Capacity and precision in an animal model of visual short-term memory

Antonio H. Lara

Jonathan D. Wallis

Temporary storage of information in visual short-term memory (VSTM) is a key component of many complex cognitive abilities. However, it is highly limited in capacity. Understanding the neurophysiological nature of this capacity limit will require a valid animal model of VSTM. We used a multiple-item color change detection task to measure macaque monkeys’ VSTM capacity. Subjects’ performance deteriorated and reaction times increased as a function of the number of items in memory. Additionally, we measured the precision of the memory representations by varying the distance between sample and test colors. We modeled the error distribution as a Gaussian function and used this to estimate the precision of VSTM representations. We found that as the number of items in memory increases the precision of the representations decreases dramatically. Additionally, we found that focusing attention on one of the objects increases the precision with which that object is stored and degrades the precision of the remaining. These results are in line with recent findings in human psychophysics and provide a solid foundation for understanding the neurophysiological nature of the capacity limit of VSTM.

Keywords: visual short-term memory (VSTM), capacity, precision, non-human primate


Introduction

It is estimated that humans can hold up to four distinct items in visual short-term memory (VSTM; Cowan, 2001; Luck & Vogel, 1997). The ability to hold multiple pieces of information in mind is paramount to complex cognitive abilities such as learning, problem solving, and language comprehension (Cowan, 2001; Cowan et al., 2005), and it is highly correlated with intelligence (Conway, Kane, & Engle, 2003). In order to understand the neuronal mechanisms underlying this phenomenon, it will be important to establish an animal model of VSTM. Recent studies have determined that non-human primates have a capacity-limited VSTM, with a capacity of about 2 to 3 items (Buschman, Siegel, Roy, & Miller, 2011; Heyselaar, Johnston, & Paré, 2011; Siegel, Warden, & Miller, 2009), raising the possibility that they may be a suitable animal model.

If non-human primates are to be used as animal models to study capacity limits in VSTM, then it is important to establish that the process behaves similarly in humans and monkeys. Recent work in human psychophysics has revealed that VSTM can be modeled as a limited pool of resources that is shared between items in memory (Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Huang, 2010; Wilken & Ma, 2004) and suggests that the capacity limit arises from an inability to store memories with enough precision to enable correct recall (but see Anderson, Vogel, & Awh, 2011; Zhang & Luck, 2008 for an alternate model).

Here, we describe experiments that we carried out to determine the nature of the capacity limit in rhesus macaque monkeys and how the precision of stored information changes as the capacity limit is approached. We adapted a color change detection task widely used in human psychophysical research (Luck & Vogel, 1997). In this task, an array of colored squares is presented followed by a short delay and subjects have to determine whether one of the squares changed color. We modified the task in a way that allowed us to determine the precision of the stored representations. By parametrically varying the color difference between the sample and test colors, we determined how the precision of stored representations changes as VSTM is increasingly taxed. This strategy is analogous to the one used by Bays et al. in order to probe how the VSTM system in human subjects breaks down as it reaches the limit of its capacity (Bays et al., 2009; Bays & Husain, 2008). If monkeys are to be a valid model of VSTM, then they too should show a reduction in the precision of stored information as the number of items in VSTM increases.

Methods

Subjects

Subjects were two male rhesus monkeys (Macaca mulatta) aged 4 to 5 years and weighing 11 to 13 kg at
the time of the experiments. Subjects’ daily fluid intake was regulated in order to maintain motivation to perform the task. During testing, subjects sat in a primate chair facing a 19-inch LCD computer screen placed at a distance of 32 cm. A pair of computers running NIMH Cortex (http://www.cortex.salk.edu) controlled the timing and presentation of stimuli. We monitored eye movements using an infrared camera with ISCAN software. All procedures were in accord with the National Institute of Health guidelines and the recommendations of the U.C. Berkeley Animal Care and Use Committee.

