. Experiment 1 VSTM capacity for color encoding of the probe-after condition. Cowan’s K s calculated from all trials were plotted in (A). Maximal K value was around 2, identical to the two-set limit of enumeration. Cowan’s K s were also calculated from trials in which the biggest set was probed (B) and in which it was not probed (C). There was no difference between (B) and (C), and both were similar to (A).

deliberately encoding the largest set to improve their performance.

Experiment 2

Results of Experiment 1 largely replicated Halberda et al. (2006): Approximately two sets can be enumerated in parallel. We also found that only two colors could be remembered while observers were performing the enumeration task, which was much lower than the three- or four-color capacity typically reported in previous VSTM studies (e.g., Alvarez & Cavanagh, 2004; Luck & Vogel, 1997). Although vivid colors were used, the enumeration displays used in Experiment 1 were quite different from typical VSTM displays used in past studies (e.g., Luck & Vogel, 1997). If VSTM capacity for color is still less than three when set enumeration is not required, it suggests that the dot displays used posed a severe limit on VSTM color encoding. The observed capacity for set enumeration therefore could simply reflect a VSTM capacity limitation for color encoding rather than a capacity limitation for enumeration per se. On the other hand, if VSTM capacity for color equates to or is greater than three when set enumeration is not required but falls under three when set enumeration is required, it would suggest a relationship between color memory and enumeration. It may be that color and numerosity are encoded together in the set enumeration task and that it is not possible to encode a color without also encoding its numerosity. In other words, each set functions as a single entry to VSTM with multiple features, much like

the object-based encoding benefit reported for VSTM for single objects (e.g., Luck & Vogel, 1997; Xu, 2002a, 2002b).

In this experiment, we repeated the design of Experiment 1, but instead of asking observers to report the numerosity of the probe color, we simply asked observers whether the probe color was present or absent in the display.

Methods

Eight new observers (two males) from the same subject pool participated in this experiment. This experiment was identical to Experiment 1 with both probe-before and probe-after conditions except that, instead of enumerating the number of dots present in the probed color, observers judged whether the probed color was present (by pressing “1”) or absent (by pressing “0”).

Results and discussion

To measure the number of colors that could be successfully retained from the briefly presented dot displays, only probe-after trials were examined (Figure 5). Cowan’s K capacities at one to four sets were 0.98 (.04), 1.82 (.08), 2.38 (.03), and 2.69 (.04). There was a significant increase in K measure between two and three sets, $t(7) = 5.22$, $p < 0.002$, and between three and four sets, $t(7) = 3.25$, $p < 0.02$. A t test against the null hypothesis that the maximum K capacity is equal to or

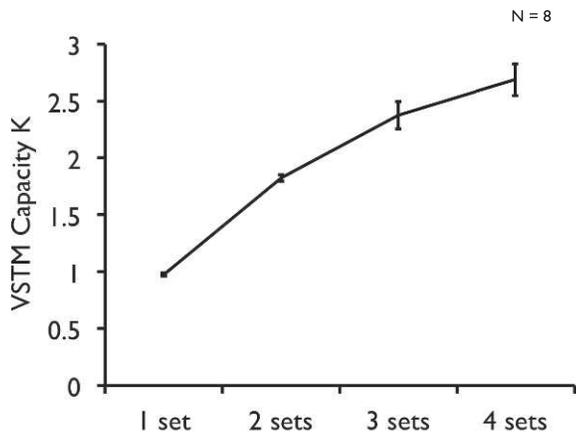


Figure 5. Experiment 2 VSTM capacity for color encoding of the probe-after condition. We repeated the design of Experiment 1, but instead of asking observers to report the numerosity of the probe color we simply asked observers whether the probe color was present or absent in the display. There was a significant increase in K measure between two and three sets and between three and four sets. These results show that when observers did not have to enumerate the sets, they could remember more than two but less than three colors from these dot displays.

greater than three was significant, $t(7) = 2.24$, $p < 0.03$. These results show that when observers did not have to enumerate the sets, they could still remember fewer than three colors from these dot displays.

These results suggest that the dot displays used here posed a severe limit on the encoding of colors into VSTM. The observed capacity for set enumeration could therefore simply reflect a VSTM capacity limitation for color encoding rather than a capacity limitation for enumeration.

Experiment 3

While the K capacity in Experiment 2 was nominally higher than that in Experiment 1, this difference could simply be caused by VSTM capacity differences between observers (e.g., Vogel & Machizawa, 2004), and VSTM capacity for color does not change whether or not the enumeration task is performed concurrently. Indeed, Feigenson (2008) have argued that numerosity is automatically encoded when an observer views sets or ensembles. This view would suggest that within the same observer, the capacity for colors, which is less than three, necessarily dictates that only two sets can be enumerated. However, it is also possible that concurrently performing the enumeration task reduces VSTM for color and that the perception of numerosity from multiple sets does not come free of cost. It could be that when

we removed the enumeration task and thus the necessity to encode all elements of a set, observers only needed to encode a few samples from each set to successfully perform the color VSTM task (see Myczek & Simons, 2008, in which observers adopted this strategy in a different task). This would result in more colors being successfully encoded into VSTM.

