The flicker paradigm provides converging evidence for a 3-item limit of visual working memory

Justin Halberda
Johns Hopkins University

Daniel J. Simons
University of Illinois

Jeffrey Wetherhold
Harvard University

Corresponding author:

Justin Halberda
Department of Psychological and Brain Sciences
Johns Hopkins University
Ames Hall
3400 North Charles Street
Baltimore, MD 21218 USA

Email: Halberda@jhu.edu
Phone: 011.33.1.44.32.23.62
Fax: 011.33.1.44.32.23.60

Contact information for Simons & Wetherhold:
Daniel J. Simons: dsimons@uiuc.edu
Jeffrey Wetherhold: jeff@wjh.harvard.edu

Key Words: Gestalt, VSTM, Capacity, Memory, Similarity, Proximity, Change detection, Top-Down

Abstract

Visual working memory appears to have a capacity of 3-4 objects, with additional limits imposed by the precision of the information that must be represented from each object. Most studies of VWM have used a “one-shot” change detection task, with capacity inferred from the success with which people detect changes to features of objects. Most have used small numbers of objects with distinctive, easily categorized features, limiting their generalizability to more subtly different stimuli or to larger numbers of objects in the scene. One-shot change detection tasks also do not assess how VWM operates over extended viewing. The flicker change detection task permits measurement of VWM capacity over time and with larger numbers of objects present in the scene, but it has been relatively underused in the literature. And, when it has been used, estimates of VWM capacity with easily discriminable objects have exceeded the typical 3-item limit. Here we use the flicker task to examine the following questions: (a) Would capacity in the flicker task be closer to the typical 3-4 items with more subtly different stimuli? (b) What is the most appropriate dependent measure in the flicker task when measuring the capacity of VWM: response time or number of changes viewed? (c) How would estimates of the capacity of VWM change for displays with far more objects than the capacity of VWM? And (d) How does VWM operate over time, with repeated opportunities to encode, retain, and compare elements in a display? Three experiments using grids of simple items varying only in luminance revealed a VWM capacity limit of approximately 3 objects that was largely impervious to changes in display on-time, off-time, and array size. This estimate converges with those from other paradigms.

Our experience of a richly-textured visual scene does not necessarily require a complete, detailed internal representation, and our awareness of the scene contents is at least partially constrained by limits on attention. For example, human adults (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999), infants (Feigenson, Carey & Hauser, 2002; Feigenson & Carey, 2003), and non-human primates (Hauser, Carey & Hauser, 2000; Hauser & Carey, 2003) attend to and remember no more than 3-4 items in parallel. Because of this limit, almost any act of seeing over time requires visual working memory (VWM) to maintain relevant information in a semi-durable store that can be compared or integrated over time (Irwin & Gordon, 1998). In essence, VWM links the present to the immediate past, serving as a bridge between online visual processing, long-term memory, and higher cognition.

VWM capacity appears limited to 3-4 unified object representations at one time (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974; Pylyshyn & Storm, 1988; Scholl & Xu, 2001; Sperling, 1960; Vogel, Woodman & Luck, 2001), though VWM may be limited to fewer than 3-4 items for visually complex or difficult to distinguish objects (Alvarez & Cavanagh, 2003), suggesting there is an information-based limit in addition to an item-based limit. Most current VWM studies use a one-shot change detection paradigm in which subjects try to detect a change to one item in an array of several items. Often, each item has multiple features that could change, but only one feature does, thereby allowing an assessment of whether or not the units of VWM are integrated objects — if subjects can detect a change to any feature of an object as well as they can detect changes to individual features in isolation, then the features might be stored bound to objects. In general, performance is nearly perfect when the display contains three or fewer simple objects. Performance declines steadily once the display contains four or more objects. This transition from accurate to degraded performance is thought to reflect an object-based limit of VWM.

The one-shot change detection task has several drawbacks as a measure of the functional capacity of VWM. First, interaction with real-world scenes always involves more than 4 simple objects and always involves viewing that unfolds over time, not a single flash. To understand the operation of VWM, it is essential to examine how items are swapped into and out of VWM and how they are used while in VWM. With only one exposure to the change and generally poor change detection performance with large numbers of objects, the one-shot task is poorly suited to examining how subjects use VWM over time when inspecting more complex displays.

An alternative change detection method — the flicker task (Rensink et al, 1997) — has mostly been used to document the phenomenon of change blindness. In this task, an original and changed version of an image alternate repeatedly until observers find the change. Provided the two images are separated by brief blank screens, people take a surprisingly long time to find the changing item, even when the changes are large and easily seen once they have been detected (Rensink et al, 1997). Despite its extensive use in the change blindness literature, with one notable exception (Rensink, 2000a), this paradigm has been less widely used to assess the capacity limit in VWM. Yet, it has several advantages (with some drawbacks as well) over one-shot tasks for measuring the functional limits of VWM. Specifically, it allows extended search through displays with many objects while still providing a useful dependent measure of change detection performance (RT).

Rensink (2000a) first used the flicker task to measure the capacity for detecting changes to the orientation or polarity of one object in an array of simple shapes. In principle, increasing display times should allow people to encode more information into VWM, provided that they have "slots" available in memory. Once VWM is filled to capacity, though, providing additional display time for encoding will not improve change detection performance. Using a measure derived from the asymptote of change detection performance as a function of display time, Rensink estimated the capacity of VWM for orientation to be approximately 5.5 objects. The function for polarity never asymptoted with the display times used, and capacity estimates of at least 8-9 items were implied. Both estimates are significantly higher than the more typical estimate of 3-4 objects in VWM, raising concerns that the task might not provide an ideal measure of VWM. Also, in Rensink's experiment, the capacity was inferred from a between-subjects comparison and from only one data point after the asymptote, making the inference less robust. The discrepancy between the capacity estimates in these studies might, however, have more to do with the nature of the stimuli than the method itself. First, Rensink's flicker experiments used only two levels for orientation (vertical, horizontal) and polarity (black, white), and subjects might be able to economize in storage by relying on verbal memory or longer-term representations. Second, subjects might store perceptual groups or clusters of objects in a single VWM "slot," artificially inflating estimates of the capacity of the store itself (see Rensink, 2000a for discussion of these issues).

