Multi-item Working Memory — A Behavioral Study

Macaque monkeys were trained to recognize the repetition of one of the images already seen in a sequence of random length. On average, performance decreased with sequence length. However, this was due to a complex combination of factors, as follows: performance was found to decrease with the separation in the sequence of the test (repetition image) from the cue (its first appearance in the sequence), for trials with sequences of fixed length. In contrast, performance improved as a function of sequence length, for equal cue–test separations. Reaction times followed a complementary trend: they increased with cue–test separation and decreased with sequence length. The frequency of false positives (FPs) indicates that images are not always removed from working memory between successive trials, and that the monkeys rarely confuse different images. The probability of miss errors depends on number of intervening stimulus presentations, while FPs depend on elapsed time. A simple two-state stochastic model of multi-item working memory is proposed that guides the account for the main effects of performance and false positives, as well as their interaction. In the model, images enter WM when they are presented, or by spontaneous jump-in. Misses are due to spontaneous jump-out of images previously seen.

Keywords: attractor neural networks, multiple memories, recency, serial order, Sternberg, working memory

Introduction

Working memory (WM) concerns the ability of an animal to react selectively based on one or more of a set of cues, following a delay between the removal of the cue (or cues) and the instance of required behavioral reaction. We use the term ‘working memory’ here as it is commonly used in the context of selective delay activity. In other contexts the term used is ‘short term memory’, which we prefer to reserve for situations in which it is the synaptic engram that is short lived, rather than the subjective interval built into the experimental paradigm. This point of view was argued in detail in Amit (1995). In humans, for example, one considers the free recall task, in which subjects are presented a sequence of words and then asked to recall as many words as possible from the list in any order. Items are not recalled with uniform efficacy, but rather the quality of recall depends on the position of an item in the sample list. The first items in the list (primacy effect) and the last items (recency effect) are more effectively recalled relative to the items in the middle of the list. That is, performance versus position is a U-shaped curve (Crowder, 1976; Baddeley, 1995). In the recognition task, in which subjects are asked to judge if a test stimulus had or had not been present in a previously presented sequence, these two effects can be dissociated by manipulating the retention interval — the time between the end of sample sequence and the test stimulus. In humans, monkeys and pigeons it was found that decreasing the retention interval reduces and eventually eliminates the primacy effect, while increasing the retention interval eliminates the recency effect (Wright et al., 1985; Castro and Larsen, 1992). These results indicate that WM has a complex dynamical structure.

Our perspective is to connect animal memory experiments and physiological recordings. This calls for behavioral paradigms that are simplified compared with those used for humans. A common example is the DMS (delayed-match-to-sample) task in which an individual trial consists of presentation of one of a set of images on a screen for a short cue period — the sample; removal of the image; a delay interval; and presentation of a test image. The animal is rewarded for responding selectively according to whether the test stimulus was the same as the sample, or different from it, within a prescribed time period after test image presentation (Miyashita and Chang, 1988). In this situation, behavioral success would suggest that an engram of the sample stimulus has been kept active during the delay interval, up to the presentation of the test stimulus, to lead to a reaction specific to the particular sample stimulus (match or non-match) and earn reward. Physiological recording has found persistent, cue-selective, elevated delay activity, which may be identified with this engram.

The DMS task mentioned above is not the only type of working memory paradigm, nor was it the first. Fuster and Alexander (1971) already used a WM task to expose neural activity in pre-frontal cortex correlated with information transmission across a delay. Monkeys were presented a scene with two boxes (the sample stimulus), and next to one of them was a piece of apple (to mark the relevant box); the scene disappears and, following a delay (up to 1 min or more), the two boxes reappear without the apple (the test stimulus). The monkey’s task was to reach for the box that was near the apple in the sample scene. One naturally assumes that some active neural signal, coding for the direction of the relevant box in the sample (marked by the piece of apple), is communicated across the delay. Indeed, elevated neural spike rates during the delay period were reported in this experiment.

There is a rich line of research in working memory of stimulus direction, (i.e. of position of a stimulus on a circle around the screen center; Funahashi et al., 1989; Goldman-Rakic, 1995). Here the paradigm is as follows: the monkey fixates the screen center; a brief cue appears on the screen, in one of eight directions (sample). Following a delay, a GO signal appears and the monkey is rewarded for making a saccade in the direction of the sample cue. Again, a neural signal representing the sample direction is observed across the delay that may guide the saccade. Closer in spirit to the original Fuster experiments are...
those of Miller’s group (Rao et al., 1997) in which the ‘what’ (image) and the ‘where’ (direction) are tied and untied.

There have been many variations on the early WM experiments, aimed at tracking the structure and the dynamics of the neural correlate of the transmitted WM engram. There is always a possibility that what is transmitted across the delay is picked up from the sample stimulus upon each presentation. This is referred to as short-term memory. In fact, in the Miyashita and Chang (1988) study they observe that while it takes extensive training to generate an observable neural correlate of working memory in the temporal lobe, the monkeys perform quite effectively with images that they have never seen before. The question posed is whether information requiring formation of a synaptic structure (required to produce the physiological correlates) is employed in the WM tasks. This has led to paradigms critical to this issue: fixed order image sequences (Miyashita, 1988; see also Orlov et al., 2002) and the pair-associate paradigm (Sakai and Miyashita, 1991; Erickson and Desimone, 1999; Rainer et al., 1999).

In the first, the individual trial is a standard DMS task, but during training the sample images are presented in a fixed sequence. During testing, however, sample images are chosen at random. It is found that the internal representations (the delay activity distributions) observed in testing, are correlated among neighbors in the training sequence. That is, for each individual neuron, delay activity may be found for more than one of the stimuli, and there is a greater probability of such overlap for stimuli that were neighbors in the original fixed sequence of trained samples. In the pair-associate task the set of images used is divided into fixed pairs — each sample and its pair-associate image. Upon presentation of a sample and following a delay, the monkey is rewarded for recognizing the presentation of the pair associate of the image, not the repetition of the same image, as in the DMS task. This requires that in the delay activity elicited by the sample there be information about the pair-associate of that image. Both experiments produce convincing evidence for long-term synaptic structuring. Moreover, some of the experiments (Sakai and Miyashita, 1991, Erickson and Desimone, 1999) go further to expose the dynamical transition during the delay interval, from sample-related to test-related delay activity (see also Mongillo et al., 2003).

All of the above experiments were restricted to observation, both behaviorally and physiologically, of the propagation of information in WM of a single item (i.e. image or direction). A step beyond has been taken by studies that probed the possible interference of the information related to different items (Miller et al., 1996). In this paradigm the monkey is presented a sequence of images, separated by delay intervals, and trained to recognize the repetition of the first image in the sequence, following an intervening set of distractor images. The monkeys succeed in mastering the task, even if among the distractors there is a repetition — the ABBA type sequences. Sequences of the latter type, however, require additional training. Single cell recordings found that selective delay activity related to the first image in the sequence is preserved in prefrontal (PF) cortex across the intervening stimuli, while in IT cortex, delay activity expresses the last image viewed (see also e.g. Yakovlev et al., 1998; Brunel and Wang, 2001). It should be remarked that all that has been observed is survival in PF cortex of the delay activity of the first stimulus, which finding does not exclude the possibility that several items of the sequence (including the first) persist (for a discussion of criteria for differentiating between these two options, see Amit et al., 2003).