To test the above two possibilities, we conducted Experiment 3 to directly compare VSTM capacity for colors with and without the enumeration task in the same group of observers. Experiment 3 was comprised of the probe-after blocks of both the enumeration task from Experiment 1 and the color memory task from Experiment 2. As probe-before trials were uninformative in VSTM capacity measures, omitting those trials allowed us to run both tasks in a single one-hour session and minimize observer fatigue.

Methods

Participants

Eight new observers (two males) participated in this experiment. All observers were Harvard affiliates (students and researchers) with normal color vision and corrected visual acuity between the ages of 23 and 32 and received monetary compensation for participating.

Materials and design

The experiment was comprised of two blocks of 268 probe-after trials. In one block, which contained trials from the probe-after condition in Experiment 1, observers were instructed to enter the approximate number of dots in the probe color (the enumeration task). In the other block, which contained trials from the probe-after condition in Experiment 2, observers simply entered “0” or “1” to indicate whether the probe color was present in the displays (the color memory task). Order of the blocks was counterbalanced across observers.

Apparatus and procedure

The testing environment was identical to that in the previous two experiments. Trial sequence and timing were also replicated. Observers received verbal instruction and practice (up to 12 trials) on each task. When performing the numerosity task, it was again emphasized that the “0” key should be used when the observer believed the probe color to be absent from the displays; observers were discouraged from using the “0” key when they were simply uncertain about the numerosity. This experiment lasted approximately 55 minutes.

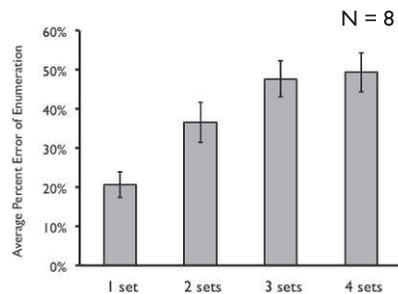


Figure 6. Experiment 3 average percentage error of enumeration in the numerosity task. Here, in all of the trials, the probe color appeared after the dot display. There was a significant rise in percentage error between one-set and two-set displays and between two-set and three-set displays. Error plateaued between three-set and four-set displays. This pattern of results replicated that of Experiment 1.

Results and discussion

Enumerating sets

As this experiment only included probe-after trials, we could not compare probe-before and probe-after performance to estimate the capacity of enumerating sets. Instead, we examined performance on probe-after trials in the numerosity task for each set size (Figure 6). A correction for the error in the “miss” trials was applied in same way as in Experiment 1, and a total of 11.4% of the color-present trials were corrected. The average percentage error at each set size for the enumeration task is shown in Figure 5. We saw a significant increase in error between one and two sets, $t(7) = 5.11$, $p < 0.002$, and between two and three sets, $t(7) = 3.86$, $p < 0.007$; performance plateaued after three sets, $t(7) = .67$, $p = 0.52$. This pattern of results replicated the pattern seen in Experiment 1, suggesting

that omitting the probe-before trials did not parametrically alter the task.

VSTM capacity for set color

As in Experiment 3, when observers performed the numerosity task the rate of misses (Figure 7C) was fairly low at smaller set sizes although we see a significant increase between one- and two-set display errors, $t(7) = 3.20$, $p < 0.02$. The rate of misses more than triples between two and three sets, $t(7) = 4.25$, $p < 0.004$, but does not increase beyond that, $t(7) = .12$, $p = 0.90$. Cowan’s K estimates for the two tasks are presented in Figure 7A. Capacity K values at one to four sets for the numerosity task were .978 (.02), 1.73 (.15), 2.20 (.45), and 2.60 (.59). For the color memory task, K values at each set size were .978 (.025), 1.79 (.14), 2.32 (.38), and 2.53 (.57). Paired t tests found no difference in capacity K between the tasks at each of the set sizes, all $t_s(7) < 1.45$, $p_s > 0.217$. We also correlated the two capacities within an observer, using the K values obtained at a set size of four (see Figure 7B). Pearson’s correlation coefficient between the two capacities was .863 ($p < 0.007$). Thus, even though observers reported that the enumeration task seemed much more difficult than the color VSTM task, the addition of the enumeration task did not significantly impact VSTM for set color, which was maximally around 2.57 in both tasks.

