Overcoming the three-item limit:
Gestalt grouping principles explain increases in change detection capacity

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

Previous work suggests that visual processing includes an object-based short-term memory that can represent information from approximately three objects in parallel. Five experiments using a change detection task with grids of simple items varying only in luminance explored what counts as an object for visual short-term memory and how visual long-term memory might contribute to performance. Experiments 1 and 2 revealed a capacity limit of approximately three objects for vSTM which is consistent with estimates from other paradigms. This limit was robust to changes in display time and array size. Experiments 3-5 examined the effect of long-term exposure to an array on change detection performance and found <u>no significant improvement</u> relative to new arrays, suggesting that the three-item limit is a fixed structural limit. Moreover, an algorithm based on the classic Gestalt principles of similarity and proximity accounted for any apparent instances of capacity greater than three items. Although perceptual grouping can change what counts as an "object" for vSTM, the three-item limit of vSTM appears impenetrable to the ameliorating effects of familiarity.

Introduction

When viewing a scene, observers shift attention from one region to another, encoding selected aspects of the scene into visual memory. Given that the capacity of visual memory appears limited, recent research has focused on the ways in which it is limited. How much of a display can we attend to at once and how much of that information is preserved in visual short-term memory (vSTM)? If attention focuses on discrete objects rather than spatial regions of a scene (Kahneman, Treisman & Gibbs, 1992; Luck & Vogel, 1997), and if encoding into vSTM requires focal attention, then the limits on visual memory will depend on the number of objects that can be encoded in a single attentional "glance" — how many objects can be attended to in parallel.

Converging evidence from a wide variety of tasks suggests a magic number of approximately 3-4 items as the capacity limit on attention and vSTM (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974; Pylyshyn & Storm, 1988; Scholl & Xu, 2001; Sperling, 1960; Vogel, Woodman & Luck, 2001; for review see Scholl, 2001). For example, adults can use attention to keep track of the motions of up to four identical, randomly moving items and distinguish them from four identical distractor items (Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999). Both human infants and non-human primates show similar limits in tracking multiple objects; 10 to 12-month-old infants can track up to three objects (Feigenson, Carey & Hauser, 2002; Feigenson & Carey, 2003) and rhesus monkeys can track up to four objects (Hauser, Carey & Hauser, 2000; Hauser & Carey, 2003). This magic number extends to other attention and memory tasks. For example, transsaccadic memory is limited to about four items (Irwin & Gordon, 1998), adults can detect changes to the properties of about four objects in parallel (Luck & Vogel, 1997), and approximately four items can capture attention in a visual search task (Yantis & Johnson, 1990; but evidence from visual marking tasks might suggest a higher number, see Belopolksy, Theeuwes, & Kramer, in press).

Despite the consistency of evidence for a limit of 3-4 items, a few tasks suggest lower or higher limits. The measured capacity for complex stimuli is often fewer than three items (Phillips &

Christie, 1977; Alvarez & Cavanagh, 2003), suggesting an additional limit on the amount of information that can be retained. With complex objects, the information limit is exceeded before the limit of 3-4 objects is reached (Alvarez & Cavanagh, 2003). In contrast, with arrays of simple objects, capacity estimates can be substantially higher than 3-4 objects (Rensink, 2000a). This increased capacity might well result from grouping or chunking processes that allow more information to be encoded into a single higher-order object. For example, three adjacent, similar items might be treated as a single object, resulting in a capacity limit greater than four individual items, but not necessarily more than four "objects" (see Cowan, 2000 for a related discussion of chunking in verbal short-term memory, and see Feigenson & Halberda, 2004 for evidence of visual chunking by infants).

In this paper, we use a change detection task to explore the effects of grouping on estimates of the capacity of visual memory and attention. In the flicker change detection task, original and modified displays alternate, separated by a brief blank screen, until subjects find the change (Rensink, O'Regan & Clark, 1995; Rensink, O'Regan & Clark, 1997; Rensink, 2000). With changes to photographs of natural scenes, observers often need many alternations in order to detect large changes, a phenomenon known as change blindness (Rensink et al, 1997; Simons & Levin, 1997). Change detection tasks provide a useful tool for the study of attention and capacity limits because attention is thought to be necessary for change detection (Rensink et al, 2000; Rensink et al, 1997). Successful change detection requires subjects to encode the pre-change display and then to compare their representation to the post-change display (Simons, 2000). If subjects could attend to and successfully encode all of the elements in a display in parallel, then they potentially could detect changes with a single exposure to the change. However, if the number of objects in the display exceeds the capacity to encode and remember them or if the display duration is too brief to complete encoding, observers will require multiple cycles to detect a change. If observers encode as many objects as possible into visual memory, then given sufficient time, they will maximize the contents of visual memory, filling it to capacity. Once the capacity limit is reached, additional exposure time will not enhance change detection because no additional items could be encoded. This asymptote of change detection performance as a function of display exposure time allows an assessment of capacity limits on the encoding of the displays.

Rensink and colleagues (2000) used this approach with simple arrays of objects to explore the capacity of visual memory for basic feature properties such as orientation and polarity. With a display time of approximately 600ms, search efficiency for an orientation change asymptoted, leading to an estimated capacity of 5.5 items (the method for estimating capacity is described more fully in the results section). In contrast, polarity changes never asymptoted with display durations up to 800ms, leading to a minimum capacity estimate of 9 items. Both of these values exceed the more typical 3-4 item limit for visual short term memory, raising the intriguing question of why subjects showed superior memory for these simple object properties. One plausible explanation is that subjects grouped items together as if they were parts of a single object, and this grouping led to inflated estimates of the number of items that could be retained and compared. One central goal of our research is to address the nature of these grouping mechanisms and their influence on estimates of capacity.

Because change detection requires the subject to compare the contents of visual short-term memory (vSTM) to a present visual scene (i.e. the 1st display to the 2nd display), this task does not distinguish between a capacity limit of vSTM itself and a capacity-limited comparison process. The empirical literature often collapses across these two potentially separable components of mid-level vision (see Rensink 2000c or Mitroff & Simons, in press for discussion). vSTM appears to be an object-limited store that encodes information in object-based rather than retinotopic coordinates (Phillips, 1974), but an object-based attention mechanism involved in the comparison process may also constrain performance. Although we remain agnostic about the relative contributions of these potential limits to estimates of 3-4 item capacity in change detection tasks, throughout this paper, we describe the limits as constraints on vSTM.

Another concern with estimates of capacity based on the flicker task is that long exposure times might allow subjects to rely on long-term rather than short-term memory to perform the task. Increasing the initial exposure time for natural scenes does not greatly facilitate change detection in the flicker task (Rensink, O'Regan, & Clark, 2000), but other evidence suggests that some information from the pre-change display can be encoded into long-term memory (Simons, Chabris, Schnur, & Levin, 2002) and that long-term memory may contribute to change detection (Hollingworth, Williams, & Henderson, 2001). If long-term memory can be used for change detection in the flicker task, it might spuriously inflate estimates of the capacity of vSTM. In several of our experiments, we systematically manipulate familiarity with the displays in order to determine whether long-term memory increases capacity estimates in a change detection task. In Experiment 3, we compared performance when a search display was repeated on many trials to performance when a different search display appeared on each trial. In Experiments 4 and 5, two observers memorized a display to a level that far exceeded the specificity required for change detection. We then estimated how this precise long-term memory influenced performance on a change detection task. Finally, we examined the extent to which grouping contributes to capacity estimates, both with and without contributions from long-term familiarity with the displays. In sum, this paper examines: (a) whether the capacity of visual short-term memory and attention can be increased beyond 3-4 items via contributions from long-term memory, and (b) whether the effects of familiarity result from an increase in the number of items that can be retained in vSTM or from grouping more information into each memory item. Unlike earlier change detection studies, we also used displays that could not be easily encoded categorically or verbally. Consequently, capacity estimates should be more closely tied to visual encoding than to other, non-visual memory mechanisms.

Experiment 1

Experiment 1 measured the capacity of vSTM using a flicker change detection task with arrays of 36 grayscale dots. If the capacity of vSTM is limited, then with a sufficient display time, subjects should completely fill this available buffer and change detection performance should asymptote.

Method

Participants

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

Displays and procedure

All subjects 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. 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 from a viewing distance of 50 cm. 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). Between displays, one dot changed luminance values by ±14 shades (70 volts). For this magnitude of change, most changed 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.

Subjects 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 had a different luminance. After another 120ms blank screen, the sequence repeated until subjects detected the change. The display duration varied across 5 blocks of trials for each subject (300, 500, 700, 900, or 1100ms), with 20 trials per block. The order of these blocks was randomized for each subject. Within a trial, the display duration was the same for both displays, and the blank screen durations were fixed at 120ms for all blocks. When subjects 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 subject pressed the space bar indicating change detection. Participants were instructed to respond as quickly as possible while prioritizing accuracy.

INSERT FIGURE 1 ABOUT HERE

Results and Discussion

All trials for which subjects 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 further removed any trials for which response latency was more than 2 SDs above a subject's mean (3.7% of trials). Mean change detection latency was calculated separately for each display duration (see Figure 2 for means and standard errors). Across all display durations, subjects detected changes after about 12.8s (STDEV = 8.5s) on average.

INSERT FIGURE 2 ABOUT HERE

Figure 2 suggests that search latency is process-limited for display durations ranging from 300-700ms; none of these conditions differed from each other in paired-samples t-tests (all t values $< .3$, all p values $> .8$; see Rensink, 2000a for a similar process-limited effect). Note, though, that shorter display durations provide more exposures to the change for the same response latency. For example, if change detection took 14000ms, observers would have had 32 exposures to the change with a display duration of 300ms, but only 16 exposures with a display duration of 700ms. Apparently, for display durations ranging from 300-700ms, the overall encoding time rather than the number of change exposures determines when subjects will detect a change. In contrast, for longer durations (700-1100ms), response latency increased linearly as a function of display duration. By 700ms, subjects apparently have filled vSTM to capacity, and longer display durations (900 and 1000ms) did not enhance change detection performance.

The finding that the number of cycles needed to detect changes and the overall response latency of change detection provide different estimates of change detection as a function of display duration is surprising given that these two measures typically are assumed to provide comparable estimates of the mechanisms underlying change detection. In fact, the results of Experiment 1 suggest that these two measures will rarely be comparable. For display durations of 300-700ms, response time remained 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 remained constant while response time steadily increased. Apparently, the number of cycles and the total change detection time are only comparable when the display duration gives subjects just enough time to load vSTM to capacity but no more.

Estimating capacity.

Following Rensink et al (2000a), we converted response latencies on each trial into an estimate of capacity and then calculated a 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; but see Horowitz & Wolfe, 1998), 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 from a single display (capacity) plus some constant:[1]

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

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 allow any measurement of either of 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:

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

This estimate is based on the idea that participants load as many items into vSTM as possible during the first display, and then compare the items available in the second display to their memory buffer for the first display. If change detection fails, subjects will store new items in vSTM and the process will continue with each new display. Capacity estimates were calculated separately for each subject for each display duration (see Figure 3). Capacity estimates increased with display durations up to 700ms, but after that point, additional display time had a minimal effect on the number of items 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 repeated measures ANOVA that yielded a significant effect of display duration, $F(4,48) = 21.60$, $p < .001$. Planned pair-wise, within-subject 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

---

[1] 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).

significant differences among display durations ranging from 700-1100ms (700 vs. 900ms, p = .785; 900 vs. 1100ms, p = .183).

INSERT FIGURE 3 ABOUT HERE

The asymptotic performance suggests a capacity of 2.5-3 items, consistent with typical estimates from other vSTM paradigms (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974; Sperling, 1960; Vogel, Woodman & Luck, 2001). Moreover, the display durations necessary to produce asymptotic performance in our study (700ms) were consistent with those found in a between-subjects analysis of orientation change detection (Rensink, 2000a).

Experiment 2

Experiment 1 revealed asymptotic performance with a 700ms display duration using a fixed number of display items. If change detection capacity is limited to 3 objects per display, then manipulating the number of items in the array should not affect this capacity estimate. Experiment 2 replicated Experiment 1 using a fixed display duration (700ms) while varying the number of display items (array sizes of 4, 9, 16, 25, or 36 items).

Method

The materials and procedures for Experiment 2 were identical to those of Experiment 1 except that all trials were presented with a display duration of 700ms and subjects completed 5 randomly-ordered blocks of 20 trials, one with each array set size (4, 9, 16, 25, or 36 items). A new array and a randomly-selected target were generated for each trial. The dots in each display were always presented in a square grid, centered on the display, with constant inter-dot distances. Subjects were 13 undergraduates who received course credit for participating.

Results and Discussion

As in Experiment 1, all trials for which subjects selected the wrong item (5.1% of trials) and all trials with response times more than 2 SDs above a subject's mean (4.6% of trials) were eliminated from further analyses. Response latency on each trial was transformed into a capacity estimate using Equation B, and subject means were calculated for each array set size. The estimated object-based capacity limit for every set size was approximately 2.5 items (see Figure 4), replicating the capacity estimate from Experiment 1. A regression analysis predicting capacity from array set size revealed no significant effect of array size (the 95% confidence interval for the slope ranged from -.008 to .02, $F(1, 63) = 1.05$, $p = .31$). Consistent with the claim that vSTM is capacity-limited to approximately 3 items, the number of items in the array had little effect on capacity estimates.

INSERT FIGURE 4 ABOUT HERE

Experiment 3

Experiments 1 and 2 suggest a capacity limit of 2.5 to 3 items for simple visual stimuli that cannot be readily verbalized or encoded categorically. If this is a fixed, structural limitation then it should be unaffected by experience or familiarity. That is, if subjects can only compare three items at a time, performance should not improve even if memory for the items in the display improves; a capacity limit based on the number of items that can be compared without disrupting memory for the items should not improve with familiarity. Alternatively, if the capacity estimate of 2.5 to 3 items results from the inability to encode simple visual stimuli well enough to retain them across the blank, then familiarity should increase capacity estimates. If this limit is purely information-based rather than object-based then familiarity might also increase capacity as encoding of familiar items could require less information. In Experiment 3, subjects viewed the same dot array on many trials and we compared capacity estimates for this familiar array to estimates derived from novel displays as in Experiments 1 and 2. Comparable capacity estimates of 2.5-3 objects for both familiar and novel displays would suggest a fixed structural limitation on the number of objects that can be attended, stored and compared across a blank.

