A dynamic neural resource model bridges sensory and working memory

Ivan Tomić[1,2]*, Paul M Bays[1]

[1]Department of Psychology, University of Cambridge, Cambridge, United Kingdom; [2]Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia

Abstract Probing memory of a complex visual image within a few hundred milliseconds after its disappearance reveals significantly greater fidelity of recall than if the probe is delayed by as little as a second. Classically interpreted, the former taps into a detailed but rapidly decaying visual sensory or 'iconic' memory (IM), while the latter relies on capacity-limited but comparatively stable visual working memory (VWM). While iconic decay and VWM capacity have been extensively studied independently, currently no single framework quantitatively accounts for the dynamics of memory fidelity over these time scales. Here, we extend a stationary neural population model of VWM with a temporal dimension, incorporating rapid sensory-driven accumulation of activity encoding each visual feature in memory, and a slower accumulation of internal error that causes memorized features to randomly drift over time. Instead of facilitating read-out from an independent sensory store, an early cue benefits recall by lifting the effective limit on VWM signal strength imposed when multiple items compete for representation, allowing memory for the cued item to be supplemented with information from the decaying sensory trace. Empirical measurements of human recall dynamics validate these predictions while excluding alternative model architectures. A key conclusion is that differences in capacity classically thought to distinguish IM and VWM are in fact contingent upon a single resource-limited WM store.

*For correspondence:
ivn.tomic@gmail.com

eLife assessment

This study presents **important** insights into the dynamical process whereby sensory information is converted from stimulus-driven activity to a working memory representation from which the information can be recalled later. The evidence supporting the claims is **convincing**, using detailed fits and model-comparison techniques applied to new and existing psychophysical data sets to evaluate a wide variety of potential mechanisms. The overall conclusion, that iconic memory and working memory are not distinct mechanisms but rather two slightly different regimes of the same circuitry, will be of interest to neuroscientists and psychologists studying sensory systems and/or working memory.

Introduction

Keeping relevant information in an easily accessible state is vital for adaptive behavior in dynamic environments. In the primate visual system, this requirement is met by visual working memory (VWM), the capacity to actively maintain visual information from milliseconds to seconds after a stimulus disappears from view (*D'Esposito and Postle, 2015*; *Pasternak and Greenlee, 2005*; *Ma et al., 2014*; *Bays et al., 2024*). While the contents of VWM are frequently updated to reflect changes in the environment and in behavioral priorities, the visual processing hierarchy itself introduces additional layers of dynamism (*Barlow, 1981*; *Van Essen et al., 1992*). The fidelity of representations therefore evolves

from the moment VWM starts accumulating evidence (*Brunton et al., 2013*; *Gold and Shadlen, 2007*) throughout the maintenance period until the information is used for action (*Schneegans and Bays, 2018*; *Panichello et al., 2019*; *van Ede et al., 2019*).

Nonetheless, within most theoretical frameworks, VWM is treated as a stationary process whereby representations are measured and modeled as fixed states of the system. One such model of WM is based on principles of neural population coding (*Bays, 2014*; *Schneegans et al., 2020*). In the Neural Resource model, visual information is encoded in the activity of a population of noisy feature-selective neurons (*Ma et al., 2006*; *Pouget et al., 2000*). The spiking activity of the neural population is constrained by normalization (*Carandini and Heeger, 2011*; *Bays, 2015*), such that the total activity is fixed but flexibly distributed between memoranda, implementing a form of limited memory resource. At retrieval, encoded stimulus values are reconstructed from the noisy spiking activity. This model has provided a quantitative account of patterns of recall error across a range of tasks and stimulus dimensions (*Tomić and Bays, 2024*; *Bays and Taylor, 2018*; *Schneegans and Bays, 2017*; *Bays, 2016a*; *Tomić and Bays, 2018*). However, despite its grounding in principles of neural coding, the basic architecture of the model lacks a temporal dimension to describe the dynamics of memory representations during encoding and maintenance.

Research on prolonged memory maintenance has demonstrated that the precision of stored representations gradually deteriorates over time (e.g. *Pertzov et al., 2017*; *Rademaker et al., 2018*). Computational models attempting to account for these dynamics have often relied on principles of diffusion within an attractor network. In such a network, information is maintained in a sustained pattern of activity, which can be visualized as a 'bump' of activity centered on the stored value. Over time, the bump diffuses along the feature dimension due to random fluctuations in neural activity, leading to stochastic changes in the encoded feature value and a gradual loss of information (*Burak and Fiete, 2012*; *Wimmer et al., 2014*). Critically, the neural code diffuses without decay in signal strength. A growing body of empirical support, both at the behavioral (*Schneegans and Bays, 2018*) and neural level (*Lim et al., 2019*; *Wolff et al., 2020*), identifies diffusion as a key mechanism of memory deterioration.

In contrast to such gradual deterioration over longer retention intervals, studies that probed memory within a few hundred milliseconds of stimulus offset revealed a precipitous decrease in memory fidelity immediately after a stimulus disappears (*Di Lollo and Dixon, 1988*; *Sperling, 1960*; *Bradley and Pearson, 2012*; *Pratte, 2018*). This early superior recall was attributed to a high-capacity but short-lived form of storage termed iconic memory (IM) (*Neisser, 1967*). An implicit assumption has often been that the behavioral advantage of early cues derives from reading out information directly from IM and circumventing capacity limitations imposed by VWM, however, this idea has not been formally modeled or tested. At the neural level, IM is thought to be supported by a brief period of decaying neural activity in early visual areas following the response elicited by the visible stimulus (*Priebe et al., 2002*; *Rolls and Tovee, 1994*; *Teeuwen et al., 2021*; *van Kerkoerle et al., 2017*). In contrast to later memory dynamics arising due to noise accumulation, early changes in memory fidelity were supported by modulation of the neural signal strength. However, little is known about the read-out of this sensory memory buffer.

Finally, memory fidelity changes during encoding while the evidence is extracted from the visible stimulus. Previous studies revealed that longer stimulus exposures have a favorable effect on the subsequent recall, but that this effect is modulated by the number of simultaneously encoded objects (*Bays et al., 2011*; *Shibuya and Bundesen, 1988*; *Vogel et al., 2006*), providing evidence for a processing or encoding limitation of VWM. As stimulus presentation duration increases, more information may be extracted from the sensory signal into VWM, increasing the fidelity of the representation. Critically, with prolonged exposure, VWM fidelity approaches a stable level that depends on the number of encoded items, suggesting that a ceiling is imposed on evidence accumulation by a shared limit on VWM resources. However, a computational framework describing information accumulation from sensory areas into VWM is lacking, and the observed encoding limit may reflect dynamics in sensory areas registering visible objects as well as VWM accumulating this sensory evidence.