We chose to extend this last paradigm to probe the depth of working memory in primates and subsequently its neural activity correlates. Clearly, cognitive behavior must rely on the ability to hold more than one distinct item simultaneously in WM. It would be required for planning complex action as much as for generating linguistic structures. We therefore modified the paradigm of Miller et al. (1996) to make it necessary for the monkey to hold in WM information representing an entire sequence of images. This was done by presenting the monkey with sequences of images of arbitrary length, selected at random from a pool of images, where the last image in each sequence is a repetition of an arbitrary image in the presented sequence. The monkey is rewarded for recognizing the repetition. The absence of predictability of which of the stimuli in the sequence will be repeated, requires preservation of a trace of as many images in the sequence as possible, to enhance performance. The random length of the trials requires repeated judgement of match/non-match for each image presented (following the first).

Here we report the success of training monkeys for the task together with the observation that it is able to maintain in WM up to 6–7 items simultaneously. We then go on to analyze the quantitative statistics of performance as a function of the different parameters of the trial: the length of the sequence in each trial; the distance between the image to be repeated (cue) and the test (match); the number of irrelevant images preceding the repeating pair (or the length of the entire sequence, for a given cue–test separation); and the length of the inter-image delay interval. The statistics concern both correct responses and errors, among which there are misses (when the monkey does not notice a repetition) and false positives (when the monkey responds as if to a repetition when it has not occurred).

Materials and Methods

Subjects and Apparatus
Two male macaque monkeys (Macaca fascicularis), weighing 5 and 9 kg, respectively, served as subjects. They were maintained under NIH and Hebrew University guidelines for use of laboratory animals for experimentation. The monkeys shared the same cage. They were fed twice daily: in the morning with dry primate ration tablets and ~6 h later, after the experiment, with additional fruit and vegetables. Water was available ad libitum from the second feeding until night. Experimental sessions lasted ~1 h — averaging 100–300 trials — and were conducted 5 days per week. The monkey sat in a primate chair and faced the PC monitor placed ~40 cm in front of him. Data collection of bar press/release and juice reward delivery was managed by a PC through either an ESRD board (Istituto di Fisica, Roma) or DAP 3400 board (Microstar Laboratories, USA).

Stimuli
The stimuli were a set of eight two-dimensional multicolored digital images, used for both monkeys, and a set of 16 images (including the first eight), used at a later stage for one monkey. Figure 1 demonstrates the stimuli used for both set-sizes. The size of the stimuli ranged from 3.5 to 5.0° of visual angle. Stimuli were presented in the middle of the PC screen.

Procedure
In each trial a sequence of images was presented to the monkey (Fig. 1, left). Each image was shown for 1 s and consecutive images were separated by an inter-stimulus-interval (ISI) chosen at random to be 1, 2 or 3 s (but fixed for each trial). The last stimulus, the ‘test’, always matched any one of the preceding ‘samples’. The trial began when the
monkey pressed a response bar, following presentation of a fixation cross. The monkey was trained to release the bar when an image was repeated. Correct bar release was rewarded with a drop of apple juice, while premature bar release ended the trial. The inter-trial-interval was 4 s for trials with a correct response, or 5 s following a miss or false positive response.

We presented sequences randomly, e.g. A-B-C-G-H-A, A-B-J-B or B-J-K-M-M. The number of sample images (not including the last, test, image which is a repetition of one of the preceding images) we called \( n \), the position of the cue within the trial we called \( q \); and the separation in the sequence between the ‘cue’ image and its ‘match’ we called \( d \). Thus, \( n = q + d - 1 \), (Fig. 1, left). The number of images \( n \) in the sample sequence (as well as \( q \) and \( d \)) varied randomly from 1 to 7 for the one monkey and 1 to 6 for the other. Thus, the monkeys needed to remember many images within the trial, in order to identify the ‘match’ as often as possible.

Previous to these tests, the monkeys were pre-trained with shorter sequences, starting with one or two samples and gradually increasing as the monkey reached an overall performance level of \( \approx 70\% \) correct (\( \approx 2 \) weeks per additional length). Data shown are from a period of \( \approx 2 \) months starting from the first trials with the maximal length achieved.

**Statistical Analysis**

Data are usually in the form of a proportion (or percent) correct. For such non-parametric data, statistical analysis depends on a confidence interval (CI), which is readily computed when the number of trials is available (see e.g. Meyer, 1965). Error bars reflect the CI. Similarly, comparisons between different measures are based on a \( Z \)-test for comparing two proportions (again, when the number of trials is known for each) giving an interval estimate by using a CI formula assuming Bernoulli distributions (see Newcombe, 1998).

**Results**

**General**

We view the behavioral results from the perspective of a multiple working memory (WM), which needs to maintain, concurrently, representations of as large a number of images as possible, to produce best possible performance. These representations must be maintained after the eliciting stimuli have disappeared. We assume that if, upon the repeat presentation of one of the images previously presented in the sequence, the image is still present in working memory, the response is positive. Otherwise no match is detected, producing a miss error (Figs 2–4). We also discuss the phenomenon of false positives (FP, see below, Figs 5–8), i.e. cases where the monkey releases the lever even though the stimulus presented is not a repeat of any of the stimuli already presented in the current trial. Such FP responses are interpreted as cases in which one of the images in the sequence is presented for the first time, and that image is already present in WM without having previously been presented in the sequence.

Table 1 presents an overview of all stimulus conditions and monkey behavior. The table presents data in terms of the number of stimuli that were repetitions of a previously presented image (so that recognition results in a hit, and the lack of response is a miss error) and stimuli that were not repetitions, so that lack of response is a correct rejection and presence of response is a false positive. The numbers are given separately for monkey F (with a set of eight images and up to seven stimuli per trial including the repetition image) and monkey M (with sets of eight and 16 stimuli and up to eight stimuli per trial). A neural-based model accounting for the hit–miss part of the data (of one of the monkeys) has been presented in Amit et al. (2003).