Method

The materials and procedures for Experiment 3 were identical to those of Experiment 1 except that all trials included 36 items and were presented with a fixed display duration of 700ms. 12 subjects completed two sessions on separate days. In the Varying Display session, subjects completed 100 trials in which a new display was generated on each trial just as was the case in Experiments 1 and 2. As in Experiments 1 and 2, the luminance of each dot and the location of the changing dot were generated randomly on each trial. In the Constant Display session, subjects completed two consecutive blocks of 50 trials. For all 50 trials of a block, subjects viewed the same initial display, and on each trial, one dot was selected randomly (with replacement) to be the changing one. In other words, subjects performed the search for a change on the same display for 50 consecutive trials (about 30 minutes). The second block of trials was identical in structure to the first except that a different randomly-generated array was used for all 50 trials. Each subject had different randomly-generated displays. The order of the Constant Display and Variable Display sessions was counterbalanced across subjects.

Results and Discussion

As in Experiments 1 and 2, error trials (2.1% of trials) and outlier response latencies (4.4% of trials) were eliminated from further analyses. Response latency on each trial was transformed into a capacity estimate using Equation B, and subject means were calculated from successive sub-blocks of 10 trials each. These sub-blocks allow an assessment of whether increasing familiarity with the display over the course of the Constant Display block produced an increase in capacity. Sub-blocks were yoked across the Constant Display and Variable Display conditions for analysis, such that the first 10 trials of the Constant Display condition were compared to the first 10 trials of the Variable Display condition. Increasing familiarity with the display did not lead to greater capacity (see Figure 5). This result was confirmed by the absence of any reliable differences in a 2 (Condition) x 5 (Sub-block) repeated measures ANOVA: effect of Display Condition, $F(1, 11) = .65$, $p = .436$; effect of Trial Block, $F(1, 11) = .84$, $p = .508$; interaction, $F(4, 44) = .18$, $p = .948$. Although Figure 5 suggests a trend toward greater capacity in the Constant Display condition on the final three sets of 10 trials (i.e. 70-80, 80-90, 90-100), the trend was not statistically reliable when tested using planned within-subject t-tests: sub-block 70-80, $t(11) = .44$, $p = .667$; sub-block 80-90, $t(11) = 1.11$, $p = .291$; sub-block 90-100, $t(11) = .45$, $p = .659$. Consistent with Experiments 1 and 2, the estimated capacity in both the Variable Display and Constant Display conditions was approximately 3 items.

INSERT FIGURE 5 ABOUT HERE

Surprisingly nearly 30 minutes of searching through the same 36-dot array did not lead to greater capacity in a change detection task. By trial number 50, participants had been searching an identical array of 36 gray dots for approximately 30 minutes. They had also been <u>attending</u> to each individual dot with focussed attention and encoding luminance information in order to find the single changing item on each of the previous 49 trials. Presumably, after 50 trials, subjects encoded the Constant Display better than would be possible with a novel display. Despite this extensive exposure, subjects were no faster to find the changing item in this familiar array than in a random array[2]. This finding suggests that long-term memory for the dots in the array does not

---

[2] Note that subjects could not have conducted the search entirely in memory because a different dot changed on each trial. Other studies have shown that subjects can sometimes perform a memory search more efficiently than a visual search when the display is constant across trials (Wolfe, Klempen & Dahlen, 2000). That approach is ineffective in a change detection task of this sort.

supplement the capacity for detecting changes to simple visual stimuli that are not readily categorized or labeled. Apprently, vSTM is relatively impenetrable to influence from long-term memory, a finding consistent with vSTM studies producing similar capacity estimates with familiar alphanumeric items (Miller, 1956; Pashler, 1988; Purdy, Eimann & Cross, 1980; Sperling, 1960).

One concern with drawing such a strong conclusion is that it is premised on the absence of an effect. Perhaps 30 minutes and 50 trials were insufficient to produce a strong effect of familiarity on capacity. Another concern is that the sort of information stored in long-term memory might not be adequate to increase the capacity for change detection with this sort of stimuli. If so, the absence of a difference between the Variable Display and Constant Display conditions could be an artifact of the particular stimuli chosen for these studies. Experiments 4 and 5 address these concerns.

Experiment 4

Experiment 4 tested whether the null result of Experiment 3 resulted from insufficient familiarity with the displays or from the impossibility of encoding our particular stimulus set into long-term memory. Two participants undertook the Herculean task of memorizing an entire grid of 36 gray-scale dots to a level of specificity that far exceeded that required for the change detection task. Successful memorization of the display constitutes an existence proof that the stimuli could be encoded into long-term memory. Presumably, complete memorization also constitutes evidence that the display is familiar. The purpose of this experiment was to provide the strongest test possible of the potential benefits of familiarity on estimates of capacity for change detection. The results of experiments 1-3 suggest, counter-intuitively, that subjects should be no faster to find a changing item in a familiar display, no matter how familiar.

Method and Results

Two participants (the author J.H. and an undergraduate volunteer G.P. who was naïve to the hypotheses being tested) completed a memorization study phase, a memorization test phase, and a visual search phase of the experiment.

Memorization Study Phase

Each subject completed 24 memorization sessions (approximately 10 hours total) over the course of 3 weeks in which they memorized the location and luminance of each dot in the 36-dot grid displayed in Figure 6 (Mean grid luminance = 117.4 volts, SD = 43.8 volts).

INSERT FIGURE 6 ABOUT HERE

At the start of each session, subjects viewed the entire grid until they felt prepared to begin the memorization testing (Mean free viewing per session was approximately 2 minutes). Subjects then pressed the space bar and all of the dots in the grid were replaced by white plus signs (+). On each of 100 trials in each session, a single randomly selected dot appeared in one of the grid locations and subjects indicated by key press whether or not this dot had the correct luminance for that grid location. For the first 12 sessions, a liberal criterion was adopted: a "yes" response was coded as correct if the probed luminance was within ±3 luminance values of the actual luminance of that dot in the memorized grid. The probe's luminance was selected at random from the 30 possible luminance values with the constraint that one half of the trials be correct by our liberal criterion. Thus, 7 of the 30 luminance values were considered to be correct (i.e. actual value ± 3). Recall that in the change detection task, the target dot changed luminance by ±14 luminance values. Thus, success in these 12 early sessions required a memory representation that was more accurate than that required by the change detection task. Response time and accuracy were recorded for each trial.

Both subjects responded faster with increasing exposure to the grid over the first 12 sessions (1 week) of memorizing: for G.P., response times decreased by 25.44ms/session on average, t (10)

= 2.427, $p < .05$; for J.H., response times decreased 45.78 ms/session, t (10) = 5.705, $p < .001$. Accuracy increased for both subjects: for G.P., by 1.46%/session, $t(10) = 4.504$, $p < .005$; for J.H., by 0.71%/session, $t(10) = 2.235$, $p < .05$ (see Figure 7). In sum, both subjects improved over the course of the first 12 sessions when using a relatively lenient criterion for correct judgments.

INSERT FIGURE 7 ABOUT HERE

In the second set of 12 sessions (Difficult Sessions), the criterion for a correct response and the selection of probes was altered to require more precise memory. The luminance of the probe dot was always within ±7 luminance values of the actual value, and only the actual luminance value ±2 values were considered correct. Thus, successful performance required subjects to discriminate between a ‘correct’ value (-2, -1, actual, +1, +2) and an ‘incorrect’ value (-7, -6, -5, -4, -3, +3, +4, +5, +6, +7). As in the first 12 sessions, the probe luminance varied randomly within this range with the constraint that half of the trials were ‘correct’. As with the first 12 sessions, the level of specificity in long-term memory required to perform this task far exceeded the precision needed for the change detection task. The difference between these levels of specificity can be seen in Figure 8. At the top is the memorized grid. Below it to the left, the same grid is presented with one dot changed by 2 luminance values (the level required for the Difficult Sessions). Below it to the right, the same grid is presented with the same dot changed by 14 luminance values (the level required for the change detection task).

The criterion chosen for the Difficult Sessions was quite stringent. Thus, rather than looking for improvement across these sessions, it is more important to note whether or not subjects could perform the task at all. Both subjects successfully discriminated correct and incorrect probe dots at better than chance levels, where chance is 50% (see Figure 7): for GP, t (11) = 11.02, $p < .001$; for JH, t (11) = 13.12, $p < .001$.

INSERT FIGURE 8 ABOUT HERE

Upon completion of all 24 sessions, both subjects wrote descriptions of how they performed the task (subjects were not allowed to discuss their strategies with one another throughout the experiment). Subject J.H. described attempting to form figures from adjacent dots. These figures were assigned some semantic label or description (e.g., "the bright dots on the left and top of the grid with the black dots inside are an upside-down ‘L’, the gray dots in the center form a counter-clockwise spiral of decreasing luminance,” etc). J.H. drew the grid and circled these groups. The majority of the groups in this drawing were formed from adjacent dots of similar luminance. Subject G.P. chose a different strategy: "I saw each row as a unit. I would think of it as a stick from left to right where each dot contrasted with the ones beside it by some specified degree. I gave labels to these contrasts like 'early autumn', 'late autumn' etc.” Thus, both subjects attempted to chunk the display into larger units, J.H. defining units based on proximity and similarity of luminance and G.P. defining units based on relative changes in contrast within a row.

Memorization Testing Phase

After the 24 memorization sessions, both subjects completed three test sessions without being allowed to view the complete grid. In each trial of each test session, subjects viewed a grid of plus signs followed by a probe dot at one of the plus locations. This dot was either the correct dot for that location or differed from that dot by ± 14 luminance values. Subjects pressed keys to indicate whether the probe dot was correct or incorrect.

Both subjects performed this task well (for G.P., M = 96% correct and 1754ms response latency; for J.H., M=100% correct and 1289ms response latency). Together, the Memorization Study Phase and the Memorization Testing Phase show that both subjects successfully committed the 36-dot array to long-term memory with a level of specificity that could potentially aid change detection performance. As further evidence of the robustness of this representation, J.H. completed

3 additional test sessions of 100 trials each after not having viewed the array for one year. In these sessions, J.H. averaged 93% correct (average RT = 1432ms).

Visual Search Phase

Following the first three memorization testing sessions, subjects participated in additional sessions designed to test visual search performance using the memorized grid. The purpose of these tasks was to verify that the long-term memory representations of the grid could subserve visual search for a changed dot. J.H. completed 3 sessions of 100 trials each examining the ability to find a changed dot in the complete grid. On each trial the 36-dot grid appeared and was either identical to the memorized grid (half of the trials) or 1 randomly chosen dot was changed by ±14 luminance values. The task was to press one key if it was the memorized grid and a different key if one dot differed.

In these sessions, J.H. detected grids with a changed dot 90% of the time, and falsely reported a changed dot only 4% of the time (mean RT = 4468ms). When a changed dot was present, search rate averaged 154.7ms/item and when no change was present, it averaged 122.9ms/item. Another way of expressing search rate is in terms of the number of dots searched per second (a compliment of the search rates given above). Expressed this way, J.H. attended to and rejected an average of 6.5 dots/second when no change was present and 8.8 dots/second when a change was present.

Recall that with a display time of 700ms, change detection performance appears limited to inspection of approximately 2.5 dots/display. Because the search through long-term memory is far more efficient than the predicted limit of 2.5 dots per 700ms display, Experiment 4 proves that long-term memory could potentially aid J.H. and G.P. with change detection for the memorized grid, a possibility tested in Experiment 5.

Experiment 5

Can long-term memory information affect the capacity limit of vSTM? If vSTM is a buffer with a structural object-based limit of 3 items, then familiarity with a display should not improve change detection performance. If vSTM is instead an information-limited buffer, then familiarity might lower the information load of each individual object, thereby allowing more dots to be stored. Experiment 5 tested these two possibilities and more generally whether long-term memory can contribute to the detection of changes in visual stimuli that are not readily categorized or labeled.

Method

Experiment 5 used the change detection task from Experiments 1-3 with display duration fixed at 700ms. Both G.P and J.H. participated in 8 sessions of 100 trials each, with each session completed on a different day. Each session consisted of ten blocks, with ten trials per block. On the odd numbered blocks of each session, subjects performed the change detection task on the memorized grid. On even numbered blocks, subjects performed the change detection task on randomized grids that were matched to the memorized grid in terms of the mean and standard deviation of the overall luminance. For memorized grid blocks, some trials started with the memorized grid and some started with the changed grid. By varying the initial display, subjects were less able to limit their search to the changed array (something they were able to do efficiently in the visual search component of Experiment 4). To do so, subjects would need to limit their search to every other display. Neither subject reported using this strategy. For blocks with the memorized grid, once subjects pressed the space bar to indicate change detection, the memorized grid was re-displayed and subjects indicated the change location with the mouse. For the randomized grid trials, the first display of the trial was shown during change localization.

Although subjects did not report attempting to search through only the memorized grid, they potentially could detect a change in this grid by noting a difference in luminance at a memorized location. That is, if the subject's attention happened to fall on the changed item when its luminance was different from its memorized value, they might detect the discrepancy from long-term memory rather than from the previous display of the memorized grid. Such change detection would amount to finding the "wrong dot" rather than finding the "changing dot." This route to detection did not exist for the randomized grids. Consequently, the capacity estimate for the memorized grid could be slightly inflated by this alternative means of change detection, and any advantage for the memorized grid blocks should be interpreted with caution.

Results and Discusion

Each session of 100 trials had 50 trials with the memorized grid and 50 with randomized grids. These trials were yoked such that the first memorized grid trial was analyzed with respect to the first randomized grid trial for each session. Thus, each subject had a possible total of 400 trial pairs across the 8 sessions. G.P. selected the incorrect change item on 3 trials and had unusually long response times on 5 additional trials. These trials and their paired trials were eliminated from the analyses, leaving a total of 392 trial pairs. J.H. selected the incorrect item on 3 trials, leaving 397 trial pairs.

After each session, subjects reported whether they had detected the change by happening to notice a dot that was the 'wrong' luminance for that location in the memorized grid. Both G.P. and J.H. reported discovering the 'changing' dot by noticing the 'wrong' dot on approximately 10% of trials with the memorized grid. This result suggests that a visual scene cannot be directly compared in its entirety to a representation in long-term memory (which potentially could have produced 100% detection). Rather, information from long-term memory may be subject to the same comparison constraints as vSTM and it is possible that information from long-term memory may need to be downloaded into vSTM in order to be compared to an incoming stimulus.

Can long-term memory information increase the capacity of change detection? Despite the concern that capacity estimates were inflated by the detection of the 'wrong' dot, G.P. had an average capacity of approximately 4 dots for both the memorized and randomized grids across sessions (see Figure 9). Note that prior to Experiment 5, G.P. had no experience with the change detection task, and his capacity was approximately 3 items for both types of grids in session 1.