Here, we investigated the temporal dynamics in the fidelity of VWM from information encoding until its recall. To map human recall fidelity to the time domain, we conducted psychophysical experiments in which we probed memory representations at different time points relative to stimulus onset and offset while simultaneously manipulating set size. To isolate memory dynamics due to changes
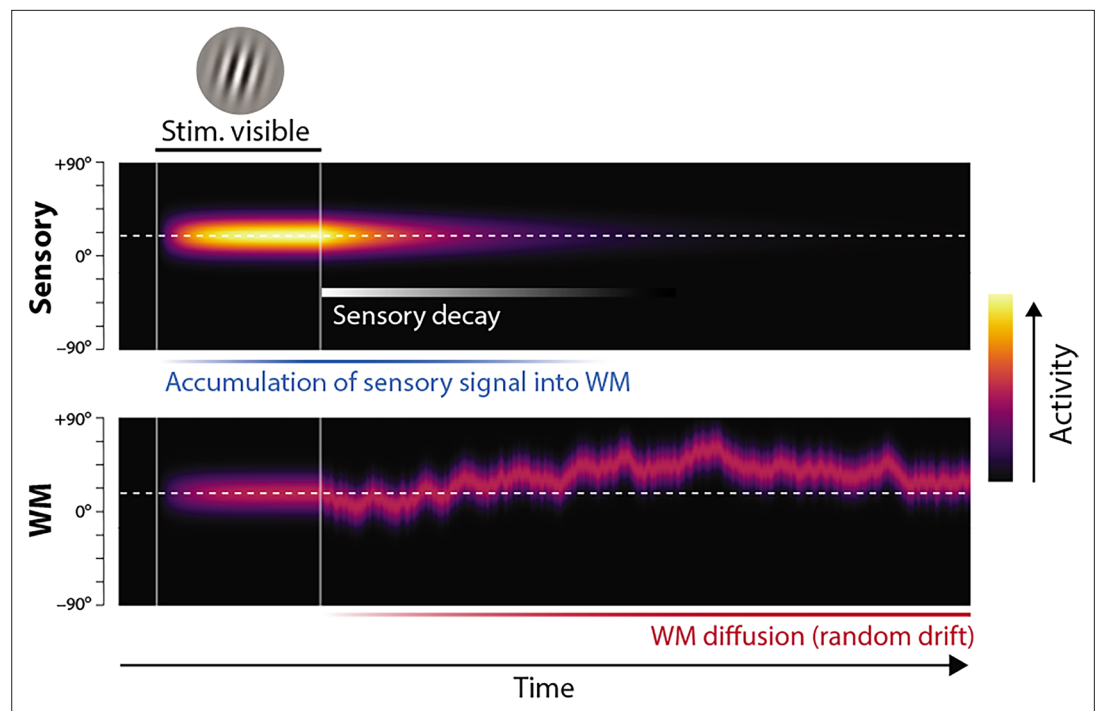
Figure 1. Proposed neural population dynamics for encoding a single orientation into visual working memory (VWM) and maintaining it over a delay. Top: Stimulus onset is followed by a ramping increase in activity (indicated by color) of sensory neurons whose tuning (indicated on *y* axis) matches the stimulus orientation. Following stimulus offset, this sensory signal rapidly decays. The sensory signal, including its decaying post-stimulus component, provides input into VWM. Bottom: At stimulus onset, the VWM population begins to accumulate activity from the sensory population. This accumulation saturates at a maximum amplitude determined by global normalization. As the sensory activity decays, the activity in the VWM population is maintained at a constant amplitude, but accumulation of random errors causes the activity bump to diffuse along the feature dimension (*y* axis) over time, changing the orientation represented by the population. At recall, when the VWM population activity is decoded, accuracy of the recall estimate depends on both the orientation represented (center of the activity bump) and the fidelity with which it can be retrieved (determined by activity amplitude).

in the representational signal, we advanced an analogue reproduction task with a novel response method specifically adapted to minimize the time cost of motor (i.e. response) processes and capture the momentary state of memory representations. This allowed us to precisely measure the time course of fidelity dynamics during representation formation (i.e. encoding) and retention (i.e. maintenance). A major conclusion is that the enhanced precision seen at very brief retention intervals depends on integration of information from the sensory store into VWM following the cue, with the result that retrieval from IM of even the simplest stimulus is subject to the temporal and capacity limitations of WM.

To explain the neural computations underlying the observed time courses, we devised a comprehensive neural model of memory dynamics whose core architecture is rooted in the Neural Resource model of VWM (**Bays, 2014**; **Schneegans et al., 2020**). The Dynamic Neural Resource (DyNR) model assumes that changes in memory fidelity reflect temporal dynamics in the sensory population registering the stimuli and from signal and noise accumulation processes of resource-limited VWM (**Figure 1**). In particular, the model prescribes how time-dependent gain control mechanisms in sensory areas produce a smooth neural response following abrupt changes in stimulus presence. As this sensory signal provides feedforward input to VWM, the dynamics in VWM activity in the temporal vicinity of stimulus presentation (i.e. onset and offset) strongly reflect not only limits in VWM, but also the dynamics of the sensory signal. Finally, once accumulated into VWM, the neural signal is subject to perturbations due to noise accumulation, resulting in degradation of internal representations with time. The DyNR model accurately reproduced the detailed empirical patterns of human recall errors in the psychophysical experiments. Based on these results, we argue that changes in memory fidelity on short time scales reflect dynamics in the gain or signal strength in neural populations representing

the stimulus, while changes on longer time scales are dominated by corruption of the representation by accumulated noise.

Dynamic Neural Resource (DyNR) model

The DyNR model generalizes an established neural population account of VWM, originally proposed by *Bays (2014)*, and inspired by similar models of attention and perceptual decision-making (*Jazayeri and Movshon, 2006*; *Ohshiro et al., 2011*; *Reynolds and Heeger, 2009*). In the original model, memorization and recall of visual stimuli is achieved by encoding and decoding of spiking activity in idealized feature-tuned neurons. The limited capacity of VWM to hold multiple object features simultaneously is reproduced by a global divisive normalization that constrains total spiking activity, implementing a continuous memory resource (*Carandini and Heeger, 2011*; *Bays, 2014*). The DyNR model (illustrated in *Figure 1*) extends this stationary encoding-decoding model with a temporal dimension. First, to capture encoding dynamics, stimulus information enters the VWM population (*Figure 1*, bottom) indirectly, by accumulation of neural signal from a separate sensory population (top), which receives the visual input. The signal strength in the VWM population at any point in time jointly depends on the history of the signal in the sensory population and the number of features competing for representation in VWM. Once the sensory signal is gone, the VWM signal is maintained at its maximum attained amplitude, but the stimulus value encoded by the signal gradually diffuses due to accumulation of random noise. Recall error depends on both the stimulus value represented at the time of retrieval (*what* is encoded) and the signal amplitude at that time, read out in the form of spikes (*how precisely* it can be decoded).