### Multiple-item change detection task

Subjects were trained on a color change detection task adapted from the human literature (Luck & Vogel, 1997), as illustrated in Figure 1. At the start of the trial, a fixation square (0.5° × 0.5° of visual angle) appeared in the center of the screen. Subjects had to maintain their gaze within 1.5° of the center of the fixation spot for 1000 ms. After subjects achieved fixation, a sample array of 1 to 4 different colored squares appeared on the screen for 500 ms. At the end of the 500-ms sample period, the array disappeared from the screen for 1000 ms. During this time, subjects had to keep the color of all the squares in VSTM. At the end of the delay period, one of the squares in the array was presented again and the subject had to move a lever upward if the color of the square had changed or move the lever downward if the color had remained the same. Subjects were free to respond as soon as the test square appeared on the screen. Correct responses were rewarded with 0.5 ml of orange juice delivered directly to the subjects’ mouth. If the subject made an incorrect response, the screen turned gray for 4 s to indicate an incorrect response and no reward was given. If at any time during the sample or the delay periods subjects broke fixation, the entire screen turned red for 10 s, after which a new trial started. There was a 3000-ms intertrial interval between all trials.

### Stimuli

Stimuli were squares 3.5° × 3.5° of visual angle. They were presented in any of four fixed locations 5° away from the fixation spot on a black background. Colors were chosen from the 1976 CIE L a* b* color space. We fixed the luminance at L = 70 and varied the a* and b* parameters to produce 24 different colored squares that span a rectangle in this cross-section of the color space (Figure 1, inset). All 24 colors could be used both as sample and as test colors. The distance between two colors in this color space is given by

$$d = \sqrt{(a_j - a_i)^2 + (b_j - b_i)^2},$$

where \((a_j, b_j)\) and \((a_i, b_i)\) are the parameters for colors 1 and 2, respectively. We avoided potential similarity between the colors in the sample array by enforcing a lower bound on the distance between all the colors \(i\) and \(j\) in the array to \(d_{ij} \geq 50\). Additionally, we restricted the distance between the sample and the test colors \((\Delta E)\) to the set \(\Delta E \in (0, 40, 50, 60, 70, 80, 90, 100)\).

### Model comparison

To estimate the precision of VSTM representations, we calculated the probability of subjects detecting a change in color as a function of distance between sample and test colors. This probability is modeled by (Green & Swets, 1966)

$$p(\text{responding change}|\Delta E) = \Phi \left( \frac{\Delta E - \mu}{\sigma} \right),$$

where \(\Phi\) is the cumulative distribution function of a normal distribution with mean \(\mu\) and variance \(\sigma^2\). Equation 1 assumes that subjects have a noisy internal representation of change and will respond “change” whenever the internal representation exceeds a criterion given by \(\mu\). We incorporated a guessing parameter \(g\) into this model to account for the fact that on some proportion of trials subjects might fail to detect a change and instead guess that a change has occurred. On a certain proportion

![Figure 1. Visual change detection task. Trials began with 1 s of fixation. An array of 1–4 colored squares appeared for 500 ms. A test color appeared after a 1000-ms delay. The distance between sample colors and the test color was varied parametrically. In this example, the test color can be chosen from any of the row of colors shown at the bottom of the figure. The inset shows the subset of the CIE L a* b* color space used in the task.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932803/)
of trials, \( g \), the subject randomly guesses, while on the remaining proportion of trials, \( 1 - g \), the subject’s behavior can be described by the model given by Equation 1. This leads to the second model:

\[
p(\text{reporting change}|\Delta E) = (1 - g)\Phi((\Delta E - \mu) / \sigma) + \frac{g}{2}. \tag{2}
\]

We also examined a third model. Non-human primates are rarely willing to work at a continuous level of high performance for an extended length of time. Instead, their performance across a single testing session is typically characterized by periods of high performance and periods of low performance. In other words, there are times when they are not fully engaged in the task. Furthermore, during these periods, their responding will frequently be biased to one or other of the responses. Such periods of response bias are a common feature of non-human primate behavior across a range of tasks. To account for these biases, we added an additional parameter, \( b \), which captures any bias that subjects may have to moving the lever in one direction:

\[
p(\text{reporting change}|\Delta E) = (1 - b)\left[(1 - g)\Phi\left(\frac{\Delta E - \mu}{\sigma}\right) + \frac{g}{2}\right] + \frac{b}{2}. \tag{3}
\]

This model can be seen as a mixture of two states such that when the subjects are engaged in the task, performance is modeled by Equation 2 and when subjects are not engaged in the task, the probability of responding “change” (moving the lever in one direction) is given by the bias parameter.

We used maximum likelihood estimation (MLE) to estimate the model parameters and calculated the Akaike information criterion (AIC) for each model to assess the relative goodness of fit of each model. We fit each model separately for each set size and estimated the precision of stored representations as \( 1/\sigma^2 \). To obtain reliable parameter estimates, we pooled trials together across sessions.