INSERT FIGURE 9 ABOUT HERE

A 2 (memorized grid vs. randomized grid) X 8 (session) repeated measures ANOVA on G.P.'s performance revealed a significant effect of Grid Type: $F(1, 384) = 8.24$, $p < .005$. Collapsing across sessions, G.P. had a slightly higher capacity for the memorized grid (M=4.74 items) than for the randomized grid (M=3.91 items). Again, this result must be interpreted cautiously given that G.P. reported noticing the 'wrong' dot on approximately 10% of memorized grid trials. Planned comparison t-tests on each session revealed a significant effect of Grid Type in session 3: $t(47) = 2.17$, $p < .05$, and in session 6: $t(48) = 2.34$, $p < .05$, as well as a marginal effect in session 7: $t(49) = 1.94$, $p = .058$. Interestingly, G.P.'s capacity also increased with practice as revealed by a significant effect of Session: $F(7, 384) = 4.16$, $p < .001$. By session 8, G.P.'s capacity was approximately 6 dots for the memorized grid *and* the randomized grids! This finding shows that capacity can exceed 3 items even if extreme familiarity with a display does not increase change detection performance. But does this 'super-capacity' truly reflect an increased capacity of vSTM? A possible alternative is discussed in the "Perceptual Grouping" section below.

J.H. also had an average capacity of approximately 4 dots for both the memorized and randomized grids across sessions. Collapsing across sessions, J.H. showed a slightly higher capacity for the memorized grid (M=4.85 items) than for the randomized grids (M=4.17 items): $F(1, 389) = 6.80$, $p < .01$ (see Figure 10). J.H.'s capacity also increased for both grid types across the 8 sessions: $F(7, 389) = 5.25$, $p < .001$. And, importantly, J.H.'s capacity in session 1 was approximately 3 items for both grid types (Memorized grid = 3.7, Randomized grids = 2.8). Planned t-tests revealed that J.H. had a significantly higher capacity for the memorized grid than the randomized grids in session 4: $t(48) = 2.52$, $p < .05$, and it was marginally higher in session 3: *t*

(49) = 1.91, *p* = .062. By session 5 the advantage for the memorized grid disappeared as capacity for the randomized grids improved to the same level as for the memorized grid.

INSERT FIGURE 10 ABOUT HERE

Thus, Experiment 5 provides no conclusive evidence that change detection in familiar displays is any easier than in novel displays (see also Pashler, 1988). Both G.P. and J.H. had estimated capacities in excess of the 3 items found in earlier studies. But, this super-capacity was not restricted to the memorized grid. Rather, by session 8, both G.P. and J.H. had estimated capacities of approximately 6 items for both the memorized and randomized grids. This finding suggests that improvements in capacity were not caused by familiarity with a particular display. Rather, they resulted from experience with this *type* of display and with the nature of the change detection task. But did G.P. and J.H. really increase the capacity of vSTM itself or did these higher capacity estimates result from perceptual grouping or chunking?

Perceptual-Grouping

Just as capacity estimates for verbal short-term memory increase when subjects can group more items into the same number of memory slots (Cowan, 2000, Ericsson, Chase & Faloon, 1980), vSTM might be limited to approximately three items, but what counts as an item may change. That is, subjects might be able to group multiple dots into single memory slots, thereby increasing an estimate of capacity without fundamentally increasing the number of "items" that can be stored in vSTM. Did grouping mechanisms contribute to performance in Experiment 5, leading to capacity estimates greater than 3 items?

To test for the possible contributions of perceptual grouping to capacity estimates, we analyzed the memorized and randomized grid blocks from Experiment 5 based on two simple predictions: (1) J.H. and G.P. used explicit grouping strategies to memorize the memorized grid.

When these strategies were not effective for a particular dot in the memorized grid (because of the relative luminance of the surrounding dots), capacity should be approximately 3 items. And (2) for both the memorized and randomized grids, change detection capacity should vary as a function of the size of the perceptual group that includes the changing target (i.e. the bigger and more salient the group, the easier it should be to find the target).

To estimate the possible perceptual groups within a display, we developed a simple algorithm based on the Gestalt principles of similarity (in luminance) and proximity (Palmer & Rock, 1994; Wertheimer, 1923/1955). The luminance and location of each dot in the display was recorded throughout Experiments 5. For each display, we calculated the similarity (in luminance) of all of the dots that surrounded the target item. We then devised a cluster rating for each dot that consisted of the number of dots "connected to" it that were within ±4 luminance values of the target, divided by 35 (the total number of distractor dots). In order to calculate proximity, or "connectedness", the similarity calculation was first performed on dots immediately adjacent to a given dot (the first layer) and then for dots surrounding the immediately adjacent dots (the second layer) and so on throughout the entire grid. In order for a dot in the second layer to be counted as similar, it had to be connected to a dot in the first layer that was also similar (and so on for all possible layers in the grid).

<u>The Memorized Grid</u>

Figure 11 shows the cluster rating for each dot in the Memorized grid from Experiment 5. The results of the cluster analysis effectively distinguished dots that were similar to their surroundings from those that were singletons.[3]

[3] Cluster-rating did not significantly correlate with luminance in the Memorized Grid ($p > .18$), and in previous experiments we have found no correlation of estimated capacity with luminance of the changing target (all $p > .5$). For a grid-size of 36 dots, having 1 similar neighbor translated into approximately a .03 Cluster-rating, 2 neighbors was approximately a .06 Cluster-rating and so on.

INSERT FIGURES 11, 12, and 13 ABOUT HERE

Next, we computed subject J.H.'s estimated capacity for each target location in the memorized grid using the 397 memorized grid trials from Experiment 5 (Figure 12). Cluster rating and capacity were positively correlated, $r = .424$, $t\,(394) = 9.307$, $p < .0001$(see Figure 13), illustrating that when the changing item was part of a similar cluster of items, J.H. found it faster. Moreover, when the target was a single dot with no similar dots connected to it (i.e. cluster rating = 0), J.H.'s capacity was 3 items. Capacity estimates increased noticeably as the size of the cluster increased, suggesting that J.H. had segmented the memorized grid into perceptual units larger than a single dot on the basis of their proximity and similarity in luminance. This finding is consistent with J.H.'s reported strategy for encoding the memorized grid. Note, of course, that other grouping strategies might also contribute to and account for increased capacity estimates. We calculated only a luminance- and proximity-based grouping measure that accounted for a significant proportion of the variance in J.H.'s change detection capacity, and other plausible grouping metrics might well account for additional variance.

G.P. reported using a different strategy than J.H. for memorizing the grid in Experiments 4-5. He attempted to form an image in his mind of each <u>row</u> of dots and used the contrasts within the row to remember the luminance for each individual dot. Thus, G.P. attempted to form perceptual groups of each individual row, irrespective of the similarity of adjacent dots that cut across rows. If perceptual grouping is based on the processing of similarity and proximity (Rock & Brosgole, 1964; Rock, Nijhawan, Palmer & Tudor, 1992a; Rock, Nijhawan, Palmer & Tudor, 1992b), then Gestalt grouping mechanisms should interfere with the chunks that G.P. attempted to form. Given that G.P. focused on contrasts between dots rather than similarities among dots, our cluster rating might be negatively correlated with estimates of his capacity. In other words, given his memorization strategy, G.P. might find changes more efficiently when the target dot is <u>dissimilar</u> to adjacent dots. With high dissimilarity, interference from similarity-based grouping would be reduced and G.P.'s strategy would be maximally effective.

INSERT FIGURE 14 ABOUT HERE

For G.P., our cluster rating was negatively correlated with estimates of his capacity, $r = -.224$, $t(391) = -4.558$, $p < .0001$ (see Figure 14). This negative correlation suggests that the natural grouping of items based on similarity and proximity (as estimated by cluster rating) interfered with his strategy of row-based grouping. G.P. detected changes faster for target items that were dissimilar to all surrounding items (cluster rating = 0) than for changes that were part of a cluster of similar items (e.g., cluster rating > .2). Moreover, when G.P.'s grouping strategy was ineffective, i.e. when the target was surrounded by a large group of similar dots (cluster rating = .26), G.P.'s estimated capacity was approximately 3 items.

For both subjects, when the target location coincided with an optimal group according to their memorization strategy, they showed increased capacity. However, when the target location was poorly grouped based on their strategy, capacity was 3 items. Apparently, capacity estimates were inflated by grouping mechanisms. Grouping allowed the subjects to change what counted as an item, clustering several dots into a single memory slot. For J.H., these perceptual groups were consistent with the Gestalt principles of similarity and proximity. For G.P., these groups were the rows themselves. Importantly, when these grouping strategies failed (J.H. cluster rating = 0, G.P. cluster rating = .26), both subjects showed the predicted capacity of approximately 3 items.

The Randomized Grids

Perceptual grouping contributed to performance with the memorized grid for both subjects and explains how capacity estimates sometimes exceeded 3 items. J.H. and G.P. had memorized this grid, so they had sufficient time and exposure to form optimal perceptual groups. However, in sessions 5-8, they also showed capacities greater than 3 items for the randomized grids, suggesting that extreme familiarity with the display was unnecessary for perceptual grouping to affect capacity estimates. Did grouping strategies contribute to capacity estimates even for these novel grids?

We calculated our cluster rating for the target in each of the randomized grids in sessions 5-8 (where subjects showed greater capacity).[4] Cluster rating was positively correlated with change detection performance for J.H. (Figure 15), $r = .221$, $t\,(198) = 3.194$, $p < .002$, r-squared = .04901, suggesting that J.H. increased his change detection capacity by chunking elements of the randomized grids according to proximity and similarity of luminance. J.H. reported that he had not used an explicit strategy on randomized grid trials, so this grouping strategy may have been more implicit or automatic. Interestingly, the correlation for J.H.'s sessions 1-4 was not significant ($r = .07$, $p = .3$, r-squared = .005), suggesting that the grouping strategy was not yet effective and that practice with the task is necessary for grouping to contribute to change detection for novel displays.

For the randomized grids in sessions 5-8, G.P. showed a non-significant <u>positive</u> correlation between the cluster rating and capacity, $r = .125$, $p = .16$, r-squared = .016 (see Figure 16). Evidently, although G.P. organized the memorized grid by rows, the more natural grouping based on similarity and proximity contributed to his performance on randomized grids.

INSERT FIGURES 15 & 16 ABOUT HERE

Together, the analyses of the memorized and randomized grids from Experiment 5 suggest that perceptual grouping contributed to change detection performance, leading to increased estimates of capacity. However, when grouping strategies were ineffective, capacity remained at 3 items.

---

[4] There are a number of possible cluster-rating analyses that might be carried out on the random grid trials as targets in these trials necessarily did not have a "memorized" luminance. The analyses shown in Figures 15 and 16 are the result of running the cluster-rating algorithm on the luminance of the target dot in the first flash of the change detection task.

Conclusions on Perceptual Grouping

Our analysis of perceptual grouping suggests that long-term memory does not expand the item-based capacity limit of vSTM. Rather, information in long-term memory may augment subject's ability to chunk a display into larger perceptual units, thereby enhancing change detection. Moreover, the ability to adopt such grouping strategies does not depend on extreme familiarity with a particular display. Provided subjects have sufficient experience with the task, they can group elements in randomized displays. Interestingly, the particular grouping strategy that subject's adopted had systematic effects on their later processing of the memorized grid demonstrating that prior experience with a stimulus and top-down construal can affect the later perceptual grouping of an array (Beck & Palmer, 2002; Kimchi & Hadad, 2002; Wertheimer, 1923/1955).

General Discussion

This set of experiments revealed a capacity limit for change detection of approximately three visual items. Experiments 1 and 2 systematically varied the display duration and the number of items in the display, revealing a fixed limit of approximately 2-3 items (see also Rensink, 2000a). One particularly interesting aspect of this result was 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. This finding suggests that overall response time and the number of displays viewed are not, as has been commonly assumed, equivalent estimates of change detection performance.

Experiments 3-5 demonstrated, surprisingly, that familiarity with an array of items does not increase the capacity of change detection. Even when subjects memorized an entire array of items to a level of specificity that far exceeded the demands of the change detection task, their change detection performance was not substantially improved! Experiments 3-5 suggest that familiarity and experience with a task can lead to improved grouping of elements of a display into larger perceptual groups and that this clustering can facilitate change detection when the target is part of such a cluster. Using a simple estimate of perceptual grouping, we found that clustering predicted performance for both memorized and randomized displays, and in the absence of a larger perceptual cluster, change detection capacity was again limited to 3 items. Although the capacity for change detection is apparently limited to 3 items, Experiment 5 showed that what counts as an item for change detection can change. When the items in a display become part of larger perceptual groups, more of them can be stored and compared across a delay without exceeding the 3-item limit of vSTM.

This suggestion parallels a similar argument from the verbal short-term memory literature, where estimates of capacity occasionally exceed a 3-item limit (Cowan, 2000). Our findings with simple, non-categorically defined stimuli are consistent with other change detection work designed to estimate capacity. For example, changes to orientation produce capacity estimates of approximately 5 items (Rensink, 2000a). Note also that capacity estimates for orientation and polarity stimuli used in earlier change detection studies might also be inflated due to grouping, categorical encoding, or even verbal labeling. It remains possible that capacity estimates for these other stimuli would be comparable to our estimate of 3 items were grouping and other encoding strategies eliminated.

Our evidence that change detection is no better with a memorized grid than with a randomized grid suggests that familiarity with the display does not contribute directly to change detection performance (see also Rensink, O'Regan, & Clark, 2000). This finding is consistent with earlier evidence that changes to familiar alphanumeric characters (Pashler, 1988) are not detected any more efficiently than changes to random dot arrays (Phillips, 1974).

In contrast to these findings from change detection research, in standard visual search tasks, search for a highly familiar target becomes highly efficient (Suzuki & Cavanagh, 1995; Wang, Cavanagh, & Green, 1994). One way to account for this discrepancy is that change detection requires attention to the target both before and after the change in order to detect it (Rensink, 1997, 2000a, 2000b). In contrast, the effects of familiarity on classic visual search appear to operate without a need for focused attention to the distractor items (Wang, Cavanagh & Green, 1994). With the exception of perceptual grouping, this sort of "pre-attentive" processing would provide little aid to change detection given the need for focused attention to the changing item during encoding, active maintenance of the item in vSTM during the blank delay, and active comparison of the pre-change and post-change item (Simons, 2000). We also note that this aspect of change detection leaves open the possibility that the 3-item limit observed here may reflect a limit on the comparison process itself rather than a limit on perceptual encoding or vSTM *per se*. The possible differentiation of the theoretical constructs of vSTM, parallel attention, and the active comparison of remembered stimuli to currently attended stimuli remains an important challenge for future research (see Mitroff & Simons, in press).