We accompany presentation of the data with the help of a much simplified and transparent model, some details of which are given in the Appendix. Basically, we assume that upon...
presentation of each image in a sequence, that image enters WM. But the presence of image representations in WM is not permanent: they can hop out of WM stochastically, and if an image does hop out, upon repeated presentation it will be missing, leading to a miss error. Such jumps are assumed to be caused by natural noise in neural dynamics together with perturbations caused by successive image presentations. These transitions are assumed to be stochastic and to take place with

Figure 3. Performance (a, b) and reaction time (RT; e, f) versus cue position $q$ for each sequence length $n$ and cue-to-match separation $d$ for the behavioral experiment, as well as simulation results (c, d) for comparison to behavioral performance (a, b). Each performance point is the ratio of the number of correct responses (Hits) and the total number of trials with given $q$ and $n$ or $d$ (Hits + Misses). (a, c, e) Points connected have the same sequence length $n$; thus, $d$ decreases as $q$ increases ($d = n - q + 1$). (b, d, f) The same data, with points connected for fixed cue-to-match separation $d$; thus, $n$ increases with $q$. (a, b) Behavioral performance significantly decreases with sequence length $n$ or cue-to-match separation $d$ at fixed cue position $q$. Alternatively, performance significantly increases with cue position, $q$, for fixed sequence length — and thus decreasing $d$ (a) or fixed cue-to-match separation — and thus increasing $n$ (b). The latter is equivalent to the surprising finding that performance increases with the addition of stimuli in the sequence before the cue. For clarity of presentation, we do not show confidence intervals for each performance level in the figures; they are given in Table 2. (c, d) Theoretical (simplified model) hit-miss performance. Simulation performance decreases significantly with sequence length and with cue-match separation, as found for experimental performance data; there is no increase with sequence length, at fixed cue-match separation, because of the assumed fixed transition probability (see Appendix). (e, f) RT (shown only for correct responses) significantly increases with increasing sequence length $n$ or with increasing cue-match separation $d$, for each fixed cue position $q$. RT follows performance, inversely: the better the performance, the shorter the reaction time. For clarity of presentation, we do not show standard errors for each RT level in the figures; they are given in Table 3.
a certain probability per unit time (in the present implementation these transitions are allowed only during image presentations, as is suggested by the analysis of the data, below). Similarly, it is assumed that items of the training set can spontaneously hop into WM, with another transition probability (these can occur during or between image presentations). These hops, if not followed by a subsequent hop out, would lead to FP errors, since upon the first presentation of that image its representation is already found to be in WM.

It is quite intuitive that a stochastic WM, as described above (see also Appendix), will lead to an increasing rate of miss errors, as well as of FPs, as the sequence grows longer, simply because images will have had more opportunity to jump out or in. The data are used to evaluate the relative roles of time and stimulation in transitions; to judge whether transition probabilities remain constant during a trial; whether there are correlations between items in WM; and whether other ingredients affect performance, such as image representations remaining in WM between successive trials, or perceived similarities between different images in the training set.

Correct Performance and Miss Errors
To obtain a preliminary picture of monkey performance of the task, we combine all data, irrespective of the length of the delay intervals or different image set sizes, for both monkeys and over the entire task training history. Figure 2 demonstrates the effect on performance of increasing the length of the sequence prior to repetition, and hence the number of stimuli which the monkeys have to hold in working memory. Performance (hits relative to hits plus misses) deteriorates as sequence length increases. Average performance for the shortest trial \((n = 1)\), a cue followed by a match, is 97%, while average performance for the longest trial \((n = 7)\), 7 samples and a match, drops to 77%. The dependence on sequence length is found to be significant [analysis of variance (ANOVA), \(P < 0.001\)]. The form of the decline in performance with trial length appears to be bi-modal, sharply declining at first and then reaching a relatively fixed plateau. Note that the finding that misses are not avoided even for the longest possible trial length implies that the monkeys were not able to count this high or use the result of this count to know that the last image must be a repetition. As we show below, performance dependence on sequence length is found to be ascribable to several causes: (1) more stimuli are stored in
When trials of a given sequence length \( n \) are plotted against cue position \( q \), one observes that performance decreases rapidly as the position of the cue moves toward the beginning of the sequence, increasing the cue–match separation, \( d \). This is shown in Figure 3a, where the data points connected by lines are those of equal sequence length \( n \). For example, for the longest trials \( (n = 7) \), there is a significant (see below) drop in performance from 99% when the cue is at the last position within the trial sequence \( (q = 7) \) to 49% when the cue is at the very beginning of the trial sequence \( (q = 1) \). Given the simple account of stochastic WM, this is just what one would qualitatively expect, since as the cue moves to the head of the sequence, for a sequence of a given length, the distance between the cue and the match increases and the chance of the item escaping from WM, before the match arrives, increases.

For a given cue–match separation \( d \), the earlier the cue position in the trial (the smaller \( q \)), the shorter the sequence length \( n (n = d + q - 1) \). The curves of Figure 3b express performance dependence on sequence length at constant cue–match separation \( d \). The data are the same as in Figure 3a, but lines connect data points of equal \( d \). Hence, as \( q \) increases along a given line, so does \( n \). Surprisingly: increasing sequence length for fixed \( d \) enhances performance; (lines connecting points of equal \( d \) increase with \( q \)). For example, for \( d = 3 \), if the cue position is \( q = 1 \) (sequence length \( n = 3 \)), performance is 71%, while at \( q = 5 \) (\( n = 7 \)) performance is 86% (Z-test, \( P < 0.01 \)). Yet in the latter case there are more stimuli in each trial; more items held in memory; or more irrelevant images preceding the cue–match pair — all of which would intuitively lead to poorer, not better, performance!

A two-way ANOVA over all 28 data points, for all pairs of values of the two independent variables \( n, q \), reveals that the main effects are significant \( (P < 0.001) \), and the interaction between them is not significant \( (P = 0.2) \). A quantitative indication of these trends is estimated by a least-squares fit to a linear function of \( n \) and \( d \); performance decreases by ~8% for...
each additional image between cue and match for fixed \( n \), and increases by \( \sim 2\% \) for each irrelevant image in the sequence at fixed \( d \).

In terms of our simplified-model account, the finding that for fixed \( d \) performance is enhanced with sequence length indicates that the transition probability to hop out of WM decreases as the trial becomes longer. This is required, because, with a fixed transition probability, there could be no dependence of performance on the number of images that preceded the cue. In this account, the probability that an item is lost before it is repeated depends on the number of stimulations intervening between the two events. The rise of performance with sequence length implies a decrease in the out-jump probability as the trial proceeds. In Amit et al. (2003), this effect was called ‘reward expectation’.

**Figure 8.** Perceived similarity test: conditional probability matrices for (a) monkey F, eight-image set, (b) monkey M, eight-image set and (c) monkey M, 16 image set. The eight-image sets are the same subset of the 16-image set (see Fig. 1b). Perceived similarity between \( X \) and \( Y \) should produce high probabilities for symmetric conditions: \( P(X|Y) \) and \( P(Y|X) \). The high conditional probabilities (in brown/red) for the pair \( 8,13 \) are evident and symmetric for both monkeys when using the eight-image set (plots a and b). For the 16-image set (c), \( P(8|13) \) is significantly in excess, but \( P(13|8) \) is larger than average but not significantly so.
The inter-stimulus interval (ISI) was varied from trial to trial, but performance versus length of delay intervals
effect), or parallel, with a speed that depends on the ‘age’ of
presented images could be either serial, non-exhaustive, moving
mechanism that compares each image with previously pre-
compares each image presented with images still in WM. The
account nor the neural based account (Amit
Preparation of long RT findings. The reason for this is that neither the simplified
Introduction and Appendix) is not helpful for interpreting the
for short ISI (93%) (Z-test, \( P < 0.001 \)), for \( n = 7 \) performance is higher for short ISI (86%) than for long ISI (66%) (Z-test, \( P < 0.001 \)).
These exceptions lead to a significant ISI effect in the ANOVA (\( P = 0.01 \)), and there is also a significant interaction between
length of ISI and \( d \) (\( P < 0.001 \)). This interaction is such that it implies a faster decrease with \( d \) for longer ISI or equivalently a stronger ISI dependence for larger \( d \).