In all, our results suggest that enumeration capacity as it has been measured by Halberda et al. (2006) is necessarily limited by color VSTM capacity. We consistently found that the average K capacity for color with these displays was below three. This was manifested in the much greater number of trials containing more than two sets in which observers failed to encode the color of the probed set. When one

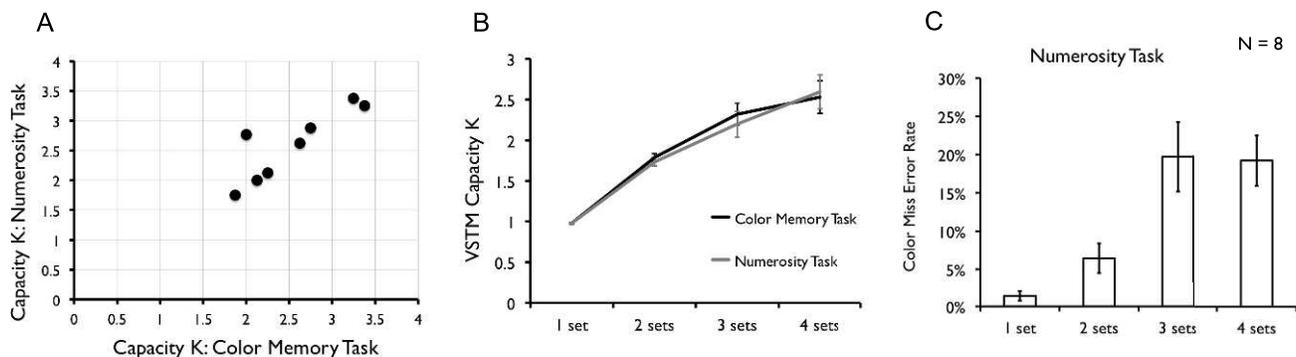


Figure 7. Experiment 3 Cowan’s capacity K estimates for set colors with and without the enumeration task being performed concurrently within the same group of observers. (A) The average capacity K at each set size. There was no difference between the K measures for set colors in the numerosity task and the color memory–only task. It appeared that the addition of the numerosity task did not decrease VSTM capacity for set colors. (B) The correlation between the two K measures, which were highly correlated ($R = .86$). (C) The average rate of miss errors in the numerosity task block at each set size.

enumerates multiple color-defined sets simultaneously in this design, it is critical to hold the colored sets in memory before making a response. If less than three colors can be held in VSTM, then it is not possible that more than two sets can be enumerated after the masked delay. These results provided further evidence that retaining set color in VSTM is a prerequisite for successful set enumeration such that VSTM capacity for set color determines the upper limit of the enumeration capacity but that the addition of the enumeration task has minimal impact on VSTM capacity for set color.

General discussion

Halberda et al. (2006) reported that observers could reliably retain and enumerate approximately three sets of dots at a glance: two colored sets and the superset of all dots on the screen. With the brief display presentation duration and the masking used in Halberda et al., the task taxes heavily the representation capacity of VSTM. Because observers need to encode and retain the colors of the different sets in VSTM before they can select and enumerate all the dots in each set, it is possible that VSTM capacity limit for colors, rather than the capacity limit of set enumeration per se, determines how many sets observers can enumerate successfully from a given display. In other words, if observers cannot represent a particular color in VSTM, it would not be possible for them to enumerate the dots appearing in that color. On the other hand, given that previous literature using highly salient colors has reported VSTM capacity to be about three to four colors (Alvarez & Cavanagh, 2004; Luck & Vogel, 1997; Pashler, 1988; Vogel et al., 2001), it is possible that the colors of three or four subsets could be retained but that only two of these subsets could be enumerated. In other words, although enumeration depends on successful VSTM retention for set color, enumeration capacity could still be dissociable from VSTM capacity for set color. Thus, to understand what determines the limitation in set enumeration, it is important to know the association between the capacity for set enumeration and VSTM capacity for set colors.

In Experiment 1, we found that observers could more or less enumerate two sets of dots, largely replicating the results of Halberda et al. (2006). Interestingly, even with the vivid and highly discriminable colors that we used, observers could retain only about two colors from the same displays while performing the enumeration task, which is lower than the three- or four-color capacity typically reported in previous VSTM studies (e.g., Alvarez & Cavanagh,

2004; Luck & Vogel, 1997). In Experiment 2, in a different group of observers using the same displays and a similar procedure but without the enumeration task, we found that VSTM capacity for color from these dot displays still to be less than three. In Experiment 3, we directly compared VSTM capacities for colors measured with and without the enumeration task within the same group of observers. We found that capacities for color did not differ between the two tasks and were highly correlated within observers. These results suggest that VSTM capacity for color is not affected by the addition of the enumeration task and that the number of colors and the number of sets that can be remembered are both less than three.