In 3 experiments, we use the flicker task to estimate the capacity of VWM. Our studies are motivated by the need to confirm the 3-item limit using alternative paradigms to the typical one-shot technique. We refine Rensink's approach by using a large numbers of objects that are visually simple, not easily verbally encoded, and that do not easily combine into perceptual groups. We also used within-subjects comparisons with enough display times to clearly show an asymptote. In doing so, we address the following questions: (a) is capacity in the flicker task closer to the typical 3-4 items for stimuli that are not easily categorized or grouped? (b) What is the most appropriate dependent measure in the flicker task when measuring the capacity of VWM: response time or number of changes viewed? In most flicker task experiments, these measures are treated as equivalent, but they might not be when using the flicker task to measure capacity. (c) How might estimates of the capacity of VWM change for displays with far more objects than the capacity of VWM? And (d) how does VWM operate over time, with repeated opportunities to encode, retain, and compare elements in a complex display?

Experiment 1

Experiment 1 measured the capacity of VWM using the flicker task by systematically varying display on-time. If the capacity of VWM is limited, then with a sufficient display time, participants should completely fill this available buffer and change detection performance should asymptote, thereby providing an estimate of the number of items being held in VWM over the course of each alternation (Rensink, 2000a). Unlike in Rensink's earlier studies, we used arrays of 36 grayscale dots that were not readily categorized or grouped.

Method

Participants

Thirteen college students who reported normal or corrected-to-normal vision participated in exchange for course credit.

Displays and procedure

All participants were tested using Macintosh iMac computers (CRT monitors with viewable area measuring 29.5 X 22.5 cm). Viewing distance was unconstrained, but averaged approximately 50 cm. (All visual angle measurements below are based on a 50cm viewing distance.) On each trial, participants viewed an original and modified array of 36 grayscale dots arranged into a 6x6 grid against a black background (see Figure 1). The diameter of each dot subtended approximately 1 degree of visual angle and the centers of adjacent dots were separated by 2.66 degrees. Each dot was randomly assigned one of 30 equally-spaced luminance values ranging from 4.65 to 61.51 cd/m2 (possible values ranged from 55 to 200 in 5 volt intervals).

Participants viewed the first array of dots, followed by a 120ms black screen and then by a second array that was identical to the first except that one dot changed luminance by ±14 shades (70 volts). For this magnitude of change, most items could either increase or decrease in luminance within the allowable range, but not both. For those cases in which the changed item was capable of becoming either lighter or darker, the direction of change was selected at random for that trial. After another 120ms black screen, the sequence repeated until participants detected the change. The display duration varied across 5 blocks of trials for each participant (300, 500, 700, 900, or 1100ms), with 20 trials per block. Block order was randomized for each subject. Within a trial, the display duration was the same for both the original and changed array. The blank screen durations were fixed at 120ms for all blocks. When participants detected the change, they pressed the space bar to stop the alternation and then used the mouse to click on the changing dot. Response times were recorded from the onset of the first display until the participant pressed the space bar. Participants were instructed to respond as quickly as possible while prioritizing accuracy.

INSERT FIGURE 1 ABOUT HERE

Results and Discussion

All trials for which participants selected the wrong item were eliminated from further analyses (5.3% of trials). To eliminate unusually long response times that might have resulted from factors unrelated to the task, we removed any trials for which response latency was more than 2SDs above a participant's mean (3.7% of trials) for that display duration, with latency measured as the time from trial onset until the participant pressed the space bar. Thus, latency includes both display on-time and blank screens. Across all display durations, participants detected changes after about 12.8s (STDEV = 8.5s) on average (Figure 2a).

INSERT FIGURE 2 ABOUT HERE

Latency to find the changing item was constant for display durations from 300-700ms (Figure 2a); none of these conditions differed in paired-samples t-tests (all t values < .3, all p values > .8). Giving linearly more time allowed participants to store linearly more information, leading to a constant RT to find a change (see Rensink, 2000a for a similar result). Thus, for these display durations, change detection was process-limited — participants did not have enough time during each display to complete all mental operations involved in the task and load VWM to full capacity. Consequently, giving more time leads to additional items in memory —limitations are due not to the capacity of VWM but to the amount of time subjects are given to load information into VWM.

Interestingly, the constant change detection latencies across this range means that shorter display durations (e.g. 300ms) provided more exposures to the change than longer durations (e.g.

700ms). For example, if change detection took 14000ms total, observers would have had 32 exposures to the change with a display duration of 300ms (i.e. = 32 * (300ms+120ms) + 420ms), but only 16 exposures with a display duration of 700ms. Figure 2b plots the number of alternations (i.e. display+ISI) needed for change detection for each display duration. For display durations ranging from 300-700ms, the overall encoding time rather than the number of change exposures determined when participants detected the change.

In contrast, for longer durations (700-1100ms), response latency increased linearly as a function of display duration (Figure 2a), but the number of alternations needed for change detection remained constant at approximately 15 cycles (Figure 2b). Thus, for display durations of 700-1100ms, performance appears to be capacity-limited. After looking at a display for 700ms, participants have filled VWM to capacity and longer display durations (900 and 1000ms) did not enhance change detection performance. In other words, an additional 400ms of viewing for each display did not allow participants to encode any additional information that could be held across the blank-screen ISI and used to aid change detection. Shifting attention or the eyes to a new set of objects or scrutinizing the display in more detail during that additional time was of no benefit because, presumably, no more information could be held in VWM and used following the blank-screen ISI.