The worse performance for longer ISI in the longest trials could derive from a combination of two effects. First, as we shall see below in the discussion of false positive responses, spontaneous in-jumps of images into WM depend on elapsed time rather than number of images presented. Secondly, as discussed there, if we postulate a saturation of WM, then, as WM is filled, new items may not be preserved. Thus, for longer ISI, WM saturation will occur at lower \( n \), leading to the increase in miss rate that we find.

The worse performance for shorter ISI in the shortest trials, of length \( n = 1 \), i.e. when there are two successive presentations of the same image, one as cue and one as match, could be accounted for by an occasional (low, \(< 5\%\) probability) failure to separate between these two presentations when the delay interval is too short. In this case the monkeys may consider the two presentations as a single one. As suggested below, in the discussion of false positive errors, the monkeys strongly avoid responding upon first presentation, because they are never rewarded at that position in the trial. Hence, they take particular care not to respond too early in the trial.

In summary, our findings imply that the transition probability for an out-jump depends mainly on the number of intervening stimuli, rather than on elapsed time. Since the duration of stimulus presentation itself was not varied, we cannot exclude the possibility that the transition probability for an out-jump depends also on this duration (for an interpretation of this possibility, see e.g. Amit et al., 2003; see also Appendix).

Table 3

<table>
<thead>
<tr>
<th>Q</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>684</td>
<td>3</td>
<td>711</td>
<td>679</td>
<td>5</td>
<td>4</td>
<td>726</td>
<td>711</td>
</tr>
<tr>
<td>SE</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Mean</td>
<td>745</td>
<td>736</td>
<td>706</td>
<td>679</td>
<td>10</td>
<td>11</td>
<td>729</td>
<td>721</td>
</tr>
<tr>
<td>SE</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Mean</td>
<td>726</td>
<td>731</td>
<td>712</td>
<td>700</td>
<td>693</td>
<td>658</td>
<td>736</td>
<td>729</td>
</tr>
<tr>
<td>SE</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>SE</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

| Reaction Times |

The reaction time (RT) is defined as the time elapsed from the onset of the match stimulus to the release of the bar by the monkey. For every fixed sequence length \( n \), we find that RT increases with increasing cue-match separation \( d \) or decreasing cue position \( q \). A detailed account is given in Figure 3c, where trials of each length are separated and plotted according to \( q \), as in Figure 3a. In Figure 3f we present a plot corresponding to Figure 3b, that is, connecting the data points for fixed \( d \). Comparison of Figure 3e, f with Figure 3a, b indicates a corre-
respondence of long RT with low performance — that is, this is the opposite of a speed versus accuracy trade-off. In particular, for \( d = 1 \), performance is flat at maximal level, and RT is flat and at minimum time. The RT behavior is also similar to the dependence of performance in that for a given \( d \), the longer the trial (or the higher \( q \)), the shorter is the RT. Hence, preceding irrelevant stimuli enhance the performance level and lead to faster decisions. For example: for \( d = 3 \), RT = 727 ms for a sequence of length \( n = 3 \), \( q = 1 \), and 680 ms for \( n = 7 \), \( q = 5 \) (t-test, \( P = 0.01 \)).

For each of the three pairs of independent variables (\( n-q \), \( n-d \), \( d-q \)) an ANOVA test was carried out over all data. It reveals that in all three cases the main effects are significant (\( P < 0.001 \)), and that there is no significant interaction between the variables of any of the pairs (\( P = 0.4 \)). This implies that the dependence of RT on any of these variables is approximately linear. A least-squares fit reveals that RT decreases on average by \(-5\) ms for each irrelevant image in the sequence at fixed \( d \) and increases by \(-16\) ms for each additional image between cue and match for fixed \( n \).

Note that the account employed to follow the data (see Introduction and Appendix) is not helpful for interpreting the RT findings. The reason for this is that neither the simplified account nor the neural based account (Amit et al., 2003) proposes a readout mechanism, that is, a mechanism that compares each image presented with images still in WM. The mechanism that compares each image with previously presented images could be either serial, non-exhaustive, moving backwards through the sequence (to explain the recency effect), or parallel, with a speed that depends on the ‘age’ of images in WM (see Discussion).

Performance versus Length of Delay Intervals

The inter-stimulus interval (ISI) was varied from trial to trial, but remained fixed within each trial — set at 1, 2 or 3 s. To investigate the effect of ISI length on performance, we plot separately, in Figure 4, performance for trials with different ISIs.

Data for each sequence length is averaged over \( q \). We find that for all trial lengths, except for the shortest and the longest sequences (\( n = 1,7 \)), performance is not significantly different for different ISI (Z-test, \( P > 0.05 \)). A 3-way ANOVA test over \( n \), \( q \), ISI reveals that the \( n \) and \( q \) effects are highly significant (\( P < 0.001 \)), and there is no interaction between them, confirming what we have seen in the preceding section. If the shortest and longest sequences are excluded, the significance of the dependence on \( n \) and \( q \) remains (\( P < 0.001 \)) and there is no significant ISI effect (\( P = 0.2 \)). Taken together, these findings imply that the transition probability for an out-jump depends mainly on the number of intervening stimuli, rather than on elapsed time. Since the duration of stimulus presentation itself was not varied, we cannot exclude the possibility that the transition probability for an out-jump depends also on this duration (for an interpretation of this possibility, see e.g. Amit et al., 2003; see also Appendix).

There are two exceptions to the lack of ISI dependence. For \( n = 1 \), performance is higher for long ISI (98%) than for short ISI (93%) (Z-test, \( P < 0.001 \)), for \( n = 7 \) performance is higher for short ISI (86%) than for long ISI (66%) (Z-test, \( P < 0.001 \)).

False Positive Errors

In 17% of the trials with monkey M and 29% of the trials with monkey F there was a false positive (FP) response, i.e. the monkey responded upon seeing an image stimulus in the sequence as if it were a repetition when in fact it was not. The correct FP error rate should be a proportion not of the sequence as if it were a repetition when in fact it was not. In 17% of the trials with monkey M and 29% of the trials with monkey F, the reaction time to the FP was longer than the reaction time to the correct stimulus, presumably because the monkeys took longer to verify the match against the image in WM. On every presentation of the same image, one as cue and one as match, could be accounted for by an occasional (low, \(< 5\%\) probability) failure to separate between these two presentations when the delay interval is too short. In this case the monkeys may consider the two presentations as a single one. As suggested below, in the discussion of false positive errors, the monkeys strongly avoid responding upon first presentation, because they are never rewarded at that position in the trial. Hence, they take particular care not to respond too early in the trial.