The above analysis allows the precision of stored representations to vary as a function of set size, analogous to “resource” models designed to explain psychophysical measures of VSTM in humans (Bays et al., 2009; Bays & Husain, 2008; Huang, 2010; Wilken & Ma, 2004). However, we also examined whether our data could be explained by a fixed-precision slots model, which holds that VSTM consists of a number of discrete “slots,” each of which is a fixed precision, \( \sigma_0 \) (Zhang & Luck, 2008). In this model, the probability of responding change when the number of “slots” available, \( k \), is greater or equal or less than the number of items in the display, \( N \), is given by (Cowan & Rouder, 2009):

\[
p(\text{reporting change}|\Delta E) = \begin{cases} 
\Phi \left[ \frac{(\Delta E - \mu)}{\sigma_0} \left( \frac{k}{N} \right)^{1/2} \right], & k \geq N \\
\left( \frac{k}{N} \right) \Phi \left( \frac{\Delta E - \mu}{\sigma_0} \right) + \left( 1 - \frac{k}{N} \right) g, & k < N.
\end{cases} \tag{4}
\]

In this formulation, the slots model states that when \( k \geq N \), subjects store all the items in VSTM with precision \( \sigma_0 \), and if there are more slots than items, then an item can be stored in more than one slot. On the other hand, when \( k < N \), the subject stores only a fraction of the items \( (k/N) \) in VSTM with precision \( \sigma_0 \) and guesses if the probed item is not in memory. Additionally, to allow for a fair comparison with the resources model (Equation 3), we incorporated a bias parameter to account for any bias that the subjects have for moving the lever in one direction. Thus, the full form of the slots model we used is given by

\[
p(\text{reporting change}|\Delta E) = \begin{cases} 
(1 - b)\left[ (1 - g)\Phi\left(\frac{\Delta E - \mu}{\sigma_0}\right) \left( \frac{k}{N} \right)^{1/2} \right] + \frac{b}{2}, & k \geq N \\
(1 - b)\left[ \left( \frac{k}{N} \right) \Phi \left( \frac{\Delta E - \mu}{\sigma_0} \right) + \left( 1 - \frac{k}{N} \right) g \right] + \frac{b}{2}, & k < N.
\end{cases} \tag{5}
\]

**Microsaccade analysis**

During the sample period, subject G tended to make microsaccades in the direction of one or more of the items in the array. We defined a microsaccade as a fixational eye movement whose velocity exceeded 6 standard deviations from mean velocity and did not leave the fixation area. To examine whether these microsaccades affected the precision with which items were stored, we sorted trials according to whether the subject made a microsaccade in the direction of the item that was eventually tested (congruent), made a microsaccade in the direction of an item that was not tested (incongruent), or did not make a microsaccade. Due to the relatively low number of trials with microsaccades in them, we pooled the data from all sessions in order to estimate precision and performance for each microsaccade condition. Since this procedure...
yields single values of precision and performance, we performed a permutation test to test which values were significantly different from one another. Briefly, for each set size, we shuffled the labels for the microsaccade condition and recalculated the performance and precision estimates. We then calculated the difference between the values for each condition using the shuffled data and repeated this procedure 10,000 times to generate seven distinct distributions of differences (one for each test of interest). We then compared the observed difference to the shuffled distributions and calculated a one-sided p-value as the proportion of shuffled differences whose value was greater or equal than the value of the observed difference.

All analyses were done in MATLAB (Mathworks, Release 2011b).

## Results

We trained two subjects on the visual change detection task shown in Figure 1. For subject G, we collected the data across 76 sessions with ~700 trials per session. For subject I, we collected the data across 72 sessions with ~600 trials per session. To determine if subjects had learned the task and could discriminate all colors, we calculated the subjects’ performance for each color using trials with set size one and for different values of \( \Delta E \). Figure 2 shows that both subjects could remember all colors at a high level of performance. The mean performance across sessions was 89 ± 3% and 80 ± 5% for subjects G and I, respectively (black error bars in Figure 2). For all colors, performance was generally better for larger values of \( \Delta E \).