The finding that response latency and the number of change exposures differ as measures of change detection performance is surprising given that these two measures often are treated as equivalent. In fact, the results of Experiment 1 suggest that these two measures are rarely comparable, at least when used to measure the capacity of VWM. For display durations of 300-700ms, response time was constant while the number of cycles required to detect a change steadily decreased. In contrast, for display durations of 700-1100ms, the number of cycles required was constant while response time steadily increased. Thus, the number of cycles and the total response time are only comparable when the display duration gives participants just enough time to load

VWM to capacity but no more. For our displays and task, that critical duration was approximately 700ms (see Rensink 2000a for a similar estimate).

Both response latency and the number of alternations required for detection allow us to derive an estimate of the capacity of VWM. Because information about the display must be stored in VWM over the course of each ISI (blank screen), the capacity of VWM will place an upper-bound on performance in the flicker task. By adding additional display time, other factors such as crowding in the display, possible saccades made during encoding, and the time required to decide that no change has been detected can be eliminated as limiting factors in our task. Once sufficient time has been added, these processes should reach completion and additional display time should no longer improve performance because no additional information can be held in VWM across the ISI. One concern about this task is that capacity limits might reflect a capacity-limited comparison process or limits on retention over the course of an ISI. Note, though, that <u>any</u> change detection task is subject to the same criticism, so this concern does not favor the one-shot task over the flicker task as a measure of capacity provided that the display times are long enough to reach an asymptote in performance.

<u>Estimating capacity.</u>

We used a method similar to Rensink (2000a) to derive an estimate of the capacity of VWM from RT. Optimal performance in the flicker task should be attained when, within a single display on-time, participants compare information about some number of items on the screen to information stored about those items in VWM, reach a decision that no change had occurred for those items, move attention to a new subset of items <u>without replacement</u>, and encode and consolidate information from the new set of items into VWM in preparation for the next display[1].

[1] Rensink 2000a suggests that search for a changing item is random without replacement, whereas Horowitz & Wolfe (1998) suggest that more traditional visual search for a target among distractors is random <u>with</u> replacement. Assuming there is some forgetting during the ISI, the resulting search in the flicker task is likely somewhere between accurate search without replacement and search with replacement. Any component of search with replacement would result in slightly higher (up to double) capacity estimates than we provide here, so the present estimates based on without replacement search can be treated as a lower bound.

Following Rensink (2000a), response latencies on each trial were converted to an estimate of capacity, and these estimates were used to calculate the mean capacity for each subject at each display duration. This capacity estimate reflects the number of items that subjects could hold in memory across the blank interval and compare to the subsequent display. Response times are a function of the total number of dots in an array (array size), display duration, and capacity. If search for the target is random without replacement (Rensink, 2000a), on average a subject will have to search (Array Size + 1)/2 individual dots before finding the target. Accordingly, the response latency is the amount of time it would take to search approximately half of the items in the array divided by the number of items that can be encoded, held, and compared from a single display (capacity) plus some constant:[2,3]

A) $$\text{RT} = \frac{[(\text{array size} + 1)/2] * (\text{display duration} + \text{blank duration})}{\text{capacity}} + \text{constant}$$

The constant in this equation must be at least equal to the on-time + ISI because no change could be detected during the first display and subsequent ISI. The constant also would increase as a result of any non-search-related contributions to response latency and due to any time spent verifying the presence of a change after detection. Because our procedure does not measure these factors, we assume for simplicity that the constant is equivalent to the first display duration + blank duration. From Equation A we can derive an estimate of capacity:

B) $$\text{Capacity} = \frac{[(\text{array size} + 1)/2] * (\text{display duration} + \text{blank duration})}{\text{RT} - (\text{display duration} + \text{blank duration})}$$

We computed a capacity for each trial for each subject, and Figure 2c displays the mean of subject means. Capacity increased for display durations up to 700ms, at which point it asymptoted at approximately 2.5 – 3 items (Figure 2c). Additional display time did not change the amount of

---

2 This equation is a more general form of the equation provided by Rensink (2000a): search slope = (alternation time)/hold. Similarly, our Equation B is a more general form of Rensink's hold = (alternation time)/(search slope).

3 George Alvarez notes that this equation fails to capture the intuition that when capacity equals array size participants should always find the changing target after a single cycle. To capture this intuition, Alvarez suggests a modification of our Equation A in which the number of cycles required before change detection = [(array size/capacity) + 1] / 2. This modification returns estimates of capacity that are not significantly different from the present Equation A over a wide range of # of cycles required, so we will use Equation A.

information that could be encoded and compared from one display to the next; performance was time-limited until 700ms but capacity-limited with longer display times. These effects were confirmed by a significant effect of display duration in a repeated measures ANOVA, $F(4,48) = 21.60$, $p < .001$. Planned t-tests revealed greater capacity for 500ms display durations than for 300ms display durations ($t(12) = 4.04$, $p < .01$) and greater capacity for 700ms than for 500ms display durations ($t(12) = 4.01$, $p < .01$) with no significant differences among display durations ranging from 700-1100ms (700 vs. 900ms, $t(12) = -.28$ $p = .785$; 900 vs. 1100ms, $t(12) = 1.41$ $p = .183$).