In summary, our findings imply that the transition probability for an out-jump depends mainly on the number of intervening stimuli, rather than on elapsed time. In addition, the slope of the dependence on \( d \) includes a factor depending on ISI length (time elapsed).

False Positive Errors

In 17% of the trials with monkey M and 29% of the trials with monkey F there was a false positive (FP) response, i.e. the monkey responded upon seeing an image stimulus in the sequence as if it were a repetition when in fact it was not. The correct FP error rate should be a proportion not of the number of trials, but of the number of opportunities in which the monkey could have made an FP error. On every presentation
of an image, which is not a repetition of a cue, the monkey could either correctly pass on (correct rejection, CR), or make a FP. If we normalize the FP rate in this way, counting each non-match presentation within the trials, M had 6.2% and F 11.5% FP errors.

**Dependence of FP Rate on Position and Image Set Size**

The dependence of FP error rate on the position in the sequence of the images that led to the FP is presented in Figure 5a. To obtain an estimate of the probability of an FP at a given position, the number of FPs is normalized by counting all instances that could have produced that FP, i.e. all trials with a FP or a CR at that position. Data are separated for the two monkeys, F and M, and for M are further divided according to image set size (M8 — image set of eight images; M16 — image set of 16 images).

Both monkeys made very few FP errors on the first item in the sequence. This is related either to internalization of the task structure, since reward was never obtained for responding to the first image, and/or to the fact that events in the ITI have rendered WM empty. Below we present evidence in support of the first hypothesis. As the sequence increases in length, the probability of an FP increases significantly (ANOVA, $P < 0.001$). The rise of FP probability with sequence length is consistent both with the option that images hop into WM with the passage of time, or with the number of image presentations in the sequence. In both cases the probability for an in-jump of an image increases with the trial length. This effect is analyzed further by studying the dependence of FP rate on ISI length (see Effect of Delay Interval Length on FP).

For monkey M, we observe that, for each FP position, the FP probability is independent of the size of the image set (8 or 16), except for FP position 7. For this point the difference is found to be significant ($Z$-test, $P < 0.02$). It is also found that without the data for FP position 7, the image set-size effect is not significant ($Z^2$-test, $P = 0.07$). It remains significant if this point is included ($Z^2$-test, $P = 0.01$).

In the stochastic WM account, as the number of images in the set increases, one expects a larger number of in-jumps, because there are more items, each with the same transition probability. But the probability of FP remains constant, because there is also a (proportionately) smaller probability for a given image to be presented. This seems to be the case also for the experimental data.

A possible alternative is that we have an upper bound for the number of items that can be concurrently present in WM, due, for example, to inhibition that rises with increased loading of the network (see e.g. Amit and Mongillo, 2003; Amit et al., 2003). In this case, WM will tend to be overloaded, following presentation of several stimuli in addition to some images that have spontaneously jumped into WM. At this point, other images can no longer jump in and upon presentation of a sample image, some image will have to leave WM. A larger image set provokes more in-jumps prior to overloading, and overloading would be reached earlier. Hence we would expect a saturation, or even a decrease, of FP probability with successive stimulations, that will occur earlier for a larger image set, as happens for monkey M with the 16-image set, Figure 5a. Another implication (suggested by an anonymous referee) would be a faster decrease in performance for larger image sets. In fact, we found decreases of ~3% for sequences of $n = 6$ or 7, but the decreases are not statistically significant for the number of trials available in the current data.

Note that even in absence of this bound we expect saturation of FP probability, because of the eventual equilibrium between in-jump and out-jump transitions. But this saturation is independent of the image set size and would take place, for the given parameters, at longer sequences.

There may be additional sources for FP errors:

- an image in the current sequence ‘matching’ an image of the preceding trial, which has been preserved in working memory and appears as a repetition;
- a perceptual perceived similarity between a pair of images in the set, which are presented in the same sequence, and which the monkey perceives as a repetition.

These two cases are analyzed in the following sections.

**Effect of Images Included in the Preceding Trial on FP**

In considering the potential effect of images of the preceding trial remaining in WM as a cause for FPs in the present one, we return briefly to the comment made in the preceding Dependence of FP Rate on Position and Image Set Size, concerning the very low number of errors upon presentation of the first stimulus. As mentioned there, this result may be related to the monkeys’ learning of the rule implied in the task, that there is never a reward for responding to the first image. But if instead processing were automatic and detection of a repetition was based only on the presence of an item in WM, then the absence of errors on the first stimulus would have been an indication that no items from the preceding trial infiltrate the current trial. As we shall see below, items have a non-zero probability to survive the ITI; hence we reject the second hypothesis.

A first step is to separate the FP data into trials in which the FP image was or was not present in the preceding trial. In 69% of FP trials (902 of 1305) the FP image was present in the sequence of the preceding trial. In contrast, only 43% (12766 of 29962) of images presented in all trials are also present in the preceding trial. Hence, if images that entered WM during the preceding trial do not persist to the present trial, FP errors would be independent of whether the current image had entered WM during the preceding trial, and one would expect that only 43% of FP errors would have the FP-image present in the preceding trial. The difference (from 69%) is found to be highly significant ($Z$-test, $P < 0.001$), which implies an interaction between images in successive trials. We repeated the same analysis for correlations with the last but one trial, and found that FP rate with repetition is not significantly different ($Z$-test, $P = 0.3$) from the rate of an image repetition (42%). Results for older trials are similar. We conclude that images of trials older than the preceding trial do not persist into the present trial.

If WM carries images of the preceding trial to the present trial (Ó Scalaidhe et al., 1997; Yakovlev et al., 1998), we would expect a distribution of FPs depending on the distance between the position of the FP stimulus in the current trial (false ‘match’) and the position of the same image in the preceding trial (false ‘cue’), analogous to the distribution of performance dependence on cue-match separation. Averaging over positions in the preceding trial, and ignoring the in-jumps, we would expect a decrease of FP rate with FP-position in the current trial. As mentioned above, the probability of making a FP increases with the FP position (Fig. 5a), i.e. the opposite behavior. This indicates that in-jumps have a stronger role than does image persistence across the ITI. To analyze the FPs that are 'false
matches’, we decompose the data of Figure 5a, into FPs (and CRs) of images, which were present (‘rep’) or were not present (‘nonrep’) in the sequence of the preceding trial. The same increasing trend is found in both cases, as demonstrated in Figure 6a,b (for monkey F and M, respectively). The fact that the slope of the rep curve is steeper than that of the non-rep curve for Monkey M is considered in the Discussion.

As suggested above, there is indeed a higher FP rate in the ‘rep’ case. Instead, in absence of between-trial correlations, we would have expected a lower rate in the ‘rep’ case, because a given image has a probability of 4.3% (<50%) to be a repetition of an image of the preceding trial. One possibility is that each active item in WM at the end of a trial can survive across the ITI, with some probability and be found in WM during the next trial. Crossing the ITI may reduce the persistence probability of an item, relative to the probability of its persistence in the preceding trial, so that at the beginning of the new trial, images presented in the preceding trial have a low (but nonzero) probability to be present in WM. The difference between ‘rep’ and ‘nonrep’ would be a rough estimate of the probability of crossing the ITI, averaged over images of the preceding trial. Most significant is the fact that there are no differences for FP rate in position 1, despite the fact that there is evidence for WM persistence across the ITI, when higher position FPs are observed. This is strong evidence that the monkeys learned the fact that a choice of an image in the first position is never rewarded, as suggested above, in Dependence of FP Rate on Position and Image Set Size.