Some recent evidence suggests that the capacity of vSTM may sometimes be <u>fewer</u> than 3 items (Alvarez & Cavanagh, 2003). Although vSTM might have a maximum capacity of 3 items, it may also be limited in terms of the information that can be encoded for each item. If the features to be encoded are categorical (e.g., red vs. blue) and easily distinguished, then capacity might reach its maximum of approximately 3 items (Luck & Vogel, 1997). However, the information limit may be reached before the 3-item limit when the items to be stored are complicated or require fine discriminations (Alvarez & Cavanagh, 2003). Conceivably, this information limit could be increased with experience or familiarity such that the capacity for complicated items could be increased. However, as shown here, this capacity is unlikely to exceed the 3-item limit even with extensive familiarity.

A positive proposal derived from the present experiments and the existing literature is outlined in Figure 17. This schematic of the structure of visual processing draws directly from Rensink's proposed triadic architecture (2000b, 2000c). Early visual processing operates in parallel to rapidly organize the visual scene into proto-objects (discrete packages of features such as color, orientation, and size that are potential objects but that are volatile and retinotopic; see also

Triesman & Gelade, 1980; Wolfe, 1994). Attention is required to perceive overall shape and the relations among features (Wolfe & Bennett, 1997). With the application of attention, a proto-object coheres into a fully integrated individual. It is the maintenance of a subset of these individuals (up to three) in vSTM that allows a change in to be detected (Rensink 2000b, Simons, 2000). vSTM is a buffer that holds information up to 3 individuals in parallel (the present results; *see also* Luck & Vogel, 1997; Scholl, 2001; Rensink 2000a; *but see* Alvarez & Cavanaugh, 2003). Information about the individuals in vSTM can be sent to long-term memory. The present experiments suggest that information from long-term memory cannot increase the item-based capacity of vSTM even when the items to be stored in vSTM are highly familiar. Rather, information in long-term memory exerts an effect on capacity through perceptual grouping (Beck & Palmer, 2002; Kimchi & Hadad, 2002; Wertheimer, 1923/1955). Perceptual grouping appears to operate after the processing of color and size constancy (Rock & Brosgole, 1964; Rock et al., 1992). These and related processes are likely to operate without attention (e.g., the extracting of the gist of a scene; see Oliva & Schyns, 1997; Rensink, 2000c; Swain & Ballard, 1991).

Our results suggest that representations in visual long-term memory do not participate directly in change detection (at least they do not benefit change detection beyond the capacity limits on vSTM or the comparison process). Long-term exposure to a stimulus <u>can, however,</u> influence perceptual grouping.

INSERT FIGURE 17 ABOUT HERE

Given our use of simple, non-categorically defined stimuli, our results suggest that capacity estimates from any paradigm in excess of a 3-item limit may involve perceptual grouping or recruitment of other processes that artificially inflate capacity estimates. Therefore, estimates of the capacity of vSTM that exceed a 3-item limit need to be evaluated for the potential contributions of such strategies and perceptual grouping. While perceptual grouping may allow information from multiple objects to be combined into a single slot in vSTM, the 3-item limit of vSTM appears inescapable.

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.

Beck, D.M. & Palmer, S.E. (2002). Top-down influences on perceptual grouping. Journal of Experimental Psychology: Human Perception and Performance, 28(5), 1071-1084.

Belopolsky, A. V., Theeuwes, J., & Kramer, A. F. (in press). Prioritization by transients in visual search. Psychonomic Bulletin & Review.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87-185.

Ericsson, K.A., Chase, W.G., & Faloon, S. (1980). Acquisition of a memory skill. Science, 208, 1181-1182.

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.

Feigenson, L. & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. Cognition, 91, 173-190.

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.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specified information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 9*, 761-768.

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.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object-files: Object specific integration of information. Cognitive Psychology, 24, 175-219.

Kimchi, R. & Hadad, B.S. (2002). Influence of past experience on perceptual grouping. Psychological Science, 13(1), 41-47.

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

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Mitroff, S. R., & Simons, D. J. (in press). Nothing compares 2 views: Change blindness results from failures to compare retained information. *Perception & Psychophsyics*.

Oliva, A. & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. Cognitive Psychology, 34, 72-107.

Palmer, S.E. & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. Psychonomic Bulletin & Review, 1(1), 29-55.

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.

Phillips, W. A., & Christie, D. F. (1977). Components of visual memory. Quarterly Journal of Experimental Psychology, 29(1), 117-133.

Purdy, J., Eimann, D. & Cross, H. (1980). Persistence of a briefly presented visual stimulus in sensory memory. Bulletin of the Psycho-nomic Society, 16, 374-376.

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. (1997). How much of a scene is seen? The role of attention in scene perception. Investigative Opthalmology and Visual Science, 38, 707.

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. (2000b). The dynamic representation of scenes. Visual Cognition, 7, 17-42.

Rensink, R. A. (2000c). Seeing, sensing and scrutinizing. Vision Research, 40, 1469-1487.

Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1995). Image flicker is as good as sacades in making large scene changes invisible. Perception, 24(Suppl.), 26-27.

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.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes cross brief interruptions. *Visual Cognition, 7*, 127-146.

Rock, I. & Brosgole, L. (1964). Grouping based on phenomenal proximity. Journal of Experimental Psychology, 67, 531-538.

Rock, I., Nijhawan, R., Palmer, S. & Tudor, L. (1992a). Grouping and lightness. Chapter in Rock I. (Ed) Indirect Perception. MIT Press/Bradford Bookseries in Cognitive Psychology.

Rock, I., Nijhawan, R., Palmer, S. & Tudor, L. (1992b). Grouping based on phenomenal similarity of achromatic color. Perception, 21(6), 779-789.

Scholl, B. J. (2001). Objects and attention: the state of the art. Cognition, 80, 1-46.

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. (2000). Current approaches to change blindness. Visual Cognition, 7(1/2/3), 1-15.

Simons, D. & Levin, D. (1997) Change blindness. Trends in Cognitive Sciences, 1, 261-267.

Simons, D. J., Chabris, C. F., Schnur, T. T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition, 11*, 78-97.

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

Suzuki, S. & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. Journal of Experimental Psychology: Human Perception and Performance, 21(4), 901-913.

Swain, M. J. & Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7, 11-32.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12, 97-136.

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.

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

Wertheimer, M. (1955). Gestalt theory. In W.D. Ellis (Ed.), A source book of Gestalt psychology, (pp. 1-16). London: Routledge & Kegan Paul. (Original work published 1923).

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin &

Review, Vol 1(2), 202-238.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Postattentive vision. Journal of Experimental Psychology: Human Perception & Performance. Vol 26(2), 693-716.

Wolfe, J. M; Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. Vision Research, Vol 37(1), 25-43.

Figure Captions

<u>Figure 1</u>. The trial structure for Experiment 1. The cycle of the displays repeated until the subject pressed. In this Figure, the changing dot is located in the fourth column from the left and in the third row from the bottom.

<u>Figure 2</u>. Mean response time (RT) in ms (±SE) for the 5 display durations in Experiment 1. From 300-700ms performance response times do not vary with display duration, suggesting that processing is information-limited. After 700ms, additional display time leads to a linear increase in response time, suggesting that subjects have reached their storage capacity.

<u>Figure 3</u>. Mean estimated capacity (±SE) for the 5 display durations in Experiment 1. Estimated capacity for display times of 300ms and 500ms are significantly below all other display times whereas estimated capacity is constant for display times ranging from 700 to 1100ms. This asymptote corresponds to an estimate of 2.5 objects.

<u>Figure 4</u>. Estimated capacity (±SE) for the 5 array sizes in Experiment 2. Estimated capacity was approximately 2.5 items at every array size. This estimate was lower for array size 4 (i.e. 2.3) because the mean response time for array size 4 was close to floor, artificially decreasing the estimated capacity for these trials.

<u>Figure 5</u>. Estimated capacity for successive blocks of 10 trials in Experiment 3, in which subjects saw a new display on each trial (Variable Display) or the same display for trials 1-50 and another display for trials 51-100 (Constant Display). As in Experiments 1-2, capacity was estimated to be approximately 2.5 objects. Increased familiarity with a constant display did <u>not</u> lead to an increased capacity for this array.

Figure 6. The memorized grid used throughout memorization and testing in Experiments 4-5. During memorization, G.P. and J.H. spent a total of approximately 30 min. viewing the complete array and approximately 9.5 hours viewing the array one dot at a time, arranged by position.

Figure 7. Memorization accuracy for subjects G.P. and J.H. for the 24 memorization sessions of Experiment 4. Accuracy increased for both subjects throughout the Easy Sessions. Accuracy was above chance for both subjects throughout the Difficult Sessions (chance = 50%).

Figure 8. The change in luminance used for the Difficult Sessions (+2) and the change detection task (+14) in Experiments 4-5. The top grid is the one that G.P. and J.H. memorized (memorized grid). On the bottom-left, this same grid is shown with one dot changed by 2 luminance values (the level of specificity required in the Difficult Sessions). On the bottom-right this same dot is shown changed by 14 luminance values (the level of specificity required by the change detection task). The changed dot is in the fourth column from the left and the second row from the bottom.

Figure 9. Subject G.P.'s capacity (i.e. the number of items held in vSTM across a blank interval) for the memorized grid and randomized grids (±SE) for the 8 successive sessions in Experiment 5.

Figure 10. Subject J.H.'s capacity (i.e. the number of items held in vSTM across a blank interval) for both the memorized grid and randomized grids (±SE) for the 8 successive Sessions in Experiment 5.

Figure 11. The cluster analysis for each position in the memorized grid from Experiments 4-5. Cluster rating equals the number of dots that are "connected" to the target dot and are similar in luminance to the target dot divided by 35 (the total number of distractors). The resulting cluster for

a sample dot is outlined in white. For a grid with 36 dots, having 1 similar neighbor resulted in a cluster rating of .03, two similar neighbors in a rating of .06, etc.

Figure 12. Subject J.H.'s average capacity for detecting a changing dot in the memorized grid in Experiment 5. Overlaid on top of each location in the memorized grid is the estimated capacity for visual search when that location was the changing target.

Figure 13. Subject J.H.'s average capacity for detecting a changing dot in the memorized grid in Sessions 1-8 of Experiment 5 displayed as a function of the cluster rating of the changing target. The significant positive linear regression suggests that subject J.H. had chunked the memorized grid into perceptual units that were each larger than a single dot. For dots that were singletons (cluster rating = 0), J.H. had an estimated capacity of approximately 3 items. This suggests that vSTM is limited to representing up to 3 individuals, though these individuals may be single dots or larger perceptual groups of dots.

Figure 14. Subject G.P.'s average capacity for detecting a changing dot in the memorized grid in Sessions 1-8 of Experiment 5 displayed as a function of the cluster rating of the changing target. The significant linear regression is consistent with G.P.'s chosen grouping strategy of focusing on contrasts within a row. G.P. was faster to detect targets that were dissimilar with their surroundings (cluster rating = 0). Importantly, when this grouping strategy was sub-optimal (cluster rating = .26), G.P. showed the predicted limit of 3 items.

Figure 15. Subject J.H.'s average capacity for detecting a changing dot in randomized grids in Sessions 5-8 of Experiment 5 displayed as a function of the cluster rating of the changing target.

Figure 16. Subject G.P.'s average capacity for detecting a changing dot in randomized grids in Sessions 5-8 of Experiment 5 displayed as a function of the cluster rating of the changing target.

Figure 17. The architecture suggested by the present results.

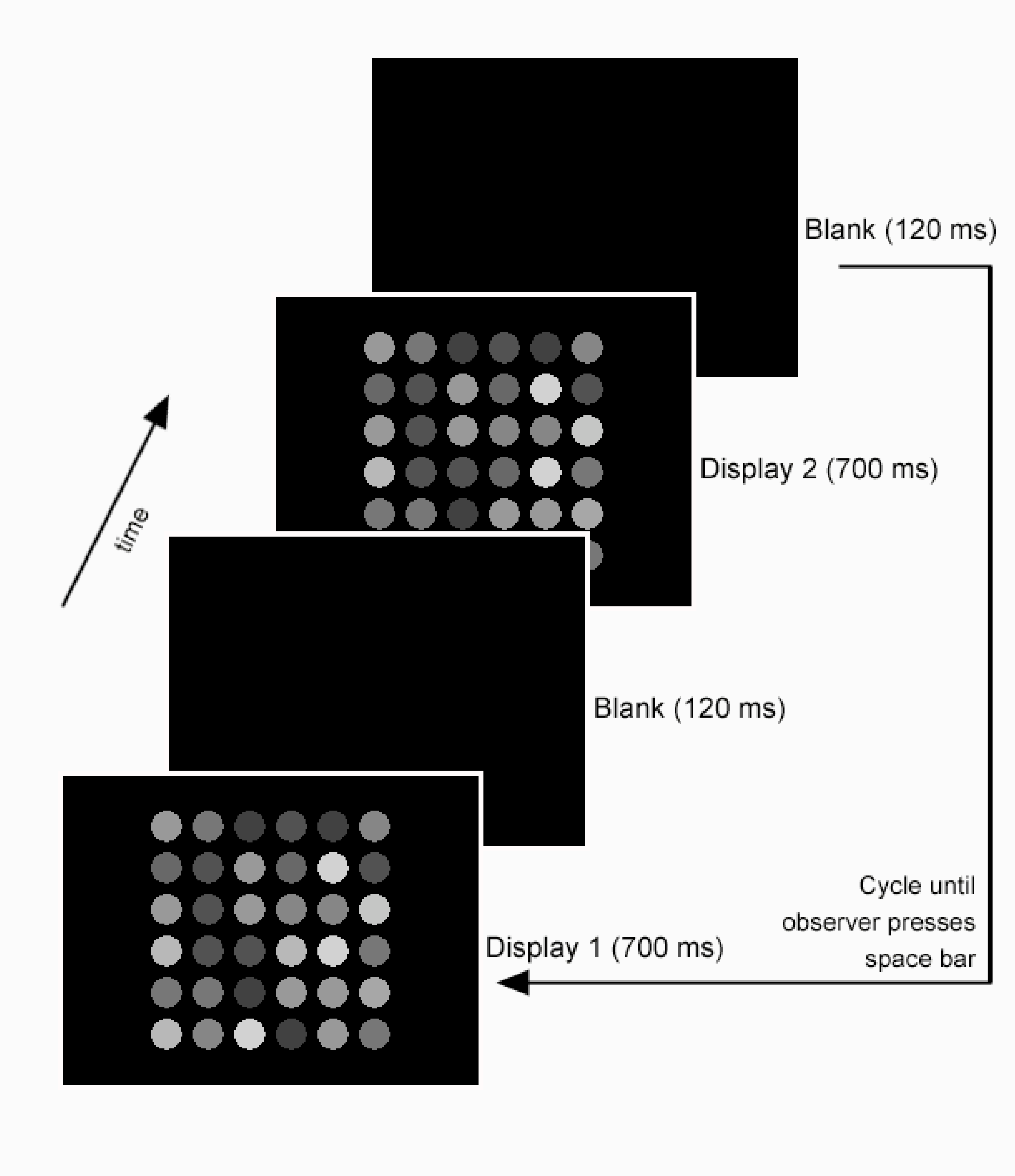

Blank (120 ms)
Display 2 (700 ms)
time
Blank (120 ms)
Cycle until
observer presses
space bar
Display 1 (700 ms)