That estimated capacity asymptoted at a display duration of 700ms makes theoretical sense given results from other paradigms. If it takes approximately 150ms to load information from three items into VWM (Woodman & Vogel, 2005; Vogel, Woodman, & Luck, in press) then the additional on-time of 550ms (i.e. 700ms – 150ms) during each display would be used by participants to compare information stored in VWM to information presently on the screen, reach a decision that no change has occurred, empty the contents of VWM, and move attention to three new items. That is, optimal performance in the flicker task would be to perform all of these operations during a single display on-time and 700ms seems a reasonable estimate of the time required to perform each of these operations: checking, moving attention, and loading memory to capacity during each on-time. That we found no change in performance with display times of 700-1100ms suggests that 700ms is enough time for these operations to complete satisfactorily.

Experiment 2

Experiment 1 produced an estimate of capacity (between 2.5 and 3 items) similar to that derived from performance on one-shot change detection tasks despite substantial differences in the measures and method. Our estimate with simple, hard to categorize stimuli was also substantially lower than the estimated capacity for orientation and polarity from Rensink's (2000a) earlier studies using the same method. Despite the similarity of our estimate to that of the one-shot display, several concerns remain about the validity of estimates derived from the flicker task: 1) with larger number of objects in each display, crowding might disrupt encoding, 2) our larger displays might require participants to saccade during each display on-time to encode the stimuli, 3) the use of a briefer blank-screen ISI than most one-shot tasks used to measure capacity might have disrupted the consolidation of information in VWM (Vogel et al, in press), 4) limitations on performance could be due to limits on the ability to compare the contents of VWM to the current display rather than to limits on VWM itself.

The first two concerns can be eliminated by increasing display on-time. As long as the change is detectable, increasing on-time permits as many shifts of attention or eye movements as needed within a single on-time. Experiment 2 used on-times as long as 1900ms, comparable to the longest (memory-test) portion of a typical single-cycle change detection paradigm (e.g. Luck & Vogel, 1997). If performance is comparable between 700ms and 1900ms then saccades and crowding cannot be the factors limiting performance in our task any more than they are for other VWM methods (likewise for any other limitation that would be ameliorated through additional display on-time).

The third concern is based on the evidence that consolidation of information into a durable form in VWM requires approximately 50ms per item for simple items in a one-shot change detection task (Vogel et al, in press). If so, then the limiting factor in Experiment 1 could have been the 120ms ISI. Experiment 2 used a 900ms ISI — the same as in most one-shot VWM tasks — to test whether consolidation was a limiting factor in Experiment 1.

The fourth concern, that the limit might be on comparison processes rather than VWM, is not easily eliminated. But, it is as much a concern for the one-shot task as the flicker task provided that the display on-times are comparable across the two paradigms. Successful change detection in the absence of a visible, low-level change signal requires a storage process and a comparison process (Simons & Rensink, 2005). Thus, whenever successful change detection is used to measure memory capacity, the estimates could reflect limits in storage or in the comparison process itself. Neuroimaging methods might be able to separate these two components (Xu & Chun, 2006).

In sum, estimates of capacity from the flicker task should be as valid as measures of capacity from a one-shot change detection task, provided that display times and ISIs are comparable. Empirically, that our Experiment 1 returned an estimate of approximately 2.6 items in VWM, convergent with estimates from one-shot tasks, adds credence to this claim.

Method

Participants

A new group of thirteen Johns Hopkins University undergraduates who reported normal or corrected-to-normal vision participated in exchange for course credit.

Displays and procedure

All procedures were identical to Experiment 1 except that the ISI was fixed at 900ms and display durations had values of 1100, 1300, 1500, 1700 and 1900ms, presented in 5 randomized blocks.

Results and Discussion

All trials for which participants selected the wrong item were eliminated from further analyses (1.4% of trials). To eliminate unusually long response times that might have resulted from factors unrelated to the task, we further removed any trials for which response latency was more than 2 SDs above a participant's mean (4.5% of trials).

Averaged across all display durations, participants detected changes after about 24.5s (SD = 5.2s). As display duration increased from 1100ms to 1900ms, the capacity estimate of VWM remained relatively constant between 3-4 items, suggesting that saccades, crowding and decision time do not limit performance. Longer display durations (1100-1900ms) did not lead to faster change detection as the number of cycles required to detect a change remained constant at approximately 12 (Figure 3b). Response latency increased slightly as a function of display on-time (Figure 3a).

INSERT FIGURE 3 ABOUT HERE

Capacity estimates varied between 3-4 items, resulting in a significant effect of display on-time on capacity, $F(4,48) = 3.71$, $p < .01$ driven mostly by the lower capacity observed (3.1 items) at display on-time = 1300ms. The average capacity estimate across all display durations was 3.6 items in VWM (SD = 1.1). This estimate is higher than the asymptotic performance in Experiment 1 (2.6 items), and the difference could be due either to the increased ISI (consolidation time) or the increased on-time (e.g. by perceptual grouping). Taken together, a VWM capacity estimate of 2.6-3.6 items is consistent with estimates from other paradigms (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974; Sperling, 1960; Vogel, Woodman & Luck, 2001).

INSERT FIGURE 4 ABOUT HERE

Experiment 3

If the increase in capacity to 3.6 items obtained in Experiment 2 was due to increasing the ISI from 120ms to 900ms, allowing for greater consolidation time, then a display duration of 700ms with an ISI of 900ms should also return a capacity estimate of approximately 3.6 items. If instead it was the longer display on-times in Experiment 2 that lead to the increase in capacity, perhaps via increased perceptual grouping, then a display duration of 700ms with an ISI of 900ms should return a capacity estimate of approximately 2.6 items. In Experiment 3 we explore the capacity of VWM with these parameters while also examining whether the number of objects in the array affects estimates of capacity. One-shot tasks typically involve arrays of 4 or fewer objects. In Experiment 3, we varied the number of display items (array sizes of 4, 9, 16, 25, or 36 items). If change detection capacity is limited to 3 objects per display on-time, then manipulating the number of items in the array should not affect the capacity estimate.