Dynamics of sensory signal strength

To model the temporal dynamics of human memory fidelity, we begin by defining computations of the sensory system registering the incoming signal. A particularly important computation is temporal filtering – a property of neurons to respond more sensitively to specific temporal patterns in stimuli. To model the signal represented in the cortical sensory level, we assume that the sensory response to a stimulus presentation of fixed duration (described as a step function in visual input amplitude, *Figure 2A and B*, left) is controlled by a monophasic temporal filter having a low-pass frequency response (*Hess and Snowden, 1992*). This choice is a natural one since it is consistent with electro-physiological studies demonstrating that a large range of temporal frequencies registered by the retina and LGN (*Derrington and Lennie, 1984*; *Lee et al., 1989*) is attenuated at higher frequencies before the signal enters the primary visual cortex (*Hawken et al., 1996*). Passing the stimulus through such a temporal filter attenuates the neural response to fast transients in the signal, and thereby produces a smooth rise and decay of neural activity in response to a uniform input signal (*Figure 2C*). In particular, we assume that the activity of the sensory population after stimuli onset and offset changes exponentially toward the maximum sensory activity and baseline activity, respectively. The choice of the filter's temporal response characteristics (i.e. its time constant) fully defines dynamics in the sensory population activity and controls the signal projected toward higher areas. The available physiological evidence suggests the temporal properties of the rising and decaying neural response are not symmetric (*Müller et al., 2001*; *Oram and Perrett, 1992*; *Ringach et al., 2003*). In particular, the neural response typically reaches the maximum activity after the onset faster than it reaches the baseline activity after the offset. Consistent with this, we allowed the sensory signal to decay at a different rate than the rising rate. The temporal dynamics in sensory population firing activity in response to a fixed input signal of duration t_{offset} is then given by:

$$\dot{\gamma}_{\text{s}}(t) = \begin{cases} (\tilde{\gamma}_{\text{s}} - \gamma_{\text{s}}(t))/\tau_{\text{rise}} & \text{for} \quad t \leq t_{\text{offset}} \\ -\gamma_{\text{s}}(t)/\tau_{\text{decay}} & \text{for} \quad t > t_{\text{offset}} \end{cases} \tag{1}$$

where $\tilde{\gamma}_{\text{s}}$ is the maximum sensory signal, τ_{rise} and τ_{decay} are rising and decaying time constants of the temporal filter, respectively.

The temporal properties of the sensory response have been shown to depend on the physical characteristics of stimuli, such as contrast and location (*Müller et al., 2001*; *Sit et al., 2009*). Similarly, previous work has demonstrated that the decaying component of the sensory response is strongly influenced by the engagement of the sensory population after stimuli offset (e.g. *Rolls and Tovee,*
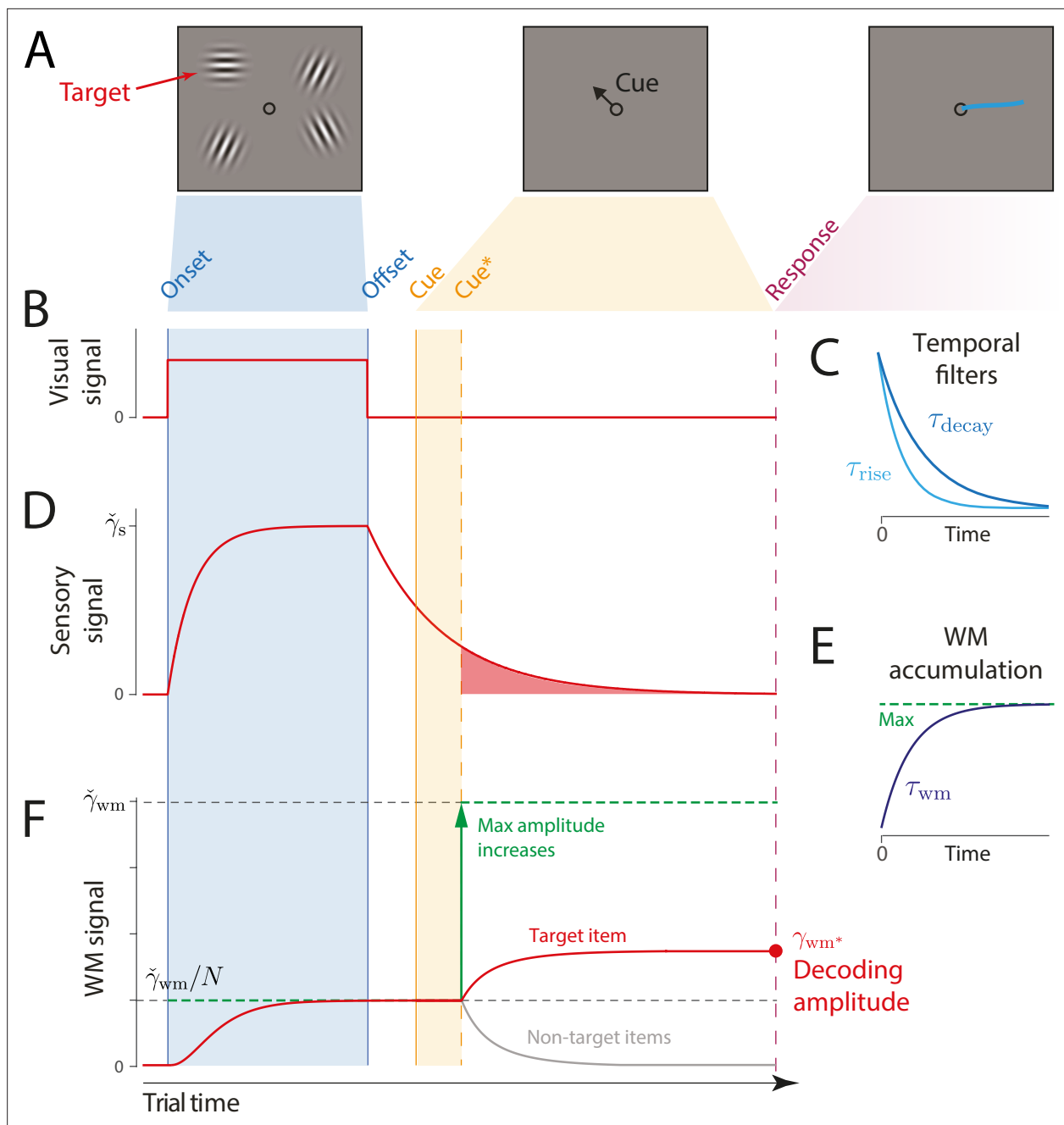
Figure 2. Schematic of signal amplitudes in the dynamic neural resource (DyNR) model during a cued recall trial. (**A**) Observers are presented with a memory array (left), followed after a blank delay (not shown) by an arrow cue (center) indicating the location of one item (the target) whose remembered orientation should immediately be reported (right). (**B**) The amplitude of the visual input associated with each item is modeled as a step function (left). The sensory response (**D**) is modeled as a low-pass filtering of the stimulus input, with different time constants for rise and decay (**C**). (**F**) Amplitude of the working memory signal reflects a saturating accumulation of activity from the sensory population (illustrated in **E**). Beginning with stimulus onset, activity associated with each item is accumulated from the sensory population into the visual working memory (VWM) population, approaching an upper bound (green dashed line) that reflects a total activity limit shared between the N items in memory. Once the cue has been presented (solid orange line) and processed (dashed orange line), uncued items can be dropped from VWM, raising the ceiling on activity available to represent the cued item (green arrow). This allows more information about the cued item to be accumulated from the decaying sensory trace (equivalent to the red shaded area in D). Response variability depends on the asymptotic VWM signal amplitude available for decoding (red circle) combined with the accumulated effects of diffusion (see text).