Figure 1

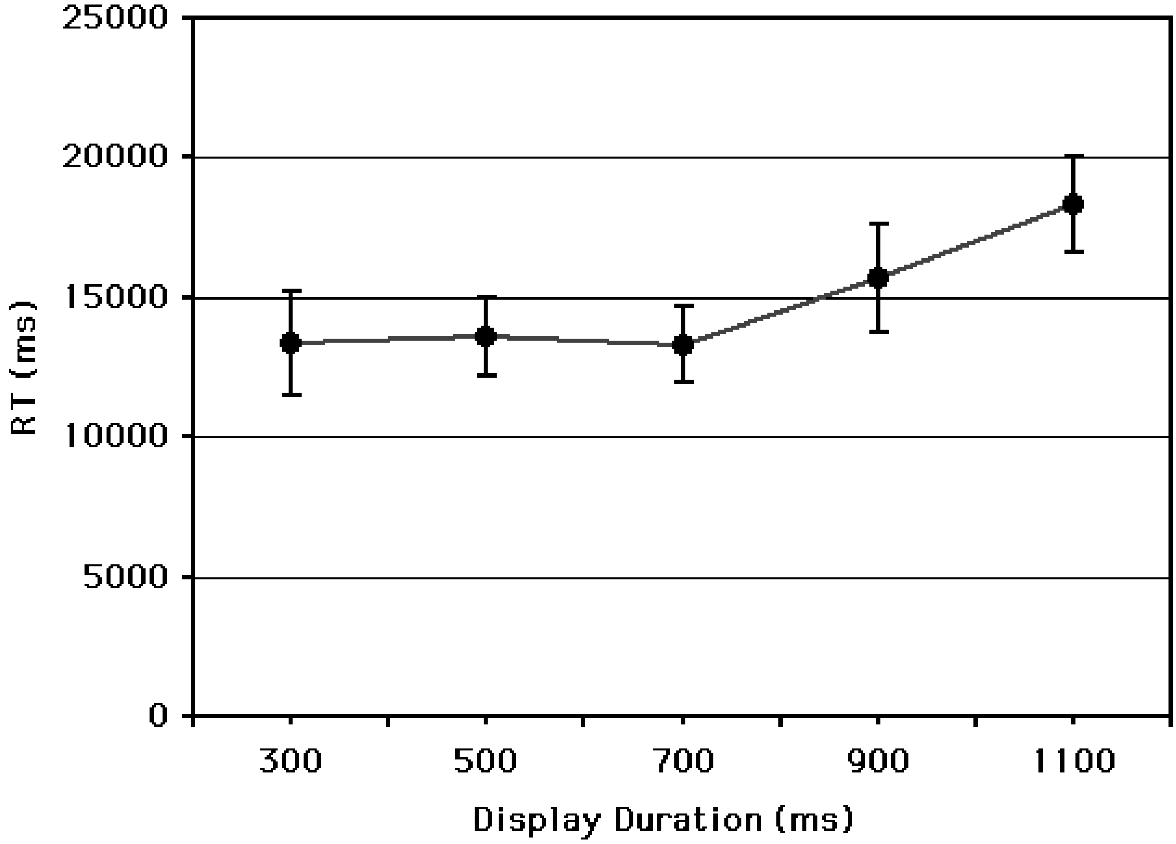
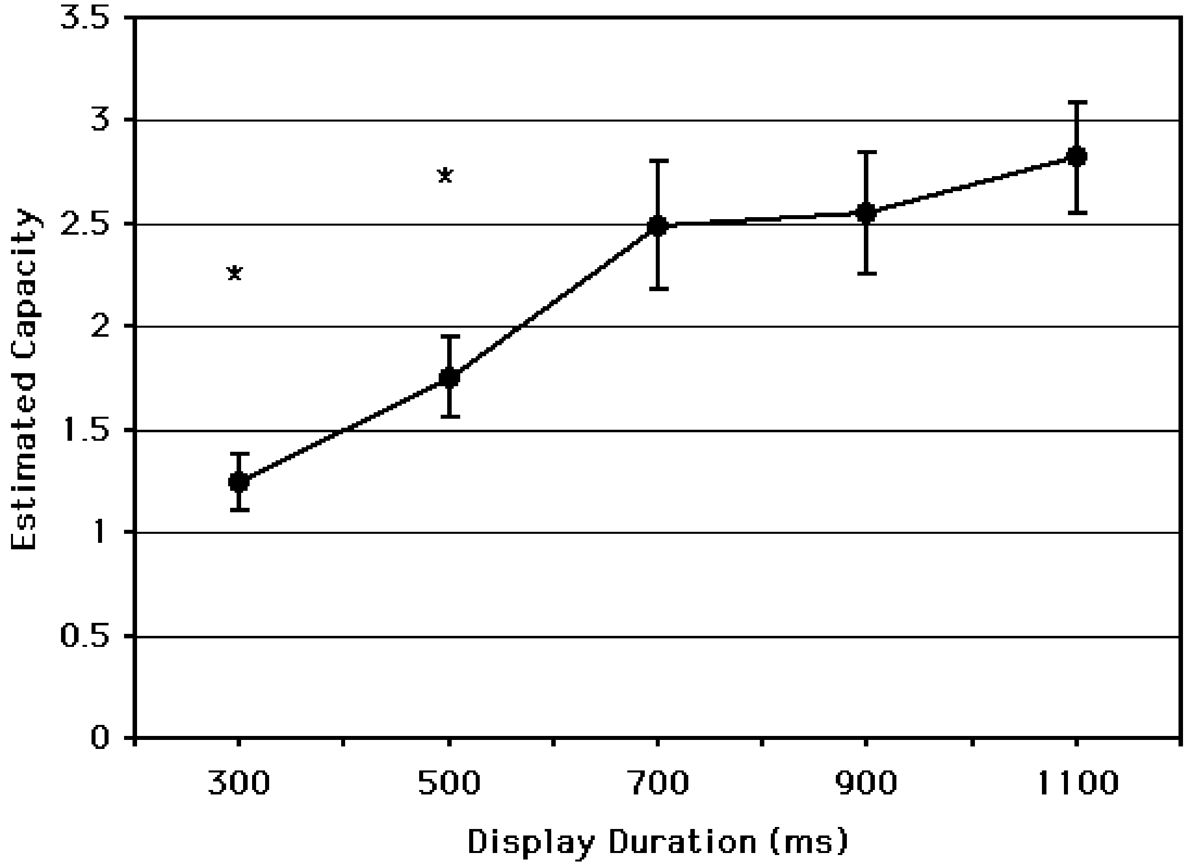

25000
20000
15000
10000
5000
0
RT (ms)
300
500
700
900
1100
Display Duration (ms)

Figure 2



Estimated Capacity
3.5
3
2.5
2
1.5
1
0.5
0
*
*
300
500
700
900
1100
Display Duration (ms)

Figure 3

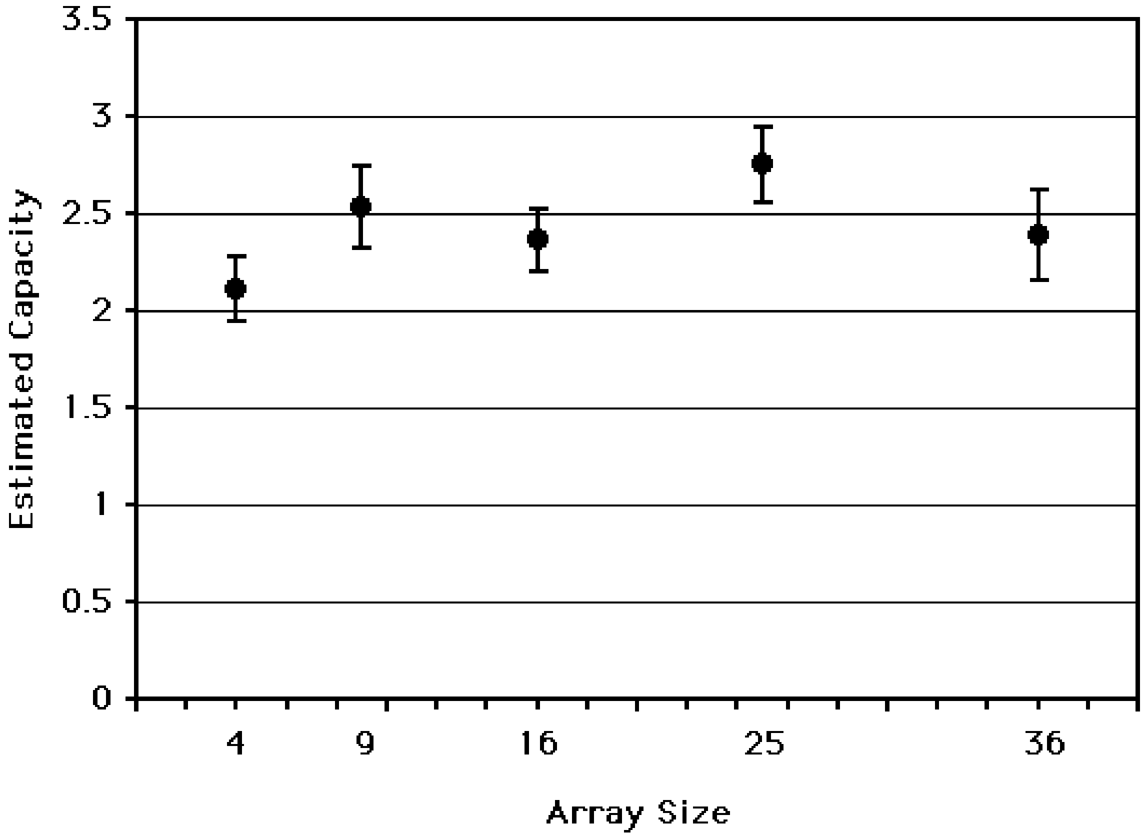
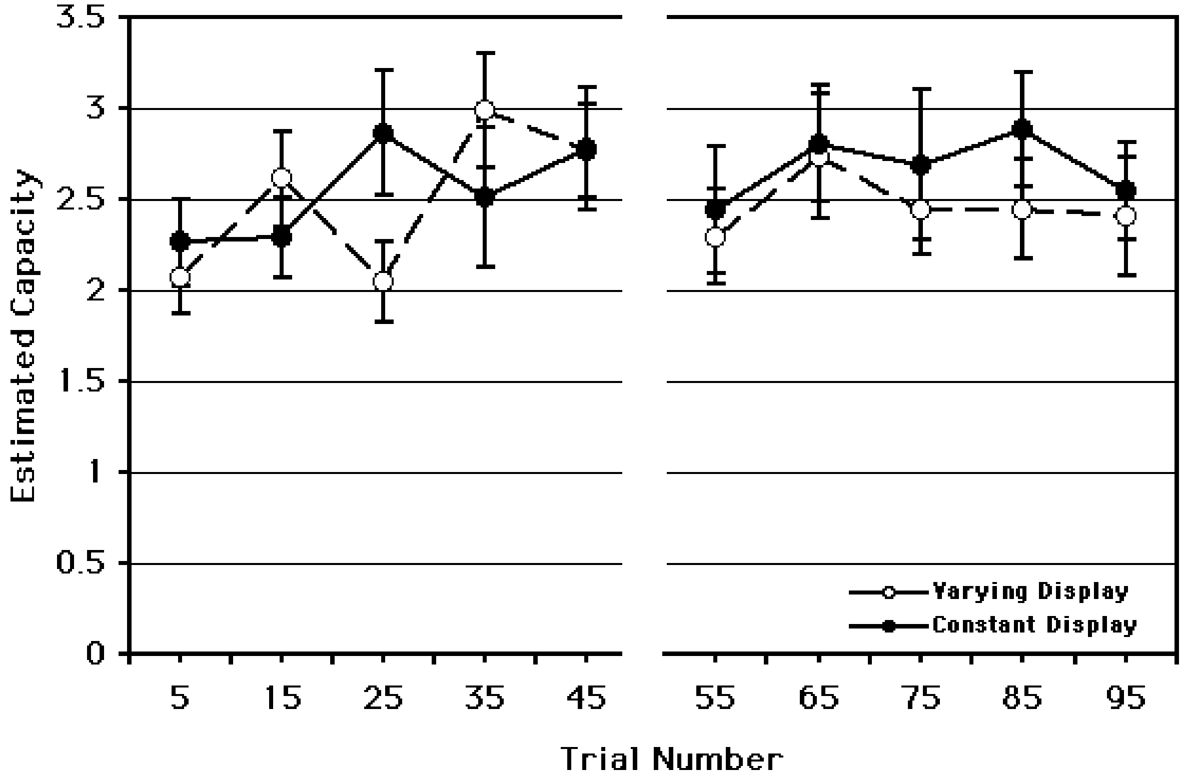


Figure 4



Estimated Capacity
3.5
3
2.5
2
1.5
1
0.5
0
5
15
25
35
45
55
65
75
85
95
Trial Number
Varying Display
Constant Display