Method

The materials and procedures for Experiment 3 were identical to those of Experiment 1 except that: (a) all trials were presented with a display duration of 700ms and an ISI of 900ms, and (b) participants completed 5 randomly-ordered blocks of 20 trials each, one with each array set size (4, 9, 16, 25, or 36 items). The dots in each display were presented in a square grid, centered on the display, with constant inter-dot distances. Participants were 13 Johns Hopkins undergraduates who received course credit for participating.

Results and Discussion

All trials for which participants selected the wrong item (11.0% of trials) and all trials with response times more than 2 SDs above a participant's mean (3.1% of trials) were eliminated from further analyses[4]. Across all array set sizes, participants detected changes after about 12.4s (SD = 8.1s) on average. Response latency on each trial was transformed into a capacity estimate using Equation B, and participant means were calculated for each array set size. The average estimated object-based capacity limit across all set sizes was 2.7 items (SD = 1.4)(Figure 5), and the number of items in the array had no effect on capacity estimates; a regression analysis of capacity and array size revealed no significant effect of array size (the 95% confidence interval for the slope ranged from -.031 to .03, $F(1, 63) = 0.000$, $p = .983$). This finding suggests that the difference between Experiments 1 and 2 derives not from the ISI duration but from the longer display on-times used throughout Experiment 2. Experiment 3 used a longer ISI just like Experiment 2, but produced capacity estimates closer to those of Experiment 1. The longer display on-times used in Experiment 2 (e.g. 1900ms) could engender higher capacity estimates (e.g. 3.6 items) by promoting

[4] Two participants had an unusually large number of wrong responses, which appeared to be from lack of effort. Results were not significantly different if these two participants were removed from the sample.

perceptual chunking of the display. Any effects of grouping or chunking were fairly minimal given that the range of capacity estimates across all three experiments was between 2.6 and 3.6 items.

General Discussion

The present experiments revealed a capacity limit for change detection of approximately three visual items using a flicker task with visually simple, hard to categorize stimuli. The experiments systematically varied the display duration and the number of items in the display, revealing an asymptotic limit of approximately 2.6-3.6 items in VWM over the course of each ISI. This result agrees with estimates of the capacity limit of VWM from other paradigms (e.g. one-shot change detection). It also extends such results to cases in which a scene contains many more than 3-4 items and in which viewing unfolds over time. The results also constrain earlier work with the flicker paradigm that had suggested larger capacities for object orientation and polarity. Our study, using a within-subjects design to more precisely measure the asymptote of performance as a function of display duration and also using less qualitatively different stimulus categories, suggested that capacity limits for orientation and polarity likely were enhanced by grouping or other chunking processes (see Rensink, 2000a for a discussion of this possibility in his data).

We also found that change detection performance was process-limited (response time) with display durations shorter than 700ms and memory-limited (# of exposures to the change) with display durations longer than 700ms (see also Rensink 2000a). This finding shows that response time and the number of changes viewed are not, as has been commonly assumed, equivalent estimates of change detection performance.

Some recent evidence suggests that the capacity of VWM sometimes is <u>fewer</u> than 3 items (Alvarez & Cavanagh, 2003), especially when the stimuli used are more subtly different from each other or when the changes are small. That is, VWM might have a maximum capacity of 3 items, but it might also be limited by the amount of information that can be encoded (Xu & Chun, 2006). With categorically encodeable and easily distinguished features (e.g., red vs. blue), performance can reach a maximum capacity of approximately 3 items (Luck & Vogel, 1997). However, with complicated stimuli requiring fine discriminations, the information limit will be reached before the 3-item limit (Alvarez & Cavanagh, 2003). The present experiments used a wide range of hard to categorize grayscale dots and large, easily detected changes in luminance (i.e. ±50% within the range of possible luminance values), yielding an estimated capacity limit of approximately 3 items in VWM. This result suggests that it is not categorizability of the features *per se* that determines capacity. Rather, VWM appears limited to storing information from approximately 3 items, irrespective of the categorizability of their features, so long as the total feature information does not exceed the information-based limit of VWM.

References

Alvarez, G. A. & Cavanagh, P. (2003). The capacity of visual short-term memory is set by total information load and by number of objects. *Article in press for* Psychological Science.

Feigenson, L. & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. Developmental Science, 6:5, 568-584

Feigenson, L., Carey, S. & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. Psychological Science,13 (2), 150-156.

Hauser, M. D., Carey, S. & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. Proceedings of the Royal Society of London: Biological Sciences, 267, 829-833.

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macques: Examinations of content and format. *Article in press for* Cogntive Psychology.

Horowitz, T. S. & Wolfe, J. M. (1998). Visual search has no memory. Nature, 394, 575-577.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. Visual Cognition, 5(1-2), 127-155.

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. Nature, 390, 279-281.

Pashler, H. (1988). Familiarity and visual change detection. Perception and Psychophysics, 44, 376-378.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. Perception & Psychophysics, 16(2), 283-290.

Pylyshyn, Z. W. & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial Vision, 3, 179-197.

Rensink, R. A. (2000a). Visual search for change: A probe into the nature of attentional processing. Visual Cognition, 7(1/2/3), 345-376.

Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. Psychological Review, 8(5), 368-373.

Scholl, B.J. & Pylyshyn, Z.W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. Cognitive Psychology, 38, 259-290.

Scholl, B. J. & Xu, Y. (2001). The magical number 4 in vision. Behavioral and Brain Sciences, 24(1), 145-146.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. Trends in Cognitive Sciences, 9(1), 16-20.