1994). In particular, a new input signal, e.g., a backward noise mask, curtails ongoing activity related to the previous stimulus, resulting in a faster decay of activity compared to the unmasked post-stimulus period (*Kovács et al., 1995*). Consistent with this, here we assume that the backward mask operates by interrupting ongoing sensory processing of stimuli, limiting the access to the sensory signal (Figure 5) (cf. integration mask) (*Turvey, 1973*).

Dynamics of VWM signal strength

The information registered by the sensory system is subsequently accumulated into a VWM population capable of maintaining activity in the absence of further input (e.g. by self-excitation, see *Aksay et al., 2001*; *Wimmer et al., 2014*; *Compte et al., 2000*; although only the resulting dynamics are modeled here). The total activity of the VWM neural population is normalized, implementing a limited resource shared out between memory items (*Bays, 2014*; *Schneegans et al., 2020*). Consequently, if the stimuli are presented for long enough, the evidence accumulated from the sensory signal into VWM will saturate at a level that reflects the total number of stimuli represented (*Figure 2D*). The dynamics in VWM population activity are given by:

$$\dot{\gamma}_{\text{wm}}(t) = \gamma_{\text{s}}(t)(\tilde{\gamma}_{\text{wm}}/M(t) - \gamma_{\text{wm}}(t))/\tau_{\text{wm}} \tag{2}$$

where $\tilde{\gamma}_{\text{wm}}$ is the maximum VWM signal amplitude, $M(t)$ is the number of items represented in VWM at time t, τ_{wm} is the time constant of accumulation into VWM.

A common assumption of VWM models is that the strength of the representational signal remains stable after encoding from a visible stimulus. This stationary view has been reinforced by typically measuring VWM sufficiently long after the stimulus disappears (~1 s) and at a single time point. In contrast, work on IM demonstrated that recall fidelity in a brief period after stimulus offset typically surpasses and then precipitously decays toward VWM fidelity level (*Coltheart, 1980*). Consistent with that, we consider how the normalized representational signal in VWM formed during encoding can be boosted in the absence of the physical stimulus. In particular, we assume a representation stored in VWM can be strengthened as long as the sensory population provides feedforward input and VWM activity is not saturated at the normalized level. Such a scenario can be achieved by cueing an item for recall in the temporal vicinity of stimulus offset, i.e., before sensory activity decays to zero. By cueing an item for recall, the remaining contents of VWM becomes obsolete and can be removed from memory (*Oberauer, 2018*). In the model,

$$M(t) = \begin{cases} N & \text{for} \quad t \leq t_{\text{cue}*} \\ 1 & \text{for} \quad t > t_{\text{cue}*} \end{cases} \tag{3}$$

where $t_{\text{cue}*}$ is the time when the item is identified for a recall and the read-out of stimulus value begins. This 'demounting' of resource from uncued items makes it available for storing additional information about the cued item, which is extracted from the residual sensory representation, increasing the representation fidelity beyond that granted by equal distribution of neural signal between items. Critically, as sensory information quickly decays, there will be less signal remaining to supplement the VWM representation of a cued item if the cue is delivered later, and at the longest cue intervals the cue will confer no advantage over the fidelity attained when all items compete equally for VWM representation (*Figure 2D*). We note that removal of uncued items cannot occur until the cue has been processed to the point of identifying 1 of the N items in the memory array. We follow *Hick, 1952*, in modeling this cue processing time as logarithmic in the number of alternatives:

$$t_{\text{cue}*} = t_{\text{cue}} + b \log_2(N) \tag{4}$$

where b is a scaling parameter. Previous work demonstrated that estimation of temporal dynamics in attention and memory could be confounded with the time needed to interpret the cue and start acting on it (*Shih and Sperling, 2002*). This is especially significant when trying to accurately capture quickly changing processes, such as decay of the sensory residual. Although the cue processing time likely fluctuates on a trial-by-trial basis due to changes in, e.g., attention, arousal, or motivation, here we focus on the influence of set size arising from a limited information processing capacity.

Diffusion of VWM encoded values

So far we have described only changes in the strength of the neural signal encoding features in memory. However, feature representations maintained over time in neural activity will accumulate noise in the absence of external input. We model this process of noise-driven diffusion as Brownian motion in feature space throughout the retention interval (*Figure 1*), contributing to variability in the decoded feature value (*Burak and Fiete, 2012*; *Schneegans and Bays, 2018*). The resulting variability is described by a wrapped normal distribution with variance σ^2 that increases linearly with time from stimulus offset, so that at time t the encoded feature corresponding to a true stimulus feature θ is:

$$\theta(t) \sim \mathcal{WN}(\theta, \sigma^2(t)) \tag{5}$$

$$\sigma^2(t) = (t - t_{\text{offset}})\dot{\sigma}^2_{\text{diff}} \tag{6}$$

where $\dot{\sigma}^2_{\text{diff}}$ specifies the base diffusion rate. While the fast decay of sensory activity after stimuli offset accounts for early dynamics in VWM fidelity, diffusion becomes prominent over longer delays, accounting for more gradual deterioration of precision with time.