We proceeded to check the dependence of FP rate on the backward serial position in the preceding trial, i.e. the distance of the item from the end of the sequence of the preceding trial. Items of higher backward serial position arrive ‘older’ to the ITI and one would expect the FP rate caused by them to decrease. The results are presented in Figure 7a for monkey F and in Figure 7b for monkey M. As expected, we found a decreasing trend for both monkeys (ANOVA, P < 0.001), except for the oldest position for M. (However, this point derivates from sparse data, only 14 FPs out of 207 trials, as evidenced by the large CI.)

In each of the two figures we present also the confidence interval of the FP rate for images that are not present in the preceding trial (‘nonrep’, averaged over current trial FP position). They are delimited by the two horizontal lines. The FP rates should also fall in this interval when the correlations with the preceding trial have died out. We observe (Fig. 7a) that the correlations vanish after five images for monkey F, while they seem not to vanish for monkey M. The FP rate corresponding to the last image of the preceding trial depends also on the behavioral result of the preceding trial: if the trial is a miss, monkey F tends to produce more FPs in the present trial than it produces if the preceding trial is correct (Z-test, P = 0.05), and more if it is correct than it is a FP (Z-test, P = 0.05). Monkey M tends to produce more FPs if the preceding trial is either miss or correct than if it is a FP (Z-test, P < 0.02). The monkeys response at the end of the preceding trial affects only FPs for the last image in that trial (i.e. a repetition of the image to which the response was given); there are no significant differences between trials ending in hits, misses or false positives (Ztest, P > 0.6) in the FP rates (in the following trial) for repetitions of non-last images.

An additional analysis was performed to check consistency with the results given above regarding performance (see Correct Responses versus Delay Interval Lengths). If the jump-out process depends on the number of image presentations rather than on elapsed time, we would expect that FPs driven by images that remain in WM from the preceding trial be independent of preceding trial ISI. This is indeed what we found (Z-test, P > 0.6).

**Effect of Delay Interval Length on FP**

The dependence of FP probability on the length of the delay intervals is plotted in Figure 5b. Data are averaged over monkeys and image set sizes, but are plotted separately according to the different delay interval (ISI) lengths: 1, 2 and 3 s. The normalization procedure is the same as in the previous section. Both effects (position and ISI interval) as well as the interaction between them are highly significant (ANOVA, P < 0.001). For sequences of length 2–5, there is a significant increase of the FP probability with increasing ISI (Z-test, P < 0.05). This indicates that the hop-in transition probability depends on elapsed time — unlike the case for hop-out processes. In fact, analysis of the data plotted as FP versus elapsed time shows that elapsed time accounts for nearly all the variance and there is no significant dependence on number of image presentations (at fixed elapsed time, Fig. 5c).

For the shortest sequences (position 1) there is no noticeable difference in FP probability, which is consistent with either hypothesis, since there is no difference in the number of items presented nor in the elapsed time. The curves tend to cross for the longest sequences (positions 6 and 7). This may be a result of greater noise in the data due to the fact that for longer sequences there was a smaller number of trials, due to trials ending in FPs. It may also be due to images remaining in WM from the preceding trial and erroneously perceived as repetitions (see preceding section), whose effect depends on number of image presentations, rather than elapsed time. Figure 5c shows the dependence on elapsed time of FPs driven by images not presented in the preceding trial, which would therefore be caused only by jumps. The data for the three values of ISI are largely overlapping (within the error bars)—there is no dependence of FP rate on ISI for this case of images that were not presented in the preceding trial — even the points for identical elapsed time represent different positions in terms of number of presented images. The full curve is the analytic solution of the model (see Appendix; parameters are as given there, and are the same as used for Fig. 3c,d). The agreement is quite striking, given that the parameters were selected by visual fit (\(p_{in} = 0.05/s; p_{out} = 0.125/s\)). We conclude that the dominating factor generating FPs in this case is jump-ins that depend only on elapsed time.

**Effect of Image Perceived Similarity on FP**

To examine the effect of potential image perceived similarity on the generation of FPs, we started by analyzing the distribution of FPs over the different images generating the FP. In absence of any perceptive relation between different images, we would expect that every image produce FPs with the same probability. The probability in such a null hypothesis would be the inverse of the number of images in the set. We therefore tested the deviations of the fraction of FPs, produced by each image in the set, from the uniform average. In particular we looked for positive deviations, which could indicate an effect of perceived image similarity.

Both monkeys, trained on the eight-image set, produced a large excess of FP frequency for image nos 8 and 13 (Ztest,
Perceived similarity of two images would imply a high probability that one of the two images provokes the FP given that the other is presented in the sample sequence, and vice versa. We consider all possible pairs and classify all FP trials in a two-dimensional array. Each entry $P(X|Y)$ is the estimate of the conditional probability that $X$ produced a FP, given that $Y$ was in the sequence, defined as the number of FP trials in which the pair $(X,Y)$ was involved, divided by the sum (over $X$) of the numbers of trials in which $Y$ was in the sequence of the FP (sum over the elements of each row in the figures). In Figure 8 we present color-coded matrices of the conditional probabilities for the eight-image set. There are 56 couples (X-Y pairs) of two different images for the eight-image set, and 240 for the 16-image set.

$P < 0.01$). This suggests that these two particular images were often confused. Note that the pair of images is identical for both monkeys. To test this hypothesis we perform an additional analysis, as follows:

Perceived similarity in a pair of images should produce symmetry about the diagonal of values significantly above average — color tending toward red. As seen in Figure 8, the pair (8,13) stands out for both monkeys. We tested the deviation from the mean of each entry $P(X|Y)$. The null hypothesis is that each entry $P(X|Y)$ is distributed around the inverse number of images different from $X$ in the set, i.e. $1/(N-1)$, that corresponds to 0.14 for eight images and 0.07 for the 16-image set. As suggested by the simpler analysis and by visual inspection, we found significant symmetric excess deviations, i.e. similarities, for one single couple (8,13), for both monkeys with the eight-image set, as shown in Figure 8a,b. For the 16-image set the total number of correct rejections in FP trials, which is the total number of events to be classified in the matrix, is 568. This leads to a very low number of FP trials and a statistical analysis at this level was rendered impossible. Nevertheless, despite the low statistics, we did see indications that images 8 and 13 stand out also in this case (see e.g. Fig. 8c). We conclude that images 8 and 13 are perceived as similar by both monkeys. In summary, our findings suggest that perceived similarity is possible, that it can lead to FPs, and that it is quite rare.