Sperling, G. (1960). The information available in brief visual presentations. Psychology Monographs, 74(11), Whole No. 498.

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, 92-114.

Vogel, E. K., Woodman, G. F. & Luck, S. J. (in press). The time course of consolidation in visual working memory. Journal of Experimental Psychology: Human Perception and Performance.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. Perception & Psychophysics, 56(5), 495-500.

Woodman, G. F. & Vogel, E. K. (2005). Fractionating working memory: Consolidation and maintenance are independent processes. Psychological Science, 16(2), 106-113.

Xu, Y. & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. Nature, 440, 91-95.

Figure Captions

Figure 1. The trial structure for every experiment. The cycle of the displays repeated until the participant pressed the space bar to indicate change detection. In this figure, the changing dot is located in the fourth column from the left and in the third row from the bottom.

Figure 2. Results of Experiment 1 (array size = 36, On-Time = varied, ISI = 120ms). (a) Mean response time (RT) in ms (±SE) for the 5 display durations. (b) Mean number of alternations required for successful change detection (±SE). (c) Estimated capacity at each of the 5 display durations[5] (±SE).

Figure 3. Results of Experiment 2 (array size = 36, On-Time = varied, ISI = 900ms). (a) Mean response time (RT) in ms (±SE) for the 5 display durations. (b) Mean number of alternations required for successful change detection (±SE). (c) Estimated capacity at each of the 5 display durations (±SE).

Figure 4. Results of Experiment 3 (array size = varied, On-Time = 700ms, ISI = 120ms). Estimated capacity (±SE) for the 5 array sizes in Experiment 3.

Figure 5. Results of Experiment 4 (array size = varied, On-Time = 700ms, ISI = 900ms). Estimated capacity (±SE) for the 5 array sizes in Experiment 4.

[5] Substituting the mean RTs or mean alternations in Figure 2 into our Equation B will not return an accurate measure of capacity as these measures are skewed towards longer RTs and greater alternations by the typical skew in these distributions. We calculated a capacity for each trial and then computed an average, which somewhat controls for this skew. For instance, calculating capacities from the median RT or median number of alternations, which also corrects for skew, returns estimates of capacity similar to those displayed in Figure 2c.

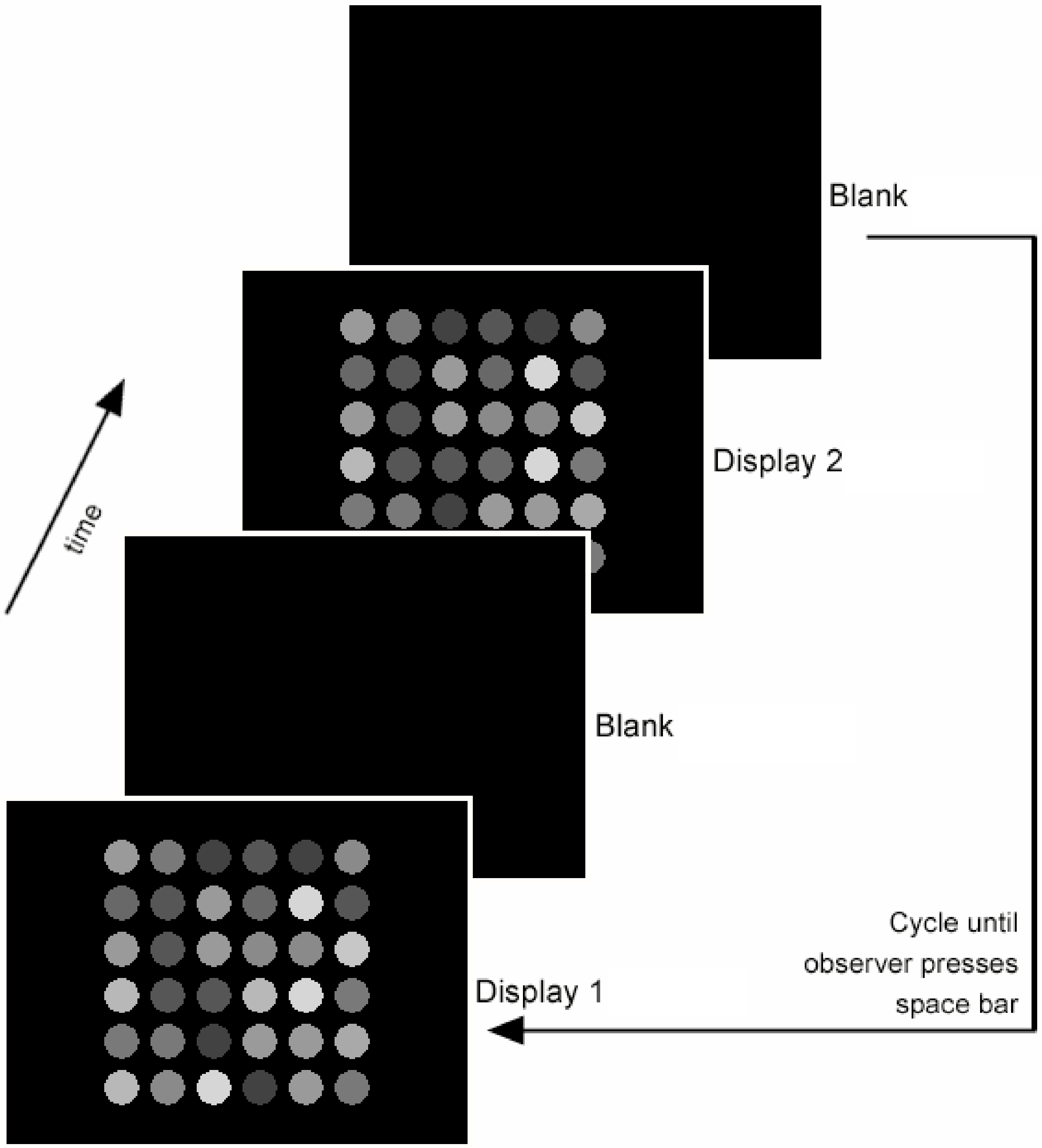
Blank
Display 2
time
Blank
Cycle until
observer presses
space bar
Display 1