Figure 5

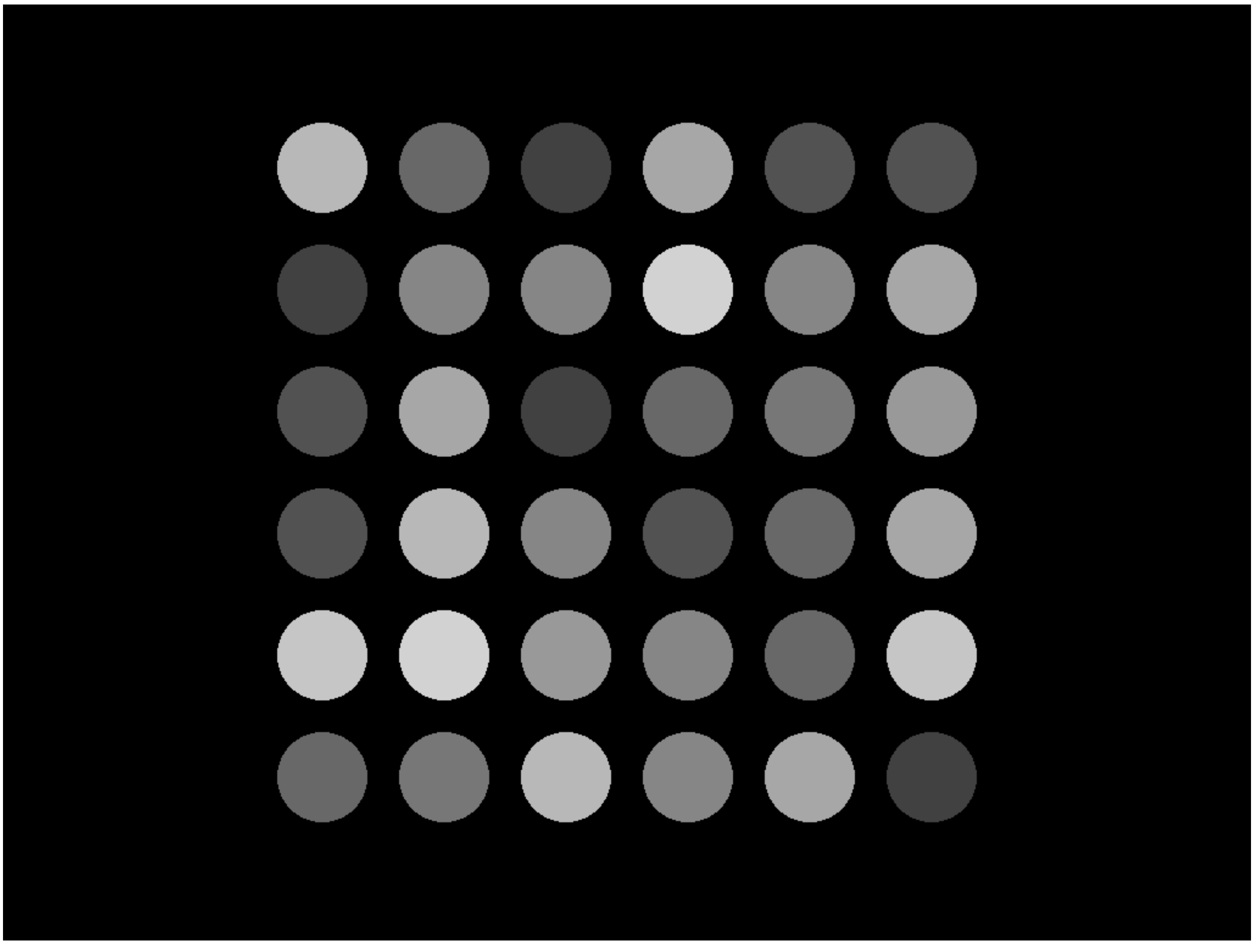


Figure 6

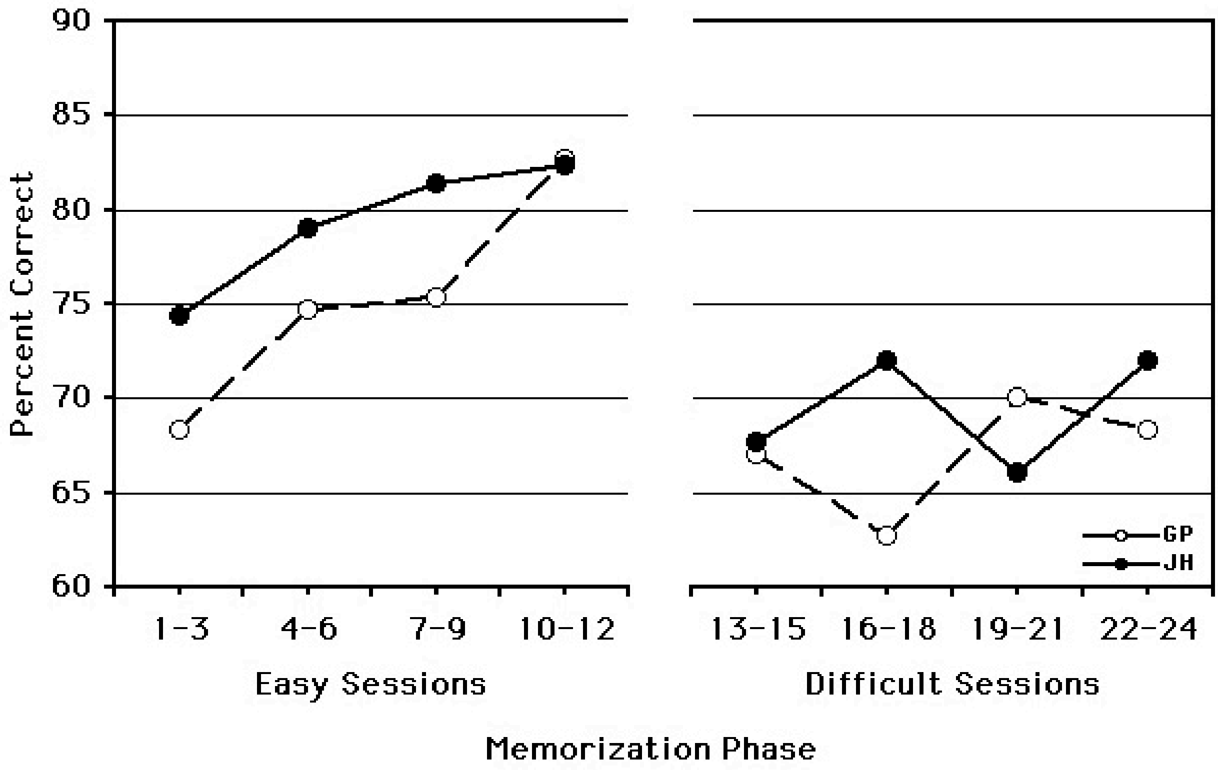

90
85
80
75
70
65
60
Percent Correct
1-3
4-6
7-9
10-12
13-15
16-18
19-21
22-24
Easy Sessions
Difficult Sessions
Memorization Phase
GP
JH

Figure 7

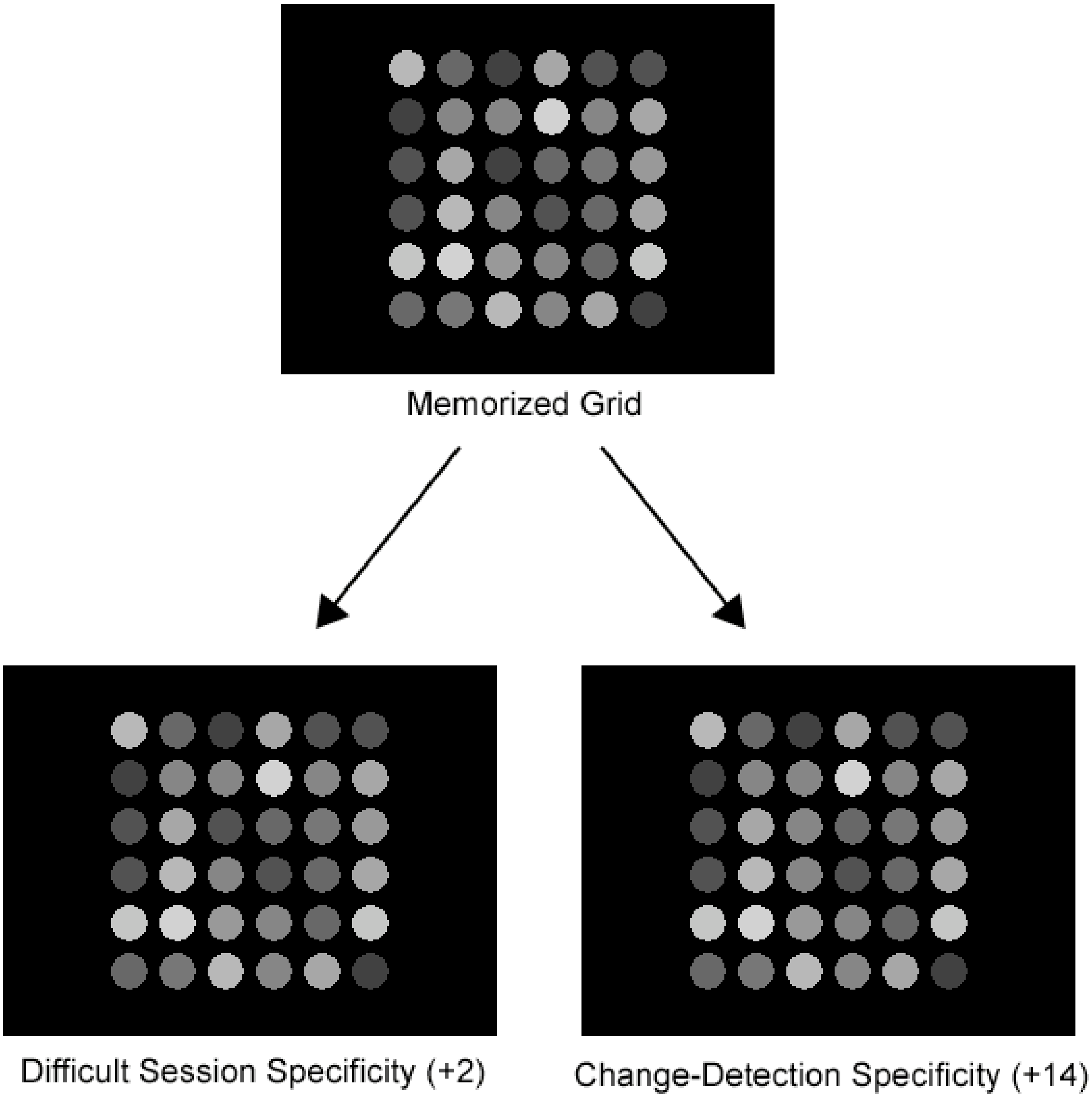

Memorized Grid
Difficult Session Specificity (+2)
Change-Detection Specificity (+14)

Figure 8

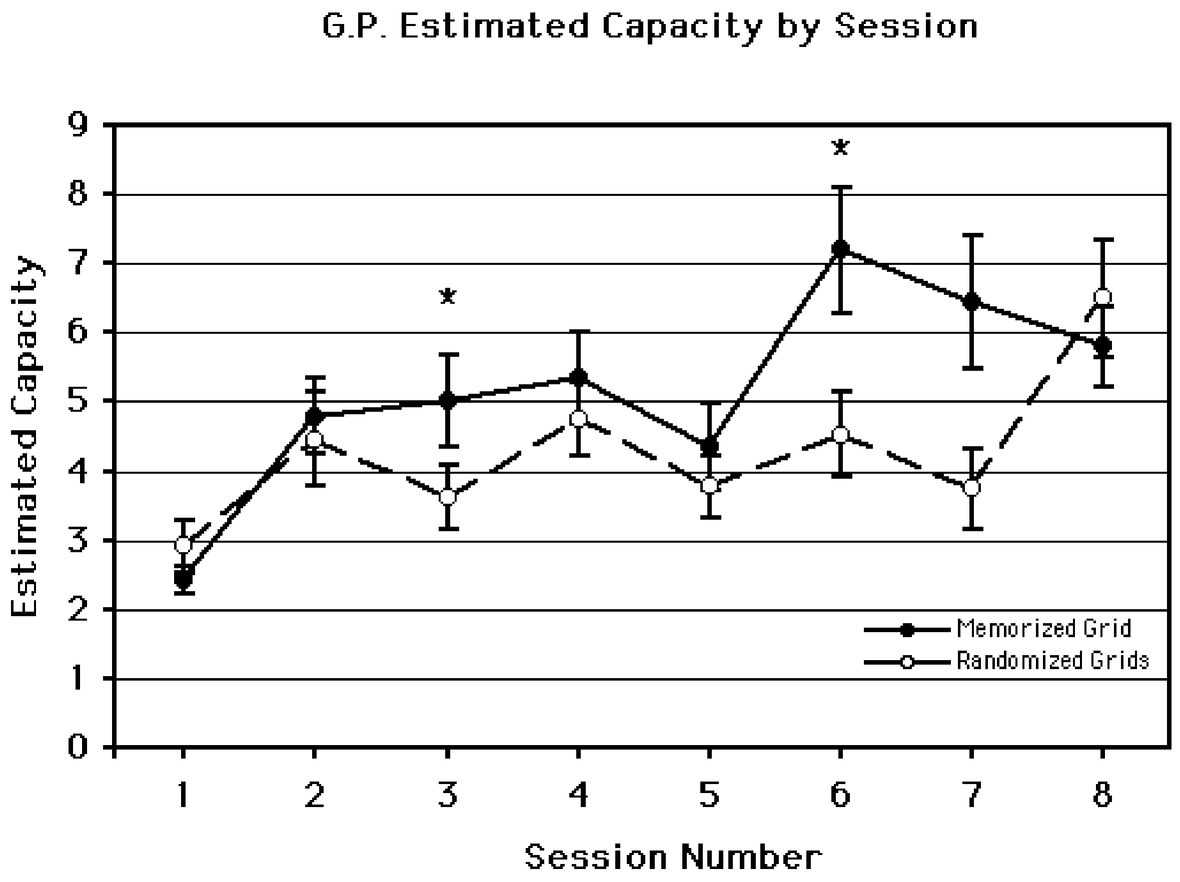

G.P. Estimated Capacity by Session
Estimated Capacity
0
1
2
3
4
5
6
7
8
9
*
*
Memorized Grid
Randomized Grids
1
2
3
4
5
6
7
8
Session Number