**Discussion**

**Summary**

We found that monkeys can be trained successfully to recognize repetition of a random item in a sequence, where the length of the sequence, the items in it and the item to be repeated are selected at random, trial-by-trial. The items (images) of the sequence are presented one after another, separated by a delay interval, which is fixed within each trial but can vary between trials. Behavioral testing with this paradigm, which requires the monkey to access multiple-item working memory, is conceived as a step toward neurophysiological exploration of the neuronal correlates. Some predictions of how such correlates could be identified have been presented (Amit et al., 2003; Bernacchia et al., 2003). Our own unpublished preliminary neurophysiological results seem to confirm the picture, via a corresponding structure of delay activity.

A simple model of stochastic working memory allows a full discussion of performance, including accuracy and FPs. It attributes missed recognition errors to spontaneous hopping of items out of WM, and FPs as due to the spontaneous hopping of items into WM. This model also allows for identification of elements that may account for more complex features of the paradigm, such as the possible dependence of transitions into and out of WM on either elapsed time or increased number of stimulus presentations (required to account for improved performance with sequence length), the upper bound for number of items in WM, and (partial) emptying of WM between trials.

Model elements establish a framework for various data trends, as follows: Images enter WM upon presentation, and remain until an out-jump. Spontaneous out-jumps depend on the number of intervening stimuli, rather than time, and during the ITI, most items leave WM. Items remaining following the ITI, as well as spontaneous in-jumps — which depend on time rather than number of intervening stimuli — lead to false positives. These features account for the performance and FP versus cue position graphs (Figs 3a,b and 5–7). A counter-intuitive finding, that to the best of our knowledge has not been described previously, is the increasing performance (and decreasing RT) with increasing $n$ at fixed $d$. A basis for this finding in the current model would require a decreasing out-jump probability with trial length.

We found an inverse relation between RT length and performance level, whereby the better the performance the shorter the RT (compare Fig. 3a and Fig. 3e). This suggests that there is a gradual process of weakening of image memory, leading to longer RTs following presentation of a repetition of an image. Modeling such a process is beyond our simple model, which assumes that items are either fully in or completely out of WM. The parameters of a modified model are constrained by the parallel between performance and RT data, that we found here. In addition, the RT-performance level inverse correlation indicates that this is not a speed versus accuracy trade-off.

**Features and Implications of Observed Behavior**

A number of features of the observed behavior should be noted, with their implications for underlying memory mechanisms:

1. Above chance performance is observed for all cases tested, even for a separation of six images between the cue (the image that is eventually repeated) and its repetition (match). Given that the monkey is not informed about the position of the image whose repetition is to be recognized, it has to keep as many items as possible in WM.

2. There is a significant recency effect, in that repetition of images shown later in the trial — i.e. closer to the repetition — have a better chance of being recognized than repetitions of images seen earlier, but there is no observable primacy effect. The absence of a primacy effect can be interpreted in light of the findings of Wright et al. (1985), that primacy is not found if either (i) the task does not require the recall of information about the entire sequence or (ii) the retention period before the response is very short. In all previous studies of performance and RT in recognition tasks, the subject was informed that the sample sequence had terminated and it was then that the retention interval began, lasting until the arrival of the test stimulus. In the present experiment the monkey does not know in advance when the test stimulus will arrive and the operation of comparison of each item with all of the preceding ones must take place upon the presentation of every stimulus. Hence, the equivalent retention period can be considered to be roughly the ISI interval, 1–3 s. The
absence of a primacy effect can be traced to the fact that 3 s is still in the range of short retention intervals (Wright et al., 1985).

3. Expectedly, performance decreases with increasing cue-match separation at fixed sequence length (explained by spontaneous jumps of images out of WM). Interestingly, performance for the shortest sequences is significantly poorer for trials with short ISI. This seemingly counter-intuitive fact we have explained by invoking a possible (slight) difficulty in separating rapidly following cue and match stimuli — a form of repetition blindness. Further tests should include even briefer ISI to test this possibility.

4. Surprisingly, performance at fixed cue-match separation increases with increasing sequence length, i.e. with increasing number of irrelevant items before the to-be-repeated cue image. The companion is a reaction time, at fixed cue-match separation, that decreases with sequence length.

5. There is a fair number of false-positives, whose percentage increases with increasing sequence length. The FPs are only marginally affected by spurious perceived similarity of different images, but turn out to be significantly affected by the images in the sequence of the preceding trial. There is a significant correlation between images that produce a FP in the current trial with those also presented in the preceding trial (Fig. 6). Moreover, there is in this case a recency effect: later images in the preceding trial have a higher chance to provoke a FP in the present trial (Fig. 7).

An intriguing finding was that for monkey M the slope of the FP rate with position in the trial was greater for items present in the preceding trial than for other images. This larger slope could be understood on the basis of a dwindling available image set as n increases, leading to an increasing chance of repeating the image(s) that remained in WM from the preceding trial. One could also invoke a novel mechanism that associates memory of images that cross the ITI with the preceding trial rather than the current one — thereby avoiding false positive responses despite presence in memory. Then, the larger slope would derive from a gradual decline in this association. This suggestion needs further experimental testing.

6. The fractions of correct and miss responses seem to be affected mostly by the intervening stimulations, and at most slightly by the time elapsed, while FPs are sensitive to the time elapsed from the beginning of the trial. We therefore conclude — and have included in the model — that jump-in probability is non-zero at every instant of the trial while jump-out probability is non-zero only during the presentation of a stimulus.

7. The main variable affecting reaction time is the cue-match separation: RT increases with increasing cue-match separation. This finding (together with the recency effect) may suggest a backward, non-exhaustive, serial scan of items in WM. Our account, in which all items are simultaneously present in WM, precludes a serial recognition mechanism, because the repeating stimulus finds a common stock of image representations in WM. If single unit recordings confirm that all images presented in the sequence (apart from those that have accidentally jumped out) reside in working memory, one would expect the mechanism leading to reaction to be parallel. The decrease of performance with cue-match separation will have to be related to an increasing difficulty in detection of the repetition with increasing ‘age’ (either in time or in intervening stimuli) of the cue image in WM, as mentioned under point 3, above.

8. The variation of the level of visual response along the trial has been put forth as a tentative ingredient for a readout signal of repetition (Brown et al., 1987; Li et al., 1993; Miller et al., 1996; Brown and Xiang, 1998). Another possibility is that the activity in IT, maintaining the trace of the last presented image, is compared with the multi-item WM activity in PF cortex to detect a repetition. None of these, so far, has been given a clear, task-dependent interpretation. A systematic theoretical account of the variation of RT would require an extension of the modeling approach, whether our simple model (Appendix) or the neural-based model of Amit et al. (2003) to include a description of the readout mechanism.

9. The RT effects found recall findings by Sternberg (1966). In his paradigm, human subjects were asked to recognize whether a test item was a member of a previously presented sequence. RT behavior was attributed to a serial and exhaustive scan, because RT for both match and non-match trials linearly increased with number of items in the sample sequence n and the RTs for both cases were essentially the same. In our paradigm, we found a strong linear dependence on d, i.e. a strong recency effect on the serial position of the cue, which would seem to exclude an exhaustive search (implying a dependence on n). Clifton and Birenbaum (1970) and Sands and Wright (1982), using the Sternberg (1966) paradigm found a serial position effect for RT, suggesting that search was not exhaustive, both in humans and in monkey. When the retention interval preceding the test stimulus is short enough, strong recency (but not primacy) effects are seen (see also Wright et al., 1985). This would hold for our experiment as well, since the retention interval is short. Also here, retention interval seems to play an important role in the recognition computation. Recall that in the present experiment there is no warning signal, and the subject does not know the time to repetition in advance.