### Performance on the multi-item change detection task

We calculated subjects’ hit and false alarms rates as well as mean performance averaged across sessions as a function of the number of items in memory (set size) using trials for which the difference between the sample and test colors was largest (\( \Delta E = 100 \)) in order to determine subjects’ maximal performance in this task. Figures 3a and 3b show that for both subjects as the set size increased, performance decreased significantly (1-way ANOVA; subject G: \( F_{3,223} = 280, p < 1 \times 10^{-15} \), subject I: \( F_{3,187} = 39.0, p < 1 \times 10^{-15} \)). For set size one, performance was close to 90% for subject G and 79% for subject I. For both subjects, each additional item significantly lowered performance (post-hoc Tukey–Kramer test, \( p < 0.05 \)), with the exception of the drop from three items to four items for subject G.

For both subjects, the hit rate significantly decreased as a function of set size (1-way ANOVA; subject G: \( F_{3,223} = 82.7, p < 1 \times 10^{-15} \), subject I: \( F_{3,187} = 18.9, p < 1 \times 10^{-9} \)) and the false alarm rate significantly increased with increasing set size (1-way ANOVA, subject G: \( F_{3,223} = 37.4, p < 1 \times 10^{-15} \), subject I: \( F_{3,187} = 3.27, p < 0.025 \)). For subject G, each additional object significantly decreased the hit rate and increased the false alarm rate with the exception of increasing from three to four items (post-hoc Tukey–Kramer tests, \( p < 0.05 \)). For subject I, the hit rate for set size one was significantly higher than all other set sizes and the hit rate for set size two was significantly higher than set size four (post-hoc Tukey–Kramer tests, \( p > 0.05 \)); other differences were not significant. Additionally, for subject I, only the false alarm rate for set size four was significantly different from set sizes one and two (post-hoc Tukey–Kramer tests, \( p < 0.05 \)).

We also examined how quickly subjects reached their decision as to whether the color had changed (Figures 3c and 3d). After the memory period, the test color was presented on the screen and subjects were able to indicate their response immediately. For all trials, we calculated subjects’ reaction time at each set size and averaged across trials with \( \Delta E = 100 \). As the number of items in memory increased, the subjects took longer to respond (Kruskal–Wallis test, both subjects \( p < 1 \times 10^{-16} \)). For both subjects, each additional item, up to set size three, significantly slowed reaction times (post-hoc Tukey–Kramer test, \( p < 0.05 \)). The addition of a fourth item did not affect reaction times.

![Figure 2](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932803/) Subjects’ mean performance averaged across each session for each color and for different \( \Delta E \). The color of the bars in the background represents the color of the sample item. The black data point on the left of each figure panel indicates the mean performance (±standard error) for all colors and all values of \( \Delta E \).
In summary, the animals’ behavioral accuracy and reaction times showed that they could successfully perform a multi-item color change detection task with up to around three items. The addition of a fourth item had little effect on performance and thus may indicate that subjects reach a performance limit beyond three items.

Previous studies that have examined the capacity of VSTM in non-human primates (Buschman et al., 2011; Heyselaar et al., 2011) have used a version of Cowan’s formula (Cowan et al., 2005) to calculate the capacity: $k = N(h - f)$, where $h$ and $f$ are the hit and false alarm rates, respectively, and $N$ is the number of items in the display. We applied this formula to our data using trials with the highest discriminability between sample and test colors ($\Delta E = 100$) and obtained an estimate for subject G of $k = 0.88$ and $k = 0.83$ for subject I. This seems to contradict previous reports that macaques have a capacity of around 2 items. However, this fixed-capacity model of VSTM makes the assumption that information is stored in VSTM in discrete fashion, i.e., that an item is either in memory or not in memory. An alternative account of the nature of VSTM argues that this need not be the case, and instead, it proposes that VSTM is a continuous resource that is split among the different items at the expense of representational fidelity. To distinguish between these two accounts, we looked at how the fidelity of the representations varied as a function of the number of items in memory as we parametrically varied the difference between the sample and test colors.

### Performance at different $\Delta E$

The results from the preceding section were calculated using trials in which the difference between the sample and test colors was highest ($\Delta E = 100$). We would expect performance on trials with small $\Delta E$ to be worse than trials with large $\Delta E$ because a small difference between sample and test colors makes it more difficult for subjects to detect a color change. For both subjects, and at all set sizes, performance was at chance for the lowest value of $\Delta E$ (Figure 4).