Figure 9

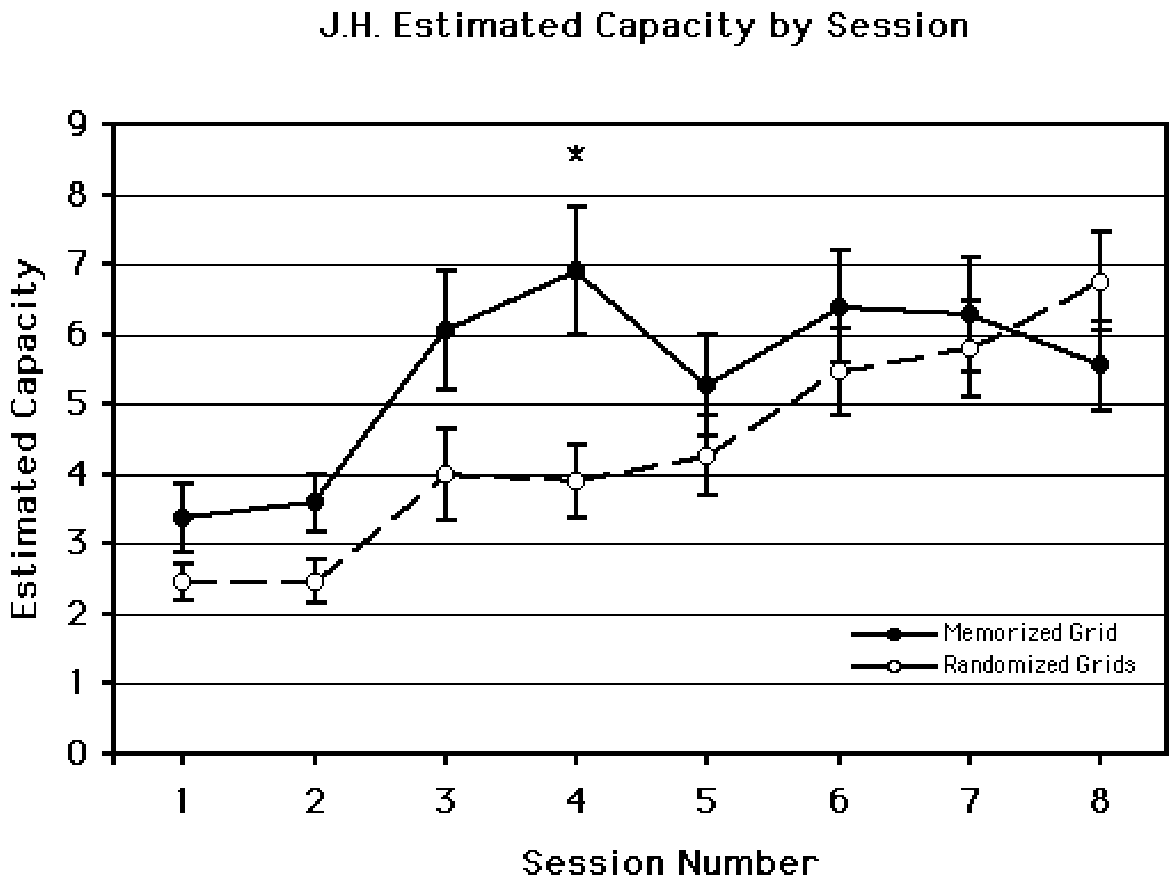

J.H. Estimated Capacity by Session
*
Estimated Capacity
0
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
Session Number
Memorized Grid
Randomized Grids

Figure 10

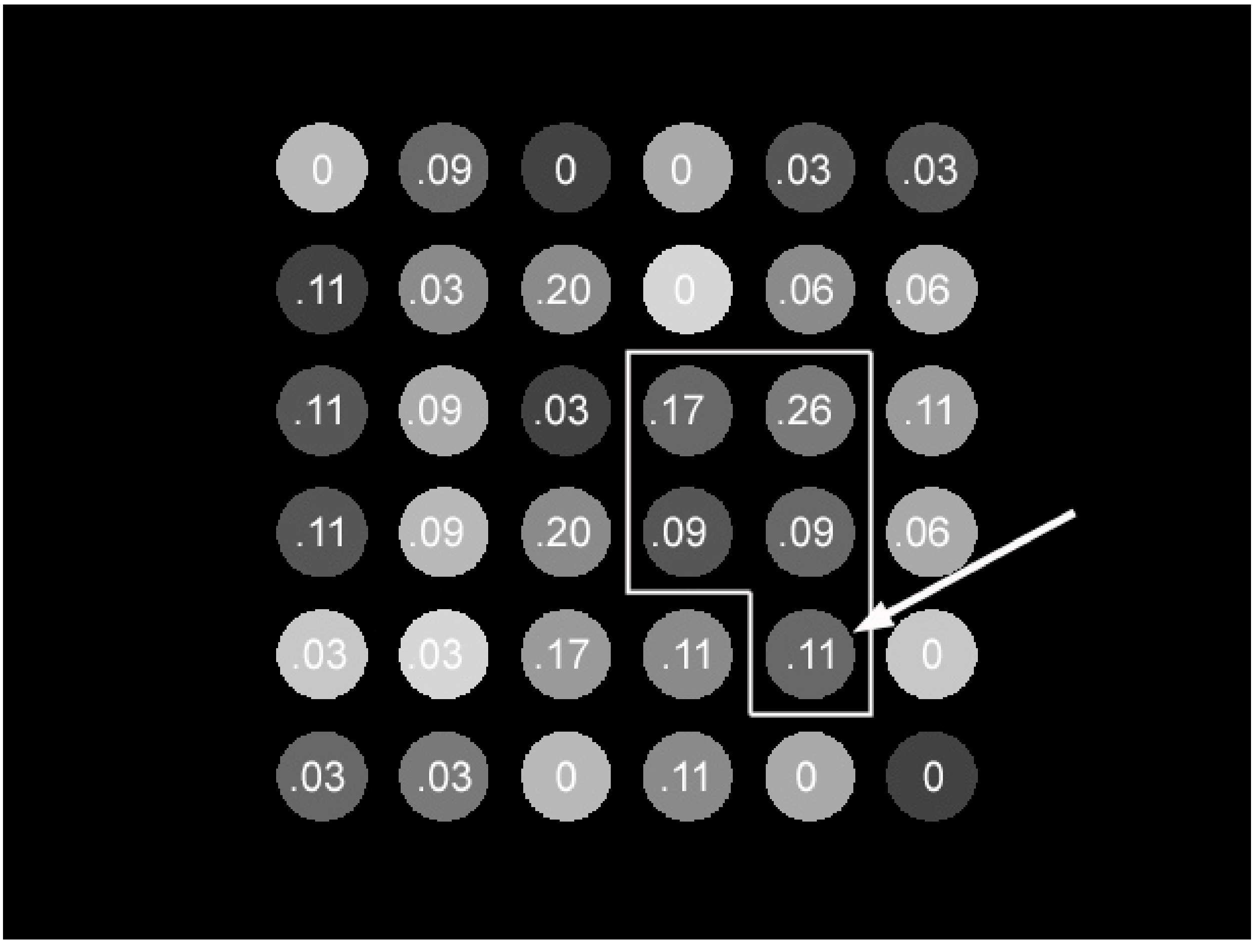

0
.09
0
0
.03
.03
.11
.03
.20
0
.06
.06
.11
.09
.03
.17
.26
.11
.11
.09
.20
.09
.09
.06
.03
.03
.17
.11
.11
0
.03
.03
0
.11
0
0

Figure 11

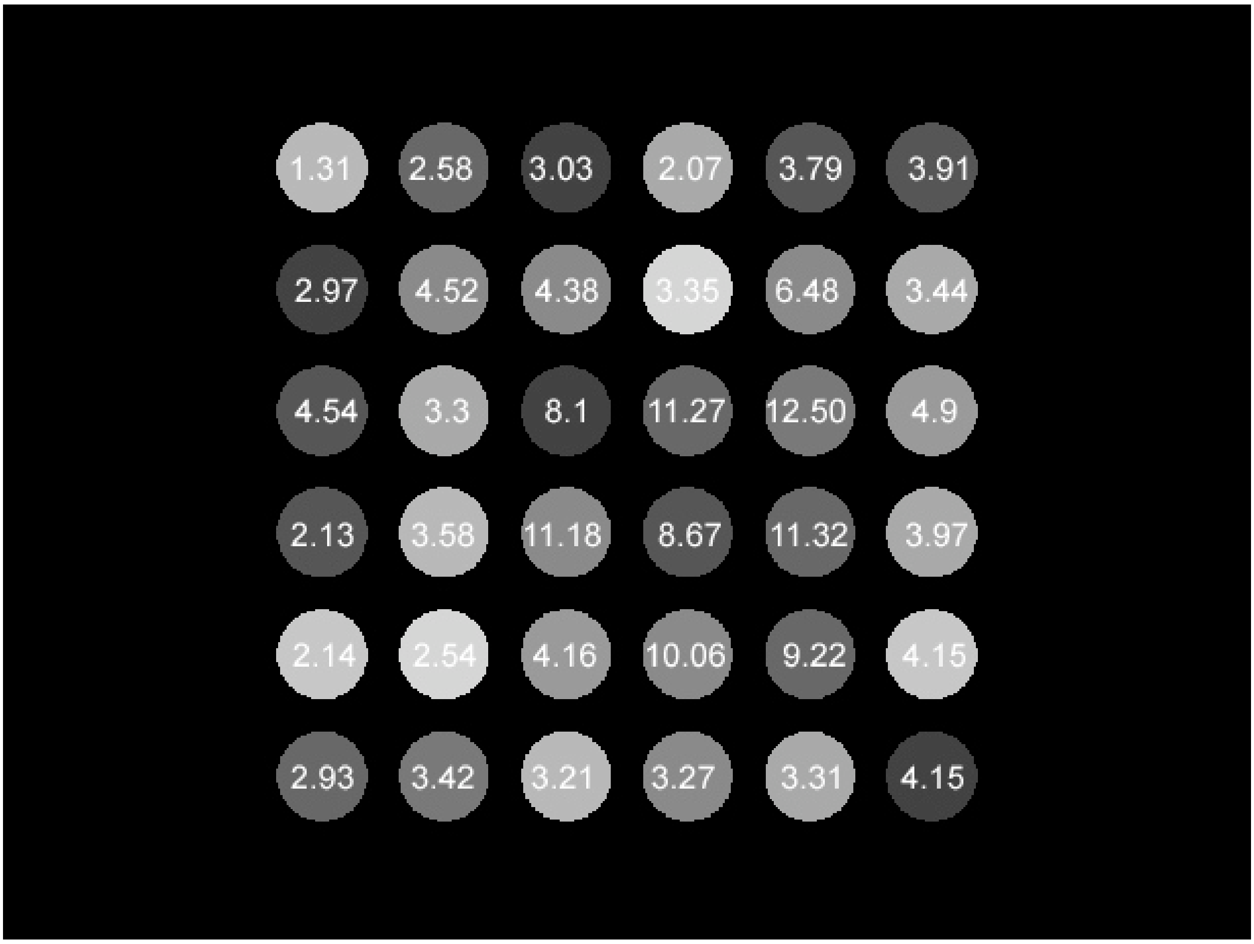
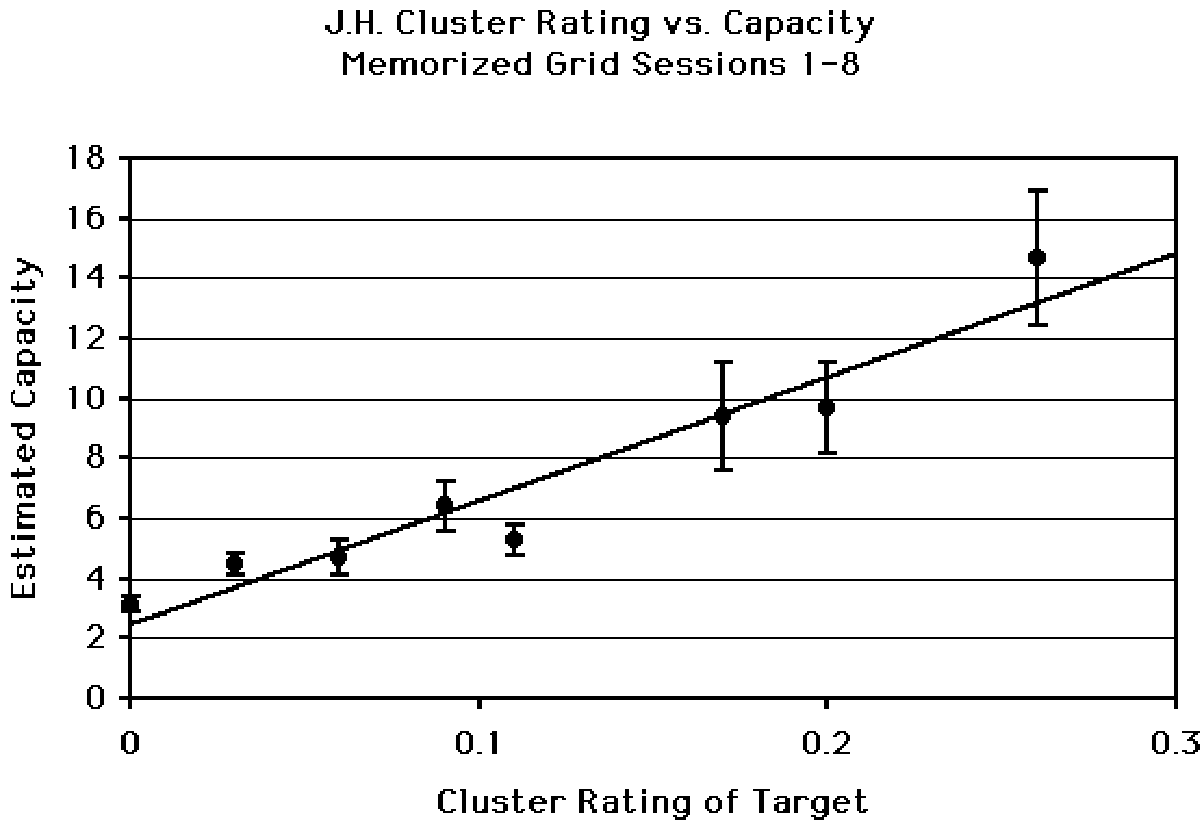

1.31
2.58
3.03
2.07
3.79
3.91
2.97
4.52
4.38
3.35
6.48
3.44
4.54
3.3
8.1
11.27
12.50
4.9
2.13
3.58
11.18
8.67
11.32
3.97
2.14
2.54
4.16
10.06
9.22
4.15
2.93
3.42
3.21
3.27
3.31
4.15