Such a diffusion account has support in the available neural evidence as well as in theoretical work. At the neural level, an electrophysiological study in monkeys performing a spatial WM task demonstrated that shifts of neural tuning curves during a memory delay predicted behavioral response errors (*Wimmer et al., 2014*). A similar finding was observed in humans where drift in the fMRI activity patterns relative to the target predicted errors in an orientation discrimination task (*Lim et al., 2019*). At a theoretical level, continuous attractor models explain diffusion as a consequence of neural variability in networks where excitatory and inhibitory connections constrain population activity to a subspace or manifold corresponding to the encoded feature space (*Burak and Fiete, 2012*; *Bouchacourt and Buschman, 2019*; *Compte et al., 2000*).

Retrieval

To model the process that leads to a response we first consider that in some trials observers may erroneously identify a non-target item as being cued. Previous work indicates these 'swap' errors occur due to uncertainty in memory for the cue features of the stimuli, in this case their locations (*Schneegans and Bays, 2017*; *McMaster et al., 2022*). We assume that changes in variability in the cue features mirror those of the memory features, leading swap frequency to decrease exponentially as a function of presentation duration and increase linearly with retention interval (*Appendix 2—figure 1*):

$$p_{\text{swap}} = (N - 1) \left[\left(\frac{1}{N} - r_{\text{spatial}} t_{\text{cue}*} \right) e^{\frac{-t_{\text{offset}}}{\tau_{\text{spatial}}}} + r_{\text{spatial}} t_{\text{cue}*} \right] \tag{7}$$

where τ_{spatial} is the time constant related to presentation duration, and r_{spatial} is the rate constant related to the retention interval.

If θ is the true feature value of the item identified as the target (i.e. the cued item with probability $1 - p_{\text{swap}}$, a randomly selected non-cued item with probability p_{swap}), then due to diffusion (*Equation 5*) the value encoded in the VWM population at the time of retrieval is given by:

$$\theta^* \sim \mathcal{WN}(\theta, \sigma^2(t_{\text{cue}*})) \tag{8}$$

We model retrieval as estimation of θ^* based on spiking activity in the VWM population that encodes the selected item. For this purpose we assume an idealized set of tuning functions, where the mean response of neuron i encoding orientation θ with population gain γ is described by:

$$f_i(\theta, \gamma) = \frac{\gamma}{n} \exp(\kappa(\cos(\theta - \varphi_i) - 1)) \tag{9}$$

where n is the number of neurons, and κ determines the tuning width. The preferred orientations of the neurons, φ_i, are evenly distributed throughout the circular space to provide uniform coverage. The spike count produced by each neuron is drawn from a Poisson distribution,

$$r_i \sim \text{Poisson}(f_i(\theta^*, \gamma_{\text{wm}*})) \tag{10}$$

and the decoded orientation estimate is obtained by ML estimation based on the spike counts:

$$\hat{\theta} = \arg \max_{\theta} \mathrm{p}(\mathbf{r}|\theta). \tag{11}$$

Additional assumptions

To fit the model to behavioral data, we make several further simplifying assumptions. We assume that the exponential decay of the sensory signal is rapid enough that there is effectively no information remaining by the time the VWM population is decoded to generate a response. This allows us to approximate the VWM activity at the time of decoding by the asymptotic VWM activity were the sensory decay to continue indefinitely:

$$\gamma_{\mathrm{wm}*} \approx \gamma_{\mathrm{wm}}(\infty) \tag{12}$$

Next, we identify diffusion in the encoded value at the time of retrieval with diffusion at the time of target item identification, justifying the use of $t_{\mathrm{cue}*}$ in *Equation 8*. We reason that the rate of diffusion

Table 1. Dynamic neural resource (DyNR) model parameters (1–9) and other variables (10–24) used in model description.

No.	Parameter/variable	Description
1	$\tilde{\gamma}_{\mathrm{wm}}$	Maximum VWM signal amplitude
2	κ	Tuning curve width
3	τ_{rise}	Rise constant of the sensory temporal filter
4	τ_{decay}	Decay constant of the sensory temporal filter
5	τ_{wm}	Time constant of accumulation into VWM
6	$\dot{\sigma}^2_{\mathrm{diff}}$	Base diffusion rate
7	τ_{spatial}	Time constant for spatial encoding
8	r_{spatial}	Rate constant for spatial diffusion
9	b	Scaling parameter for Hick's law
10	t	Time, relative to stimulus onset ($t = 0$)
11	t_{offset}	Time of stimulus offset
12	t_{cue}	Time of cue onset
13	$t_{\mathrm{cue}*}$	Time an item is identified for report
14	N	Number of items in stimulus array
15	$M(t)$	Number of items in memory at time t
16	$\tilde{\gamma}_{\mathrm{s}}$	Maximum sensory signal amplitude
17	$\gamma_{\mathrm{s}}(t)$	Sensory signal amplitude at time t
18	$\gamma_{\mathrm{wm}}(t)$	VWM signal amplitude at time t
19	$\gamma_{\mathrm{wm}*}$	VWM signal amplitude at the time of decoding
20	$\sigma^2(t)$	Accumulated diffusion at time t
21	n	Number of neurons
22	θ	True stimulus feature value
23	θ^*	Encoded stimulus feature value at the time of decoding
24	$\hat{\theta}$	Decoded stimulus feature value

is slow enough relative to the rate of sensory decay, that any additional diffusion in the brief period of post-cue sensory accumulation is negligible.

In Experiment 1 (see below), a task with a fixed 200 ms exposure period, we assume that the initial encoding of all items into VWM is complete by the time of stimulus offset, i.e., that VWM activity at this time can be approximated by its asymptotic level reflecting normalization:

$$\gamma_{wm}(t_{offset}) \approx \tilde{\gamma}_{wm}/N \tag{13}$$

Finally, in the condition of Experiment 1 where memory array and cue are presented simultaneously, we assume that only the cued feature is encoded in VWM, reaching the maximum amplitude, $\tilde{\gamma}_{wm}$, irrespective of set size. Maximum likelihood (ML) fits were obtained via the Nelder-Mead simplex
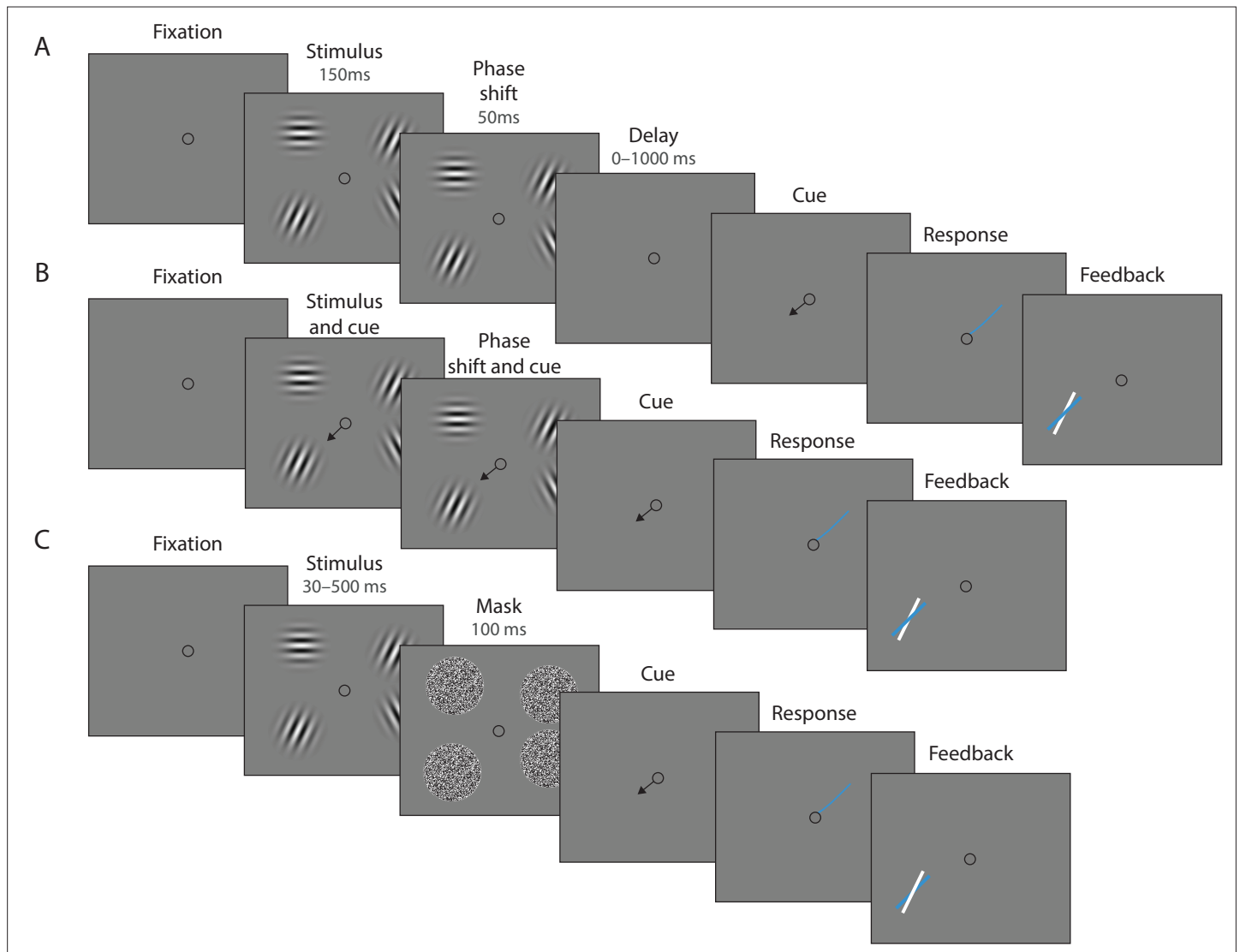

Figure 3. Experimental procedure. (**A**) Experiment 1. On each trial, a memory array was presented consisting of 1, 4, or 10 randomly oriented Gabor stimuli. In 50% of all trials, the stimuli underwent a change of phase and contrast toward the end of the exposure period intended to minimize retinal after-effects. After a variable delay, an arrow cue was shown pointing toward the location of one stimulus from the preceding array. Observers reported the remembered orientation of the cued stimulus by swiping their index finger on the touchpad. The response was followed by feedback showing the true orientation. (**B**) In a proportion of trials, the cue was presented simultaneously with the stimuli. (**C**) Experiment 2. On each trial a memory array consisting of 1, 4, or 10 randomly oriented Gabor was presented for a variable duration, and followed by a white noise flickering mask. The mask was replaced by an arrow cue pointing toward the location of one stimulus from the preceding array. Observers reported remembered its orientation and received feedback as in Experiment 1. Stimuli are not drawn to scale.

method (function *fminsearch* in Matlab). All parameters and variables used to describe the DyNR model are listed in *Table 1*.

Overview of experiments

We tested predictions of the DyNR model against empirical data collected in continuous report tasks. In Experiment 1 (*Figure 3A and B*), observers were presented with an array of oriented stimuli for a fixed duration followed after a variable delay by a visual cue identifying one of the preceding stimuli whose orientation should be reported. This experiment was designed to investigate the contribution of decaying sensory representations following stimulus offset to the dynamics of recall fidelity. Experiment 2 (*Figure 3C*) was aimed at expanding the results of the first experiment to now also assess the accumulation of information during the time the stimuli were visible. In this case, the exposure duration was varied while the delay before the visual cue was held constant. In both experiments we varied the number of stimuli in the array (set size) to assess capacity limitations affecting encoding and maintenance.

To provide additional validation of the DyNR model, we also tested its predictions against data from a previously published continuous report experiment (Experiment 1 in *Bays, 2014*) and one additional dataset collected as part of a separate study (*Tomić et al., 2024*). A detailed description of all experiments is provided in the Methods section.