Indeed, for the smallest $\Delta E$, as set size increased, performance dropped significantly below chance. This suggests that subjects were unable to discriminate between colors separated by a distance of 40 units or less and were judging both colors to be the same (thus making more incorrect responses than by simply guessing). As the sample and test colors became more discriminable (higher $\Delta E$), performance improved for set size one and eventually plateaued at around 93% for subject G and 84% for subject I. For trials with two and three items, the subjects performed better than chance only when the difference between sample and test colors was large (binomial test;
subject G: $p < 0.05$ for set size two at $\Delta E = 80$, for set sizes two and three at $\Delta E = 100$; subject I: $p < 0.05$ for set sizes two and three at $\Delta E \geq 80$).

In summary, our subjects performed the task more accurately as the difference between the sample and test colors increased. Furthermore, these effects became less pronounced as the set size increased. Taken together these results suggest an interaction between memory load and fidelity with which memory representations are held in VSTM, such that increasing load results in less precise representations.

Comparison of variable-precision and fixed-precision models

The above analysis suggests that the precision of stored representations varies as a function of set size. An alternative is that the precision of representations is fixed and the drop in performance measures is due to the fact that subjects only have a limited number of memory slots in VSTM. To determine this, we examined whether our data could be better explained by a variable-precision model or a fixed-precision slots model (see Methods section). We computed the probability that the subjects would report “change” as a function of the difference between the sample and test colors (i.e., the probability of a correct response as a function of $\Delta E$) and fitted the models specified by Equations 1–5 to this probability. To obtain as reliable an estimate as possible, we pooled our data across sessions and subjects. We calculated the AIC and the corresponding Akaike weights for each fit (Table 1). For the variable-precision models (Equations 1–3), we found that the evidence overwhelmingly favors Model 3 relative to Models 1 and 2 for all set sizes above 1. Since the Akaike weights overwhelmingly favor Model 3, we base our estimates of precision on Model 3, which we refer to as the “variable-precision” model, and do not consider Models 1 and 2 further.

We also tested the predictions of a fixed-precision slots model specified by Equation 5. Since the slots model is defined in a piecewise manner for $k$ below or above the number of items in the display, a value for $k$ is needed in order to properly fit it to the data. Since all of our

![Figure 4. Mean performance (±standard error) averaged across trials as a function of set size separated by $\Delta E$. For trials with set size one, performance is at chance only for the lowest value of $\Delta E$; for higher values of $\Delta E$, performance is significantly above chance. For other set sizes, performance is significantly above chance only for set size two and only at $\Delta E = 100$.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932803/)

<table>
<thead>
<tr>
<th>Set size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>9738 (0.0)</td>
<td>13,652 (0.0)</td>
<td>13,050 (0.0)</td>
<td>5656 (0.0)</td>
</tr>
<tr>
<td>Model 2</td>
<td>9696 (0.28)</td>
<td>13,649 (0.0)</td>
<td>13,035 (0.0)</td>
<td>5645 (0.0)</td>
</tr>
<tr>
<td>Model 3</td>
<td>9697 (0.16)</td>
<td>13,279 (1.0)</td>
<td>12,752 (1.0)</td>
<td>5514 (1.0)</td>
</tr>
<tr>
<td>Slots ($k = 1$)</td>
<td>9696 (0.28)</td>
<td>13,617 (0.0)</td>
<td>12,969 (0.0)</td>
<td>5547 (0.0)</td>
</tr>
<tr>
<td>Slots ($k$ not fixed)</td>
<td>9696 (0.28)</td>
<td>13,606 (0.0)</td>
<td>12,967 (0.0)</td>
<td>5548 (0.0)</td>
</tr>
</tbody>
</table>

Table 1. AIC values and the corresponding Akaike weights in parenthesis for each set size for the different models. The smallest values of the AIC (and corresponding highest Akaike weights) are denoted in bold. For the slots model, we tested the model with the capacity fixed at $k = 1$ and with $k$ as a free parameter.
behavioral measures show a sharp decline for set sizes larger than 1, we tested a model in which the parameter $k$ was fixed at a value of exactly 1. We set the value of $A_0$ to the value estimated using only set size one trials. Table 1 shows the AIC values and Akaike weights for this model. We found that the variable-precision model (Model 3) outperformed the fixed-precision slots model. In an attempt to improve the slots model fit, we did not restrict $k = 1$, and instead, we recalculated the fit with $k$ as a free parameter determined by MLE. We found the MLE estimate $k = 1.12$. The AIC value (Table 1) indicates that using this value gives a better fit compared to $k = 1$, even with the penalty of an extra parameter. The improved fit of this model, however, was still worse compared to the variable-precision model.