Figure 1

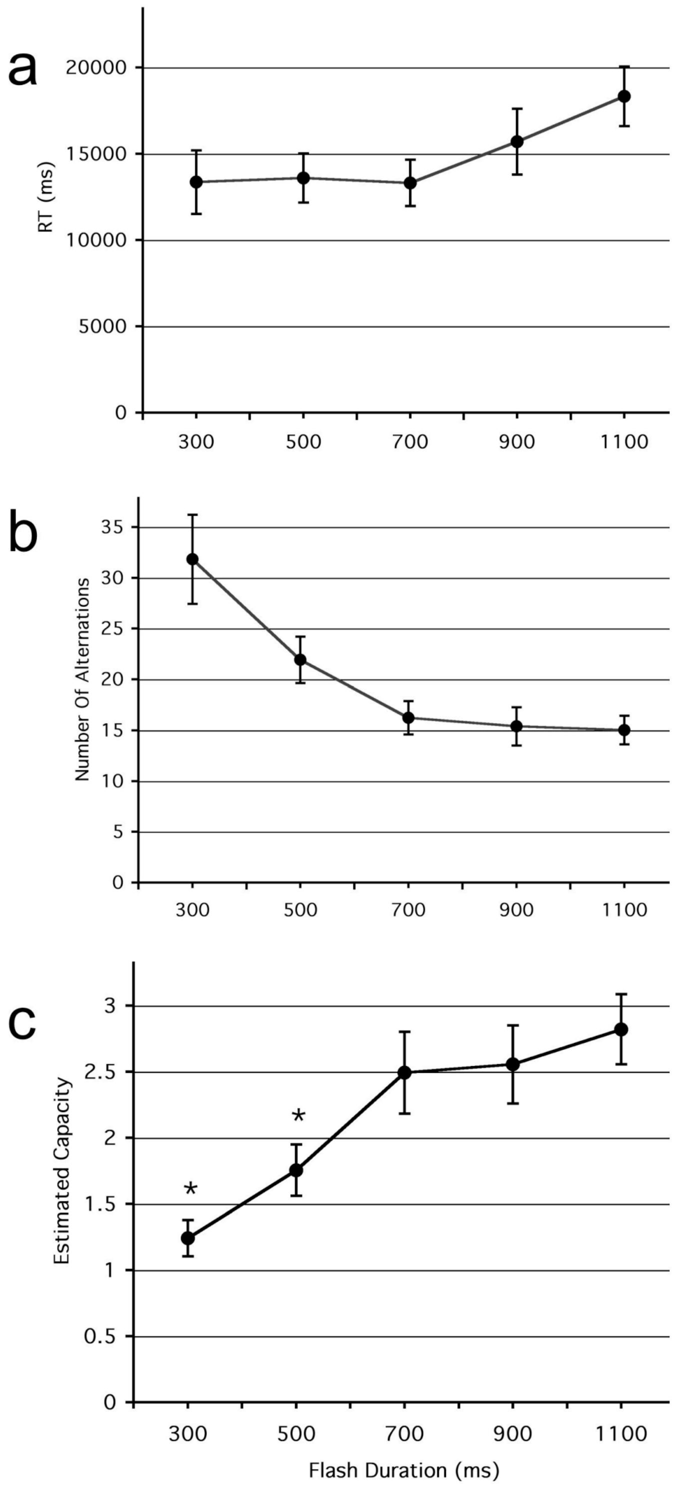
a
b
c
RT (ms)
20000
15000
10000
5000
0
300
500
700
900
1100
Number Of Alternations
35
30
25
20
15
10
5
0
Estimated Capacity
3
2.5
2
1.5
1
0.5
*
*
Flash Duration (ms)