Results

Experiment 1: Delay duration

In Experiment 1, we evaluated the time course of VWM fidelity over brief memory intervals. Previous work has demonstrated that immediately after a stimulus physically disappears, its representation briefly persists in the sensory system in the form of residual neural activity (*Teeuwen et al., 2021*). Accumulation of this lingering sensory activity into VWM could enable superior recall of information (*Coltheart, 1980*) within the constraints of a finite VWM resource that strongly limits representational fidelity (*Ma et al., 2014*). To describe these dynamics, we examined human recall of orientation stimuli presented in arrays of varying sizes and probed after a variable delay ranging from 0 ms to 1000 ms. Here, we focus on an experimental condition in which retinal afterimages were suppressed by a phase
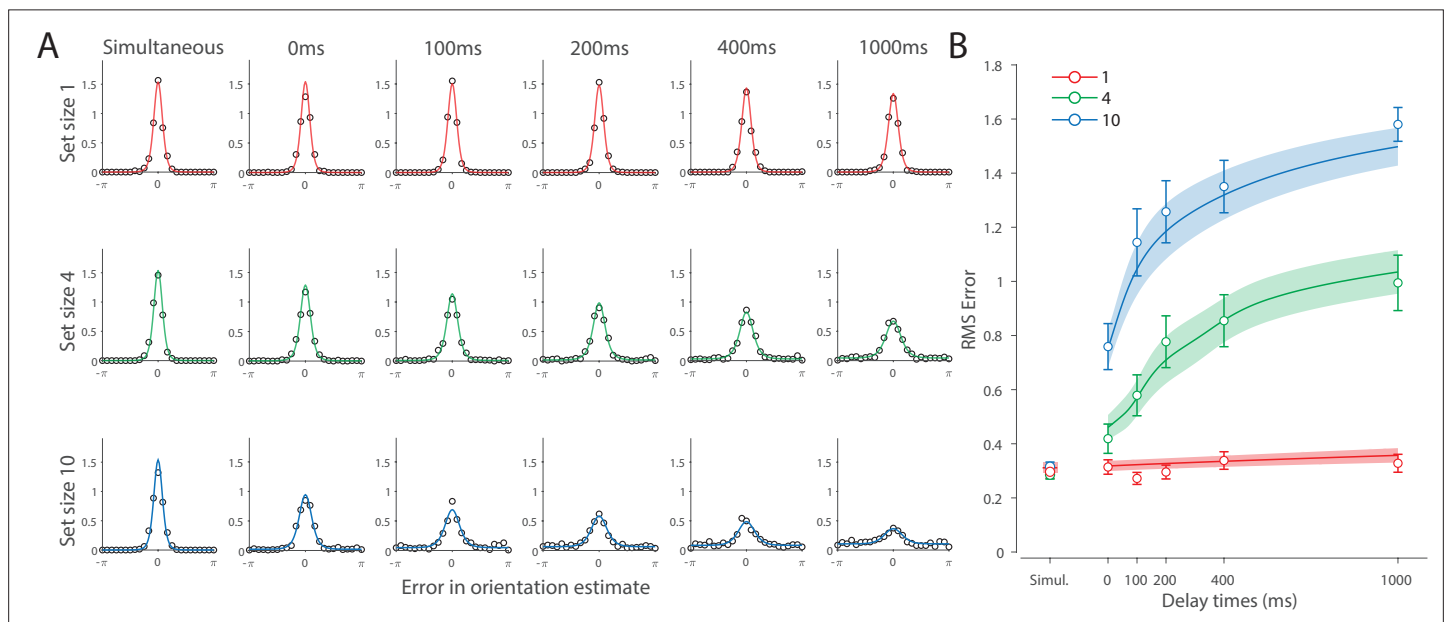

Figure 4. Experiment 1 data and model fits show the consequences of varying set size and delay duration on working memory (WM) reproduction error. (**A**) Empirical recall error distributions (black circles) and the dynamic neural resource (DyNR) model fits (colored curves). Different panels correspond to different set sizes (rows) and delays (columns). (**B**) Corresponding RMS errors from experimental data (circles and error bars) and the DyNR model fits (curves and error patches). Error bars and patches indicate ±1 SEM. N = 10.

shift toward the end of stimuli presentation. Validation of this method and results from the condition without a phase shift are provided in Appendix 1.

Experimental data

Recall error distributions and mean performance in Experiment 1 are plotted in *Figure 4A and B*. Response error (measured as RMSE) increased with both set size and delay duration. A repeated measures ANOVA revealed a significant effect of set size ($F_{(2,18)} = 117.8, \mathrm{p} < 0.001, \eta^2 = 0.44$), delay time ($F_{(5,45)} = 52, \mathrm{p} < 0.001, \eta^2 = 0.23$), and their interaction ($F_{(10,90)} = 26.7, \mathrm{p} < 0.001, \eta^2 = 0.13$) on response error. We further explored this interaction, first finding response error in the 1 item condition (red in *Figure 4*) did not change with delay ($F_{(5,45)} = 1.32, \mathrm{p} = 0.27, \eta^2 = 0.07$). This was supported by Bayesian analysis ($BF_{10} = 0.34$) which found weak to moderate evidence against modulation of 1 item recall by memory delay. In contrast, response error increased with delay for the remaining two set sizes (4 items, green; 10 items, blue; main effect: $F_{(5,45)} = 55, \mathrm{p} < 0.001, \eta^2 = 0.48$). This increase in response error consisted of an initial rapid rise (over the first 200 ms), followed by a more gradual increase as the delay between stimulus and cue increased. Next, we found a modulating effect of delay on recall for the remaining two set sizes (interaction: $F_{(5,45)} = 10.1, \mathrm{p} < 0.001, \eta^2 = 0.05$). The direct comparison revealed that the increase in response error with delay ($\Delta\mathrm{RMSE} = \mathrm{RMSE}_{1000\mathrm{ms}} - \mathrm{RMSE}_{\mathrm{Simult}}$) was greater when observers memorized more items ($t_{(9)} = 9.1, \mathrm{p} < 0.001, d = 2.88$).

One surprising result was the observed set size effect in the 0 ms delay condition ($F_{(2,18)} = 23.7, p < .001, \eta^2 = .53$) consistent with a stepwise increase in recall error with set size (pairwise comparison, $t_{(9)} \geq 2.88, p \leq .036, d \geq 0.91$, Bonferroni correction applied). Importantly, this effect was a consequence of responding based on a memory of the stimulus, since orientation reproduction was comparable across set sizes in the perceptual condition (simultaneous presentation; $F_{(2,18)} = 1.26, p = .3, \eta^2 = .04, BF_{10} = 0.47$). Previous studies have characterized IM as an effectively unlimited store, capable of holding any number of items without a consequent loss of fidelity (*Doost and Turvey, 1971*; *Sperling, 1960*). While our modeling ultimately affirmed this conception of IM, we nonetheless show that recall of information is contingent on the number of objects concurrently in memory from the moment stimuli physically disappear (see below).

Taken together, these results provide evidence that the fidelity of stored representations changes dramatically over the first few moments after stimuli offset. We next aimed to explain the neural computations supporting these dynamics. In summary, behavioral data displayed three key characteristics we aimed to explain, all visible in *Figure 4B*. First, recall fidelity for a single item remained relatively stable across changes in delay, and was the same as perceptual fidelity. Second, recall fidelity for higher set sizes showed substantial, nonlinear temporal dynamics. Lastly, recall fidelity was contingent on the number of stored items from the moment stimuli disappeared.