Having established that Model 3 gives the best fit out of the three variable-precision models and that the model with variable $k$ gave the best fit of the two fixed-precision slots models, we fit these two models separately to both subjects (Figure 5) and calculated the goodness of fit of the model. The fit for both models was significant in both subjects ($F$-test, $p < 0.05$) for set sizes one, two, and three. Furthermore, the fit was consistently better for the variable-precision model relative to the fixed-precision model in both subjects.

### Precision of memory representations

We estimated the precision of subjects’ internal color representations, as the reciprocal of the standard deviation ($1/\sigma^2$) of the variable-precision model. For both subjects, the precision became smaller as a function of set size (Figure 6). Subjects showed a significant decrease in precision for trials with more than one item ($\sigma^2$ estimates for set size >1 fall outside the 99% confidence interval for set size one). This indicates that as soon as a second item was added to memory, the representation of both objects was significantly degraded. Adding a second item to the display caused the precision to drop to almost half of the maximum for subject G and to about 70% of the maximum for subject I. Adding third and fourth items caused the precision to drop further in both subjects.
Effect of microsaccades on task performance and precision

During the sample period, subject G tended to make microsaccades in the direction of one or more of the items in the array. We observed microsaccades in 11% of all trials in all sessions. Out of those, 20% were made in set size one trials, 36% in set size two trials, and 41% in set size three trials, but only 3% of microsaccades were detected in set size four trials. Consequently, we had insufficient data to assess the effects of microsaccades on set size four trials and concentrated our analysis on trials with set sizes two and three. The subject’s low number of microsaccades on set size four trials as well as his chance performance on set size four trials suggests that the subject simply chose to guess when presented with four items rather than attempt to perform the task.

We sought to determine the effect that making a microsaccade in the direction of a particular item had on behavioral performance (see Methods section). It has been previously suggested that microsaccades can be used as a proxy for the locus of attention (Brien, Corneil, Fecteau, Bell, & Munoz, 2009; Engbert & Kliegl, 2003; Hafed & Clark, 2002; Hafed, Lovejoy, & Krauzlis, 2011). We hypothesized that if this was the case for our subject, his performance should improve in the multiple-item trials if attention is focused on one of the items. We found that this was indeed the case (Figure 7).

For set size one, we tested whether there was a difference in performance when the subject made a microsaccade versus no microsaccade (since there was only one item in the display, incongruent microsaccades were not possible). We found that there was no significant difference in performance between microsaccade and no microsaccade trials (permutation test, \( p > 0.1 \)). For set sizes two and above, subjects could make congruent, incongruent, or no microsaccades. We found that there was a significant improvement in performance on congruent trials (permutation test, \( p < 0.01 \)) and a significant decrease in performance on incongruent trials (permutation test, \( p < 0.001 \)) for both set sizes two and three.

Having established that microsaccades can affect behavior in the task, we next looked at the effect of the microsaccade on the precision of the represented information. We fit Model 3 for each set size again but separating the trials based on the microsaccade condition. Again, to obtain as reliable an estimate as possible, we pooled data across sessions and focused on set sizes 1–3 since the low number of microsaccades in the four-item displays and the poor model fit precluded an accurate estimation of the precision.

The overall trend of a drop in precision as the number of items increased remained unchanged (Figure 7b). However, for set sizes two and three, a congruent microsaccade resulted in a significantly larger precision of the stored representation compared to no microsaccade (permutation test, \( p < 1 \times 10^{-3} \)). Additionally, for set sizes two and three, an incongruent microsaccade resulted in a significantly smaller precision compared to congruent microsaccades (permutation test, \( p < 1 \times 10^{-3} \)).

Discussion

We report three main findings. First, we found that rhesus macaques can successfully perform a multi-item color change detection task up to around three items. As the number of items to be stored in VSTM increased, there was a decrease in performance and an increase in reaction times.