Notes
This study was supported by a Center of Excellence Grant ‘Changing Your Mind’ from the Israel Science Foundation and a Center of Excellence Grant ‘Statistical Mechanics and Complexity’ (SMC) from the INFM, Roma-1, as well as a grant from the National Institute for Psychobiology in Israel to VY. A.B. is grateful for a PhD grant of the SMC Center. We thank Dr Ehud Zohary for assistance throughout this study, Ms Svetlana Lein for software support for the behavioral experiments, and Dr Boris Gutkin for bringing the Haarmann and Usher model to our attention.

Address correspondence to Prof. Shaul Hochstein, Neurobiology Department, Life Science Institute, Hebrew University, Givat Ram, Jerusalem, 91904, Israel. Email: shaul@vms.huji.ac.il.
Appendix — A Null Hypothesis Simplified Model

To follow the trend of the behavioral results we use a very simple model of working memory. It is much simpler than the neural-based model of Amit et al. (2003), though it is inspired by it (see also Haarmann and Usher, 2001). While it produces rather rich behavior, it is analytically transparent and very quickly simulated. Here we describe the most elementary version of the model. More elaborate instances, activating options that will be mentioned but not developed, will be described elsewhere.

There is a ‘working memory’ (WM) in which several items can be present simultaneously and independently (in contrast to Haarmann and Usher, 2001). The system is exposed to trials as described in the Methods, except for the absence of an ITI: stimulus presentation interval, $T_{STM} = 1$ s; delay interval between stimuli, $T_{DEL} = 3$ s. At the start of a trial, WM is empty. A subset of $n$ images is selected, both the number $n$ and the identity of the images in the subset are chosen at random and independently. The image to be repeated is also selected at random. As time proceeds, items can jump into WM with probability $p_{in}$ per unit time, and out of WM with $p_{out}$ per unit time. At the beginning of each presentation interval the following takes place: (i) if the image presented is already in WM and it was not previously presented (i.e., this is not yet the repetition image, $n + 1$), the trial is a false positive and is stopped; (ii) if it is not already in WM, it is inserted into WM and the trial proceeds. (iii) If the image presented was previously presented (i.e., this is image $n + 1$), a repetition of the image presented at $q$, and the image is still in WM (i.e., it did not jump out of WM), the trial is a hit; (iv) if it is not in WM, it is a miss.

As the trial period flows, in every small time interval $dt$ there is an in-jump probability $p_{in}dt$ for each item that is not in WM and — upon presentation of a new image — an out-jump probability $p_{out}dt$ for each image that is in WM: $p_{out} > 0$ only during the presentation of a new stimulus, as suggested by the performance analysis. In this way we ensure the main feature of a stochastic subset of images in WM, the probability of losing an image increasing with the number of successive image presentations and the probability of a spontaneous entry increasing with the elapsed time. We did not vary $T_{DEL}$ given that, as suggested below, performance approximately depends only on successive image presentations and FP depend only on elapsed time. The miss-error part of this model is similar to the model of Bugmann and Bapi (2000).

To exemplify the workings of the model we ran 300 trials for every pair $n,q$, for sequences with $n = 1, \ldots, 7$, $p_{in} = 0.005/s$, $p_{out} = 0.125/s$, which implies that (on average) an image jumps into WM every 200 s and jumps out every 8 s of new image presentation (each of which is presented for only 1 sec). Figures 3c,d are the equivalent of Figures 3a,b, respectively, for the model performance level. Given that the transition probabilities are constant along the trial, we expect a flat performance level versus $q$ at fixed $d$, instead of the increasing trend of Figure 3b. The perfect performance of $d = 1$ is due to the fact that out-jumps are not allowed during ISI. The fact that performance data for $d = 1$ is slightly less than 100%, indicates the existence of a small out-jump probability during the ISI, or another source of behavioral noise. The increase of performance with $q$ at fixed $n$ is reproduced, with very similar percentages. The full curves are analytical solutions for the corresponding quantities. For the false positives, the analytic solution of the model and the data are plotted together in Figure 5c (solid line). The increase of FP probability with elapsed time is reproduced. These figures should serve only as an indication of the relevance of the simplified model. No systematic fit was performed.

The probability that an item, which entered at cue position $q$, exits WM before the end of the sequence $n$, and hence the probability of a missed response, is

$$P_{miss}(d) = \frac{p_{in}}{p_{in} + p_{out}} e^{-\frac{n}{T_{STM}} \left[1 - e^{-\frac{(n + 1) - (n + q) + 2T_{STM}}{T_{STM}}} \right]} \left(1 - \frac{(d - 1)T_{DEL}}{T_{STM}} \right)$$

which depends only on $a = n - q + 1$. The probability of a false positive at position $n$ is

$$P_{fp}(n) = 1 - e^{-\frac{(n - 1)T_{DEL} + \left(1 + \frac{a}{T_{STM}}\right)T_{STM}}{T_{STM}}} - P_{miss}(n)$$

where $(n - 1)T_{DEL}$ is the time elapsed after $(n - 1)$ presentations, while $(n - 1)T_{DEL} + T_{STM}$ is the total time elapsed. When $d$ and $n$ become very large, the probabilities saturate.

In the regime $P_{miss}(T_{DEL} + T_{STM}) + P_{out} T_{STM} < 1$, one can use a linear approximation for $P_{miss}$ and $P_{fp}$

$$P_{miss}(d) \approx \left(1 - \frac{T_{DEL}}{T_{STM}}\right)^d$$

In this approximation, performance depends only on the out-jump probability and on the elapsed time during image presentations, while FP probability depends on the in-jump probability and on the total elapsed time.

The model is open to a variety of options:

1. $p_{in}$ and $p_{out}$ may depend on elapsed time during the trial, or on the number of item presentations in the sequence (as e.g. reward-expectation; see Amit et al., 2003). A decrease of $p_{out}$ along the sequence, for example, would naturally lead to a rise of performance at fixed $d$ with increasing $n$, as observed in Correct Performance and Miss Errors.

2. There may be an upper bound on the total number of items that can coexist in WM at any given time, which would produce particular effects for long sequences and large image sets, as hypothesized in FP Dependence on Position and Image Set Size, and Performance versus Length of Delay Intervals.

3. Items in working memory may stochastically pass from trial to trial, leading to effects similar to those observed in Effect of Images Included in the Preceding Trial on FP.

All the above options merit an implementation in the context of the simple model, since they can be easily incorporated and their effect rapidly analyzed. On the other hand, they can provide useful clues to the origin of effects observed in the behavioral study. Yet, changes of this kind would not corrupt the general picture provided by the simple model, in following the observed data.

References