Figure 2

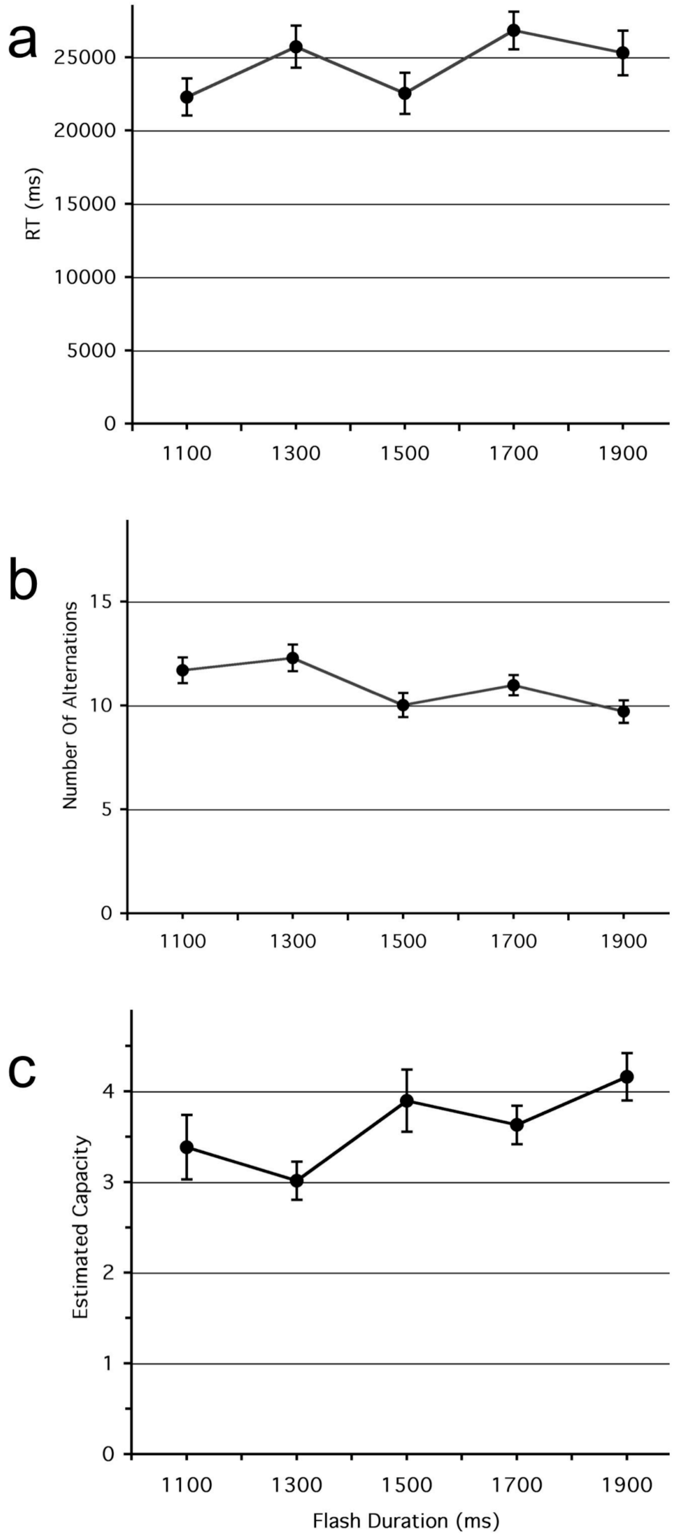
a
25000
20000
15000
10000
5000
0
RT (ms)
1100
1300
1500
1700
1900
b
15
10
5
0
Number Of Alternations
1100
1300
1500
1700
1900
c
4
3
2
1
0
Estimated Capacity
1100
1300
1500
1700
1900
Flash Duration (ms)

Figure 3

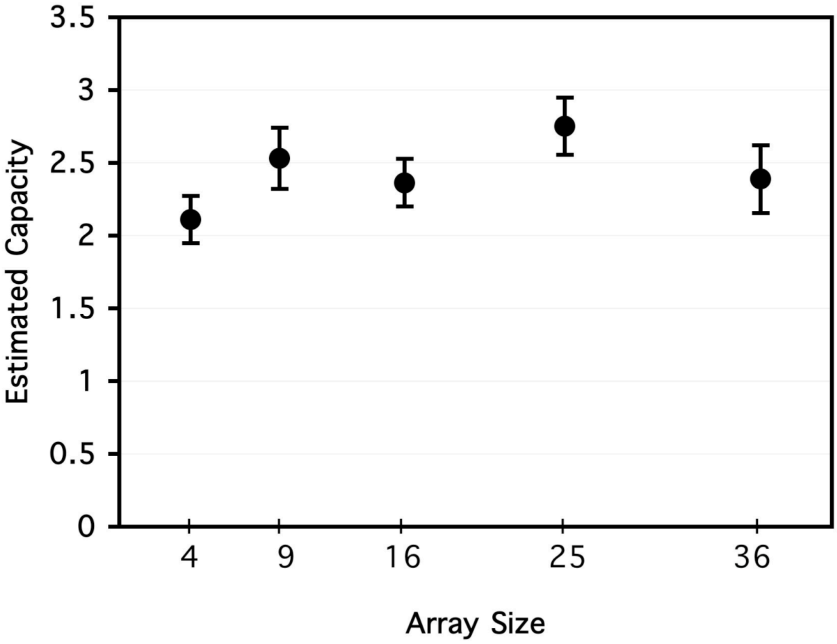

Figure 4

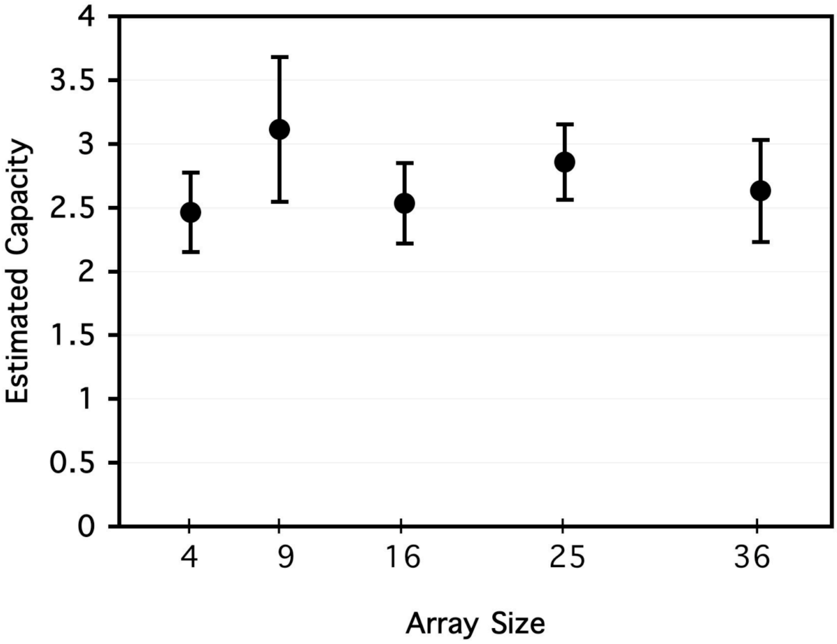

Figure 5