The less-than-three-item VSTM capacity for colors thus makes it impossible for observers to encode more than two sets of dots as successful enumeration of three sets requires the encoding of at least three colors. Consistent with the VSTM finding, in both Experiments 1 and 2, we found that the number of trials in which observers failed to encode the probed color set in the probe-after condition substantially increased in displays containing more than two sets. A single, serial process by which sets' colors are remembered before they can be enumerated may explain why capacity for set enumeration cannot exceed two in both Experiment 1 and in Halberda et al. (2006). This is consistent with work by Feigenson (2008), who also showed that enumeration is constrained by limits of memory rather than attention. These results leave open the intriguing possibility that, if VSTM color encoding capacity from the dot displays could be increased to more than three, the numerosity of more than two sets of dots may be retained. In fact, when Feigenson presented temporally interleaved sets with highly discriminable object features (e.g., toy pigs, candies), she reported enumeration capacity to be at three sets, greater than the two-set capacity reported here and in Halberda et al.

In our study, the encoding of dot color necessarily precedes the enumeration of all dots in a particular color. Moreover, once encoded, the dots of a particular color must remain in memory until the observer makes his or her enumeration response, engaging VSTM mechanisms. Because sensory memory is virtually limitless (Sperling, 1963), it is likely that the capacity bottleneck for both color and enumeration occurs at the point of information transfer between sensory memory and VSTM. While it may be suggested that VSTM capacities for both color and numerosity are independently constrained by a sensory feature of the multi-dot stimuli—such as crowding, lateral masking, or the unpredictability of dot size and location—this is unlikely given the serial processing that our data suggested. As we have described, enumeration in our experiments is an inherently serial process: Items in a set must be encoded and retained in VSTM before they

can be enumerated. Thus, sensory features of the displays would first affect item encoding into VSTM, reducing VSTM capacity for these items, which would then, in turn, influence the enumeration capacity. This would explain why the addition of the enumeration task had minimal impact on VSTM capacity for set color.

While we cannot completely rule out with the present data that a sensory feature of the stimulus set has produced these seemingly yoked capacities coincidentally, recent work in our lab using a similar paradigm (Poltoratski & Xu, under review) has investigated subset enumeration across manipulations of the displays. We found no change in enumeration capacity when the subsets were presented sequentially in individual displays nor when irrelevant distractor sets were added. These results provide some evidence that sensory features of the display alone do not determine the capacity limit for enumeration.

In Experiment 1, without probing the superset of all dots, capacity for enumeration was two sets, similar to when the superset was probed in Halberda et al. (2006). This suggests that either the encoding of the superset is automatic regardless of whether it is explicitly probed or not or that resources dedicated to encoding the superset cannot be used to encode an additional set in the enumeration task, possibly due to the fact that the encoding of the superset does not require colors to be encoded and selected first. In another study (Poltoratski & Xu, under review), we found that manipulations of bottom-up and top-down attention did not affect the enumeration of the individual subsets but significantly impacted the enumeration of the superset. Specifically, neither presenting the subsets in sequential, individual displays nor adding irrelevant distractor dots impacted the enumeration of the subsets. However, in both of these conditions, observers could no longer automatically enumerate the superset of all dots. This work also shows that changing the physical parameters of the displays, such as crowding and masking, has no effect on subset enumeration capacity.

In summary, the present studies show that while observers can enumerate multiple spatially overlapping sets, VSTM capacity for set color can pose an upper capacity limit for set enumeration.

Keywords: visual short-term memory, numerosity, psychophysics, color

Acknowledgments

This study was supported by NSF grant 0855112 and NIH grant 1R01EY022355 to Y. X. We thank members of the Harvard Vision Sciences Laboratory for their generous support in this research and Dr.

Mike Pratte for invaluable discussions at Vanderbilt University. Authors declare no competing financial interests.

Commercial relationships: none.

Corresponding author: Sonia A. Poltoratski.

Email: sonia.poltoratski@vanderbilt.edu.

Address: Department of Psychology, Vanderbilt University, Nashville, TN, USA.

Footnote

¹ In an effort to make our results comparable to those of Halberda et al. (2006), corrections for multiple comparisons were not performed here as they were not done in that study. The overall results reported herein did not change markedly if corrections for multiple comparisons were performed.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*(4), 698–707.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.
- Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361.
- Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition*, *107*(1), 1–18.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.
- Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576.

- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378.
- Poltoratski, S. A., & Xu, Y. (Under review). Enumeration of subsets but not the superset withstands temporal and attention manipulations.
- Sperling, G. (1963). A model for visual memory tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*(1), 19–31.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology Human Perception and Performance*, *27*(1), 92–114.
- Xu, Y. (2002a). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology-Human Perception and Performance*, *28*(2), 458–468.
- Xu, Y. (2002b). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, *64*(8), 1260–1280.