Figure 12



J.H. Cluster Rating vs. Capacity
Memorized Grid Sessions 1-8
Estimated Capacity
0
2
4
6
8
10
12
14
16
18
0
0.1
0.2
0.3
Cluster Rating of Target

Figure 13

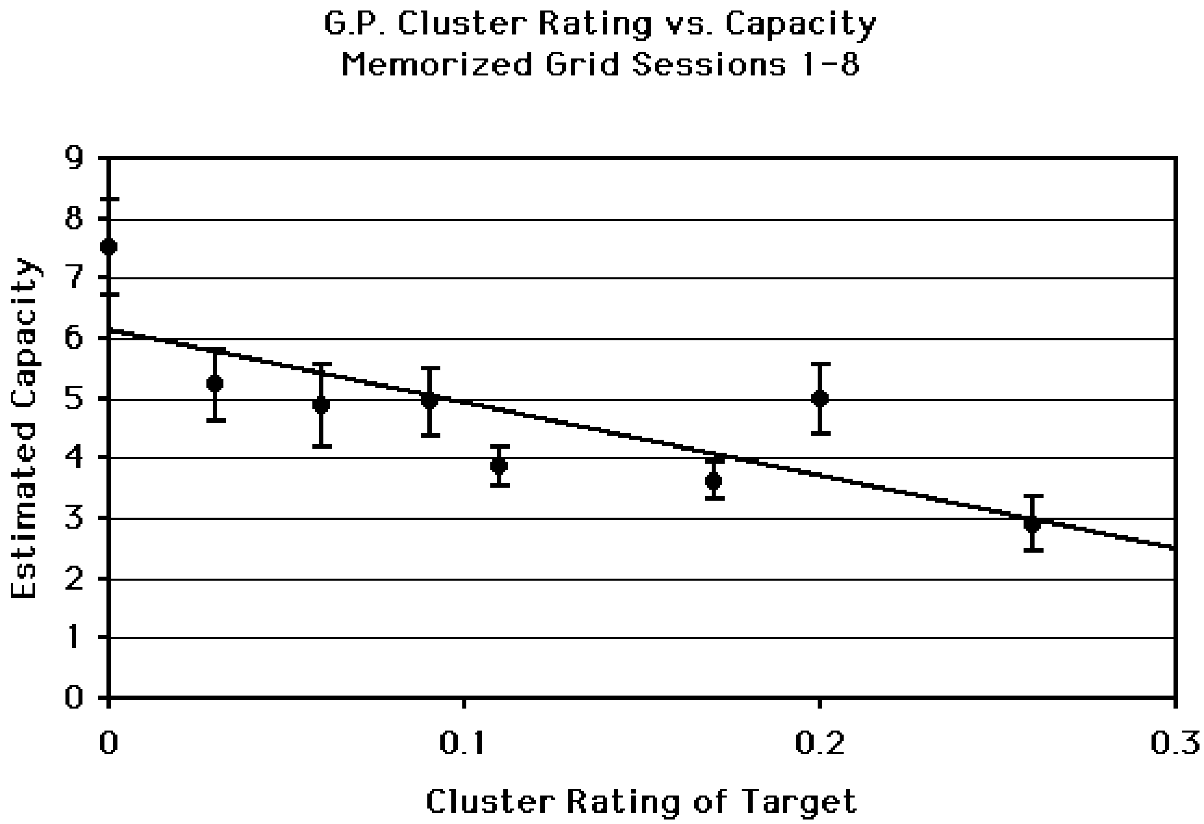

G.P. Cluster Rating vs. Capacity
Memorized Grid Sessions 1-8
Estimated Capacity
9
8
7
6
5
4
3
2
1
0
0
0.1
0.2
0.3
Cluster Rating of Target

Figure 14

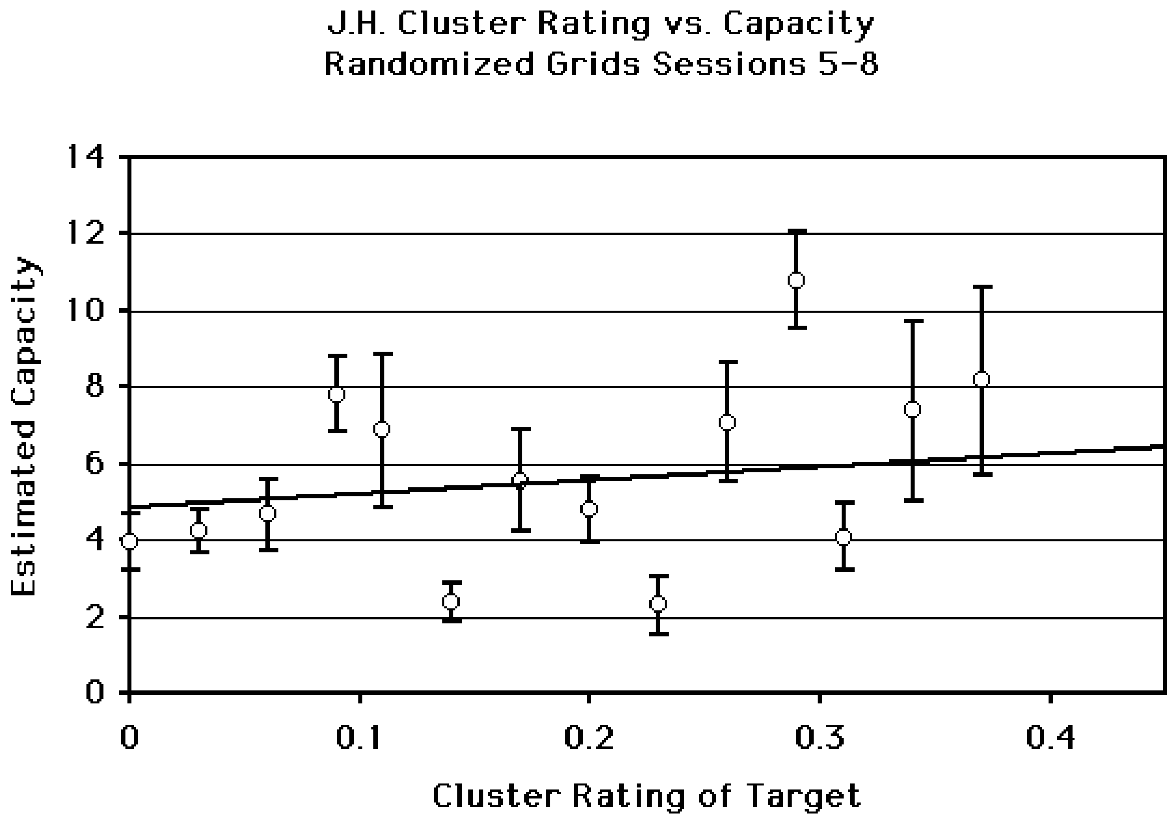

J.H. Cluster Rating vs. Capacity
Randomized Grids Sessions 5-8
Estimated Capacity
0
2
4
6
8
10
12
14
0
0.1
0.2
0.3
0.4
Cluster Rating of Target

Figure 15

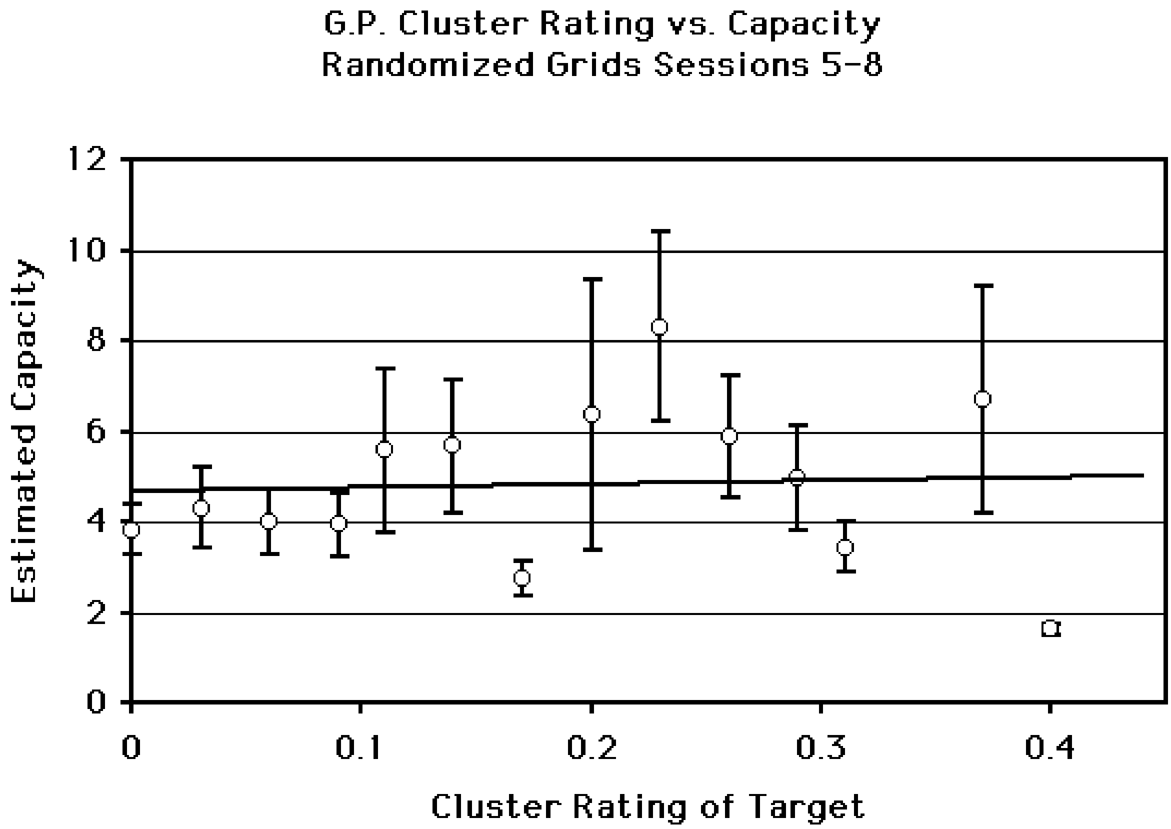
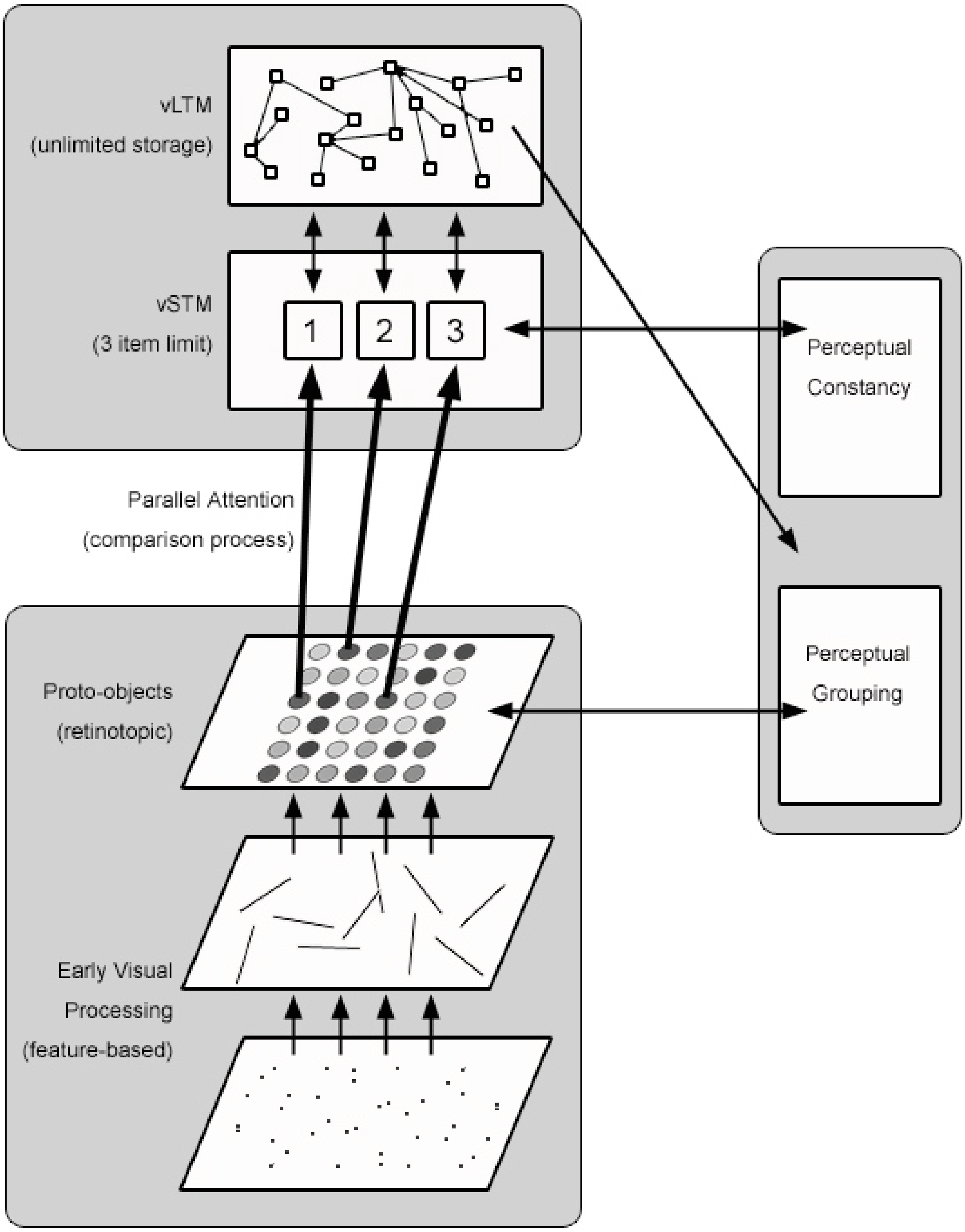

G.P. Cluster Rating vs. Capacity
Randomized Grids Sessions 5-8
Estimated Capacity
12
10
8
6
4
2
0
0
0.1
0.2
0.3
0.4
Cluster Rating of Target

Figure 16



vLTM
(unlimited storage)
vSTM
(3 item limit)
1
2
3
Perceptual
Constancy
Parallel Attention
(comparison process)
Perceptual
Grouping
Proto-objects
(retinotopic)
Early Visual
Processing
(feature-based)

Figure 17