![Figure 7](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932803/)

Figure 7. (a) Performance averaged across trials as a function of set size separated by the microsaccade condition. (b) Precision estimate as a function of set size separated by microsaccade condition. For set size one, there are no incongruent microsaccades since there is only one item in the display. Error bars denote the 95% confidence interval (*\( p < 0.01 \), **\( p < 0.001 \)).
Second, we found that the precision of memory representations decreases as the number of items loaded in VSTM increases and that a significant decrease in precision occurred as soon as a second item was added to memory. Finally, we found that covertly attending to one object had the effect of increasing the precision with which it was stored in VSTM while decreasing the representational precision with which the other objects in the array were stored.

Animal model for VSTM capacity

There has been great interest recently in establishing animal models of a capacity-limited VSTM in order to probe the underlying neurophysiology behind this phenomenon. Compared to human psychophysics, there is relatively little known about the capacity of VSTM in primates. One of the earliest studies employed an N-back task with image sequences of varying length (Yakovlev, Amit, Romani, & Hochstein, 2008). Monkeys were shown a sequence of distinct images and were rewarded for recognizing the first repeat of an image. As the number of intervening images between the sample and the target increased, overall performance declined and reaction times increased. In this way, it was determined that monkeys could retain information of about 6–7 images at a time. However, these results should be interpreted with caution, as there are several potentially confounding processes that could be at play. For example, VSTM capacity in humans was initially thought to be 7 ± 2 items (Miller, 1956), but this was later revised down to 4 to account for chunking, recency and primacy effects, and proactive and retroactive interference between objects in memory that arise when presenting a sequence of items to be memorized (Cowan, 2001).

An alternative way to determine the capacity limit is to overload the system at the time of stimulus presentation such that there is no time to form chunks or rehearse the information before the presentation time is over. This approach has been widely used in the human psychophysics literature (Luck & Vogel, 1997; Pashler, 1988; Phillips, 1974; Vogel, Woodman, & Luck, 2001; Wilken & Ma, 2004; Zhang & Luck, 2008) and has been adapted for rhesus macaques (Buschman et al., 2011; Elmore et al., 2011; Heyselaar et al., 2011). In these studies, as in ours, animals were shown an array of colored squares and were required to maintain the items in VSTM. The entire array would then reappear at the test phase, and one of the items may or may not have changed color. The subject had to make a saccade to the item whose color had changed or maintain fixation if none of the colors had changed. These studies all reported that monkeys could successfully perform a multi-item change detection task, maintaining up to around 2–3 items in VSTM, agreeing well with our own findings.

Precision of VSTM representations

Although previous studies have estimated the capacity of VSTM in monkeys, there has been less work aimed at understanding the precision with which information is stored in VSTM. One exception is the work of Elmore et al. (2011) who reported on an experiment similar to ours in which they compared the performance of macaque monkeys with human subjects in a color change detection task. Their results show that as the number of items to be remembered increases, the discriminability, $d^\prime$, (which is used as a measure of precision) decreases in a power law fashion. Similarly, our results show that the precision of VSTM representations dramatically decreases as soon as more than one item is loaded into memory. These results point to a model in which VSTM can be thought of as a limited resource that can be flexibly allocated to encode multiple items in memory. According to this view, there is a limited pool of resources that is shared out between remembered items such that adding more items to memory reduces the resources allocated to each item. This reduction is manifested as decreased precision of stored memories, which adversely affects behavioral performance in these tasks. This account of VSTM does not necessarily require that items be stored in an all-or-none fashion leading to a fixed capacity. Rather, this model allows for the flexible distribution of resources among the items in memory so that information about all or most of the items is represented, albeit not very detailed information.

An alternative to the flexible resources model argues that visual working memory has a limited number of slots that represent memories with a fixed precision (Cowan, 2001; Pashler, 1988; Rouder et al., 2008; Vogel et al., 2001; Zhang & Luck, 2008). According to this slots model, subjects are able to maintain multiple representations with a fixed precision up to the limit of the number of slots available. This implies that behavioral measures of performance should stay relatively unchanged as a function of set size until all available slots are filled, leading to a limit in VSTM capacity. When there are no more slots available, behavioral parameters would show a sudden and sharp decrease. However, the slots model also allows the same item to be stored in multiple slots, thereby allowing for a more graded drop-off in behavioral performance as the number of items increases. The precise nature of the behavioral predictions made by the slots and resources model has been the source of considerable controversy (Anderson et al., 2011; Bays et al., 2009; Bays & Husain, 2008; Rouder et al., 2008; Wilken & Ma, 2004; Zhang & Luck, 2008). Our results were better fit by a variable-precision model compared to a fixed-precision slots model.

Although a variable-precision model provided a better fit to our data compared to a fixed-precision slots model, we believe that care must be taken in interpreting this result
and applying it to humans. The fixed-precision slots model can be best tested when subjects possess multiple slots (or at the minimum two slots). This allows a determination to be made of the nature of the precision of representations below and above the number of available slots. If there is only one slot, it is impossible to measure how the precision of sub-capacity representations changes as the number of items is increased, which is a critical prediction of the slots model.

To some extent, understanding the neuronal mechanisms that underlie VSTM may help shed light on the correct way to model the process behaviorally. The results of Buschman et al. (2011) seem to support the shared resources model. They found that at the neuronal level, information is divided among the different items of the array in a graded manner. However, their study did not explicitly measure the precision of the representations either behaviorally or neuronally. Our results suggest that a fruitful avenue for future research will be to understand the neuronal mechanisms that enable the brain to allocate its resources to encode information with varying fidelity.

Role of attention in a capacity-limited VSTM

One possible way of encoding variable-precision representations in VSTM could be to differentially allocate attention to the contents of VSTM (Cowan, 1995; Desimone & Duncan, 1995; Posner, Snyder, & Davidson, 1980). Indeed, Bays and Husain (2008) found that when they cued subjects to covertly attend to one of the items in the sample array, discrimination precision was better when the item was probed compared to non-cued items. Further evidence for this view comes from findings that show that attention has a strong influence on the fidelity of subjects’ recall of spatial frequency information (Huang & Sekuler, 2010). In their study, Huang and Sekuler momentarily presented subjects two gratings and, after a delay period, asked them to report the spatial frequency of one of the stimuli. On some trials, subjects were cued to attend to one of the stimuli and ignore the other. They found that when subjects focused their attention on the cued stimulus, it greatly improved the precision with which the stimulus was later recalled compared to the trials in which both had to be remembered. We did not cue subjects to attend to a single stimulus, and all stimuli were equally likely to be tested. However, subjects could still adopt a strategy of covertly shifting their attention to one of the stimuli and ignore the others.

It has been suggested that microsaccades can provide an index of covert attention (Brien et al., 2009; Engbert & Kliegl, 2003; Hafed & Clark, 2002; Hafed et al., 2011). Although there is some debate concerning the suitability of microsaccades as an index of attention (Horowitz, Fencsik, Fine, Yurgenson, & Wolfe, 2007; Horowitz, Fine, Fencsik, Yurgenson, & Wolfe, 2007; Laubrock, Engbert, Rolfs, & Kliegl, 2007), we found that making a microsaccade significantly affected the subject’s behavioral performance, suggesting that in this case, microsaccades can indeed reveal the locus of covert spatial attention. Similar to the findings in humans, we found that focusing attention on an item significantly affects the representational precision of memories. For those trials where we were able to detect the attentional shift, we found that focusing attention on an item significantly increased the precision of that item. This suggests that increased performance on the congruent microsaccade trials might be due to the subject storing a higher fidelity representation of the object (compared to not making a microsaccade or making an incongruent microsaccade), which in turn makes it more likely that the subject will correctly report whether the tested item changed color. If the subject makes an incongruent microsaccade, he stores an even lower precision representation compared to not making a microsaccade at all. This carries with it a cost: Errors were significantly more likely when the unattended item was tested, presumably due to the degraded representation with which it was stored.

These results point to a significant involvement of attention in controlling the fidelity of the contents of VSTM. They also are relevant to the debate about the nature of VSTM representations. With respect to the slots model, it is difficult to explain how covert attention could increase the precision with which an item is stored in a slot, since the item is supposed to be either stored perfectly or not stored at all. However, the dynamic modulation of precision is a key feature of the resources model (Bays & Husain, 2008) and readily explains our data. Determining the precise neuronal mechanisms of how this selective allocation of attention gives rise to variable-precision representations could be crucial in understanding the fundamental nature of the capacity limit in VSTM. Our results show that the rhesus monkey is a viable model with which to realize this aim.

Acknowledgments

This work was supported by NIDA Grant R01DA19028 and NINDS Grant P01NS04081 to J.D.W.

Commercial relationships: none.
Corresponding author: Antonio H. Lara.
Email: homero@berkeley.edu.
Address: University of California at Berkeley, 132 Barker Hall, Berkeley, CA 94720-3190, USA.

References