DyNR model

Curves in *Figure 4A and B* show fits of the model with ML parameters (mean ± SE: population gain γ = 59.8 ± 3.3, tuning width κ = 3.21 ± 0.2, sensory decay time constant τ_{decay} = 0.21 ± 0.052, VWM accumulation time constant τ_{WM} = 0.096 ± 0.045, cue processing constant b = 0.171 s ± 0.055 s, base diffusion σ_{diff}^2 = 0.03 ± 0.017, swap probability p = 0.027 ± 0.009). The model provided a close fit to response error distributions (*Figure 4A*) and summary statistics (*Figure 4B*; see also *Appendix 2— figure 1* for reproduction of swap error frequencies), successfully reproducing the pattern of changes with set size and delay. In particular, the model accounted for the three key observations identified above.

First, the model predicted the near-constant recall fidelity observed for a single item across these short retention intervals. The neural signal associated with the target object at recall depends on the normalized signal in VWM at offset supplemented by the available sensory signal post-cue. The sensory signal is integrated into VWM after the cue to fill any unallocated neural resource that arose by discarding uncued items. In the case of a single item, the entirety of VWM resources are allocated to one object during encoding, so no resource is freed by the cue that would allow the signal to be further strengthened based on the decaying sensory representation.

Importantly, this prediction contradicts the classical view of direct read-out from IM, according to which representational fidelity should be enhanced with very short delays irrespective of VWM limitations (see *Alternative accounts* below for a formal test of such a model). Note that the DyNR model

nonetheless predicts some deterioration in fidelity over time even for a single item, due to noise-driven diffusion of the stored value. However, based on previous reports, we expected this process to be substantially slower and the impact on single-item precision relatively small on this (≤1 s) time scale. The fitted diffusion parameters and resulting shallow slope of fitted RMS error (red curve in *Figure 4B*) confirmed this.

Second, the neural model predicts the specific pattern of dynamics observed in trials with multiple items (set sizes 4, green, and 10, blue curves). Once the cue is presented, resources encoding uncued items are freed and the decaying sensory signal representing the target item is further integrated into VWM, still subject to limited total VWM resources but now without competition from other items. Due to exponential decay of the sensory signal, the increase in fidelity thus accrued changes rapidly with retention interval over the first few hundred milliseconds. At longer delays, the cue identifies the target only after the sensory signal has effectively disappeared, so the VWM signal representing the target item remains at the normalized level reflecting equal distribution between all items in the memory array, and memory dynamics consist only of the more gradual deterioration of fidelity due to accumulated noise in the encoded value.

Finally, the DyNR model predicts the presence of a set size effect on fidelity throughout the entire memory period, including the no delay (0 ms) condition in which the cue onset was coincident with stimulus offset, but not in the simultaneous cue condition. In the model, this behavior emerges as a consequence of two independent processes. First, at the end of stimulus presentation, items within smaller (lower set size) arrays are encoded in VWM with higher signal amplitude, reflecting normalization. This signal strength represents a baseline that can be supplemented by further integration of the sensory signal after an early cue. However, if the sensory decay is sufficiently rapid, then even if the cue is presented immediately the target representation will not attain the maximum amplitude (equivalent to set size of 1) starting from a lower baseline. Second, as described by Hick's law (*Hick, 1952*), it takes longer to identify the target item based on the cue as the number of alternatives increases (see *Alternative models* below for a formal test of this assumption). As a result, for higher set sizes, less sensory signal encoding the target item remains to be integrated into VWM once it has been identified.

Model variants

We next focused on alternative explanations for the temporal dynamics observed in Experiment 1. Specifically, we examined whether the observed dynamics could be accounted for either solely by post-stimulus changes in neural signal amplitude or solely by noise-driven diffusion of stored values. To pre-empt our conclusions, we demonstrate that both components are needed to explain the observed dynamics in memory fidelity. Moreover, to more closely examine the role of diffusion in WM dynamics, we fit our neural model to an additional dataset collected in our lab (*Tomić et al., 2024*; see Appendix 4 for full details). This experiment used longer delays compared to those used in Experiment 1, and therefore precluded any beneficial effect of post-stimulus sensory information, while at the same time allowing the diffusion to operate over a longer period. This experiment allowed us to test whether diffusion is sufficient to account for human recall errors with longer memory delays.

Fixed neural signal

A recent computational study on forgetting in VWM proposed that diffusion is sufficient to explain memory dynamics over delay (*Panichello et al., 2019*). To test for this, we developed two reduced versions of the DyNR model in which the diffusion process was solely responsible for memory fidelity dynamics. In both variants, the sensory signal terminated abruptly with stimuli offset, so the VWM signal encoding the stimuli was independent of the delay duration and equal to the limit imposed by normalization ($\hat{\gamma}_{wm}/N$). In the first variant, the diffusion rate was constant across set sizes, as in the full model. The formal model comparison demonstrated that the full DyNR model performed better than this simplified alternative (ΔAIC = 609.5).

In the second variant, we allowed the diffusion rate to increase proportionally with set size (for a similar proposal, see *Koyluoglu et al., 2017*). This model was again outperformed by the full DyNR model (ΔAIC = 666.4). Critically, both models tested here failed to qualitatively reproduce the observed nonlinear pattern of changes in recall error with time, notably overestimating recall error at the shortest delays by assuming no modulation in the representational signal (*Appendix 3—figure 1*).

Diffusion

We developed two variants of the proposed neural model to test the role of diffusion. In the first variant, we completely omitted the diffusion process from the model to test whether the sensory signal modulation during the retention period is sufficient to explain temporal dynamics in recall fidelity. It could be argued that diffusion accounts for only minor changes in precision over brief delays as used here and, therefore, adds unnecessary complexity to the proposed model without improving the fit substantially. However, the formal model comparison revealed that the full DyNR model provides a better fit to human recall error compared to the matching model without diffusion (ΔAIC = 17.9).

The second variant was identical to the proposed model, except that we replaced the constant diffusion rate with a set-size-scaled diffusion rate by multiplying the right side of *Equation 6* by N. The model comparison showed that the full DyNR model also outperformed this variant (ΔAIC = 29.8). While both model variants qualitatively reproduced the increase in memory error with delay and set size, the pattern of variability was better explained by the model with a constant diffusion rate across set sizes. Although a more substantial diffusion effect could become apparent with longer delays than those used here, previous work demonstrated that noise-driven diffusion causes representations to deteriorate throughout the entire retention period (*Bouchacourt and Buschman, 2019*).

Finally, we examined the role of diffusion with longer memory intervals in a separate experiment using variable set sizes and memory intervals (1 s and 7 s) (for full details, see Additional dataset 1 in Appendix 4). We demonstrated that, once sensory information decayed completely, an accumulation of error during retention interval accounted for continuing memory deterioration. Together, the results presented here corroborate findings on the role of diffusion in temporal dynamics of recall fidelity (*Schneegans and Bays, 2018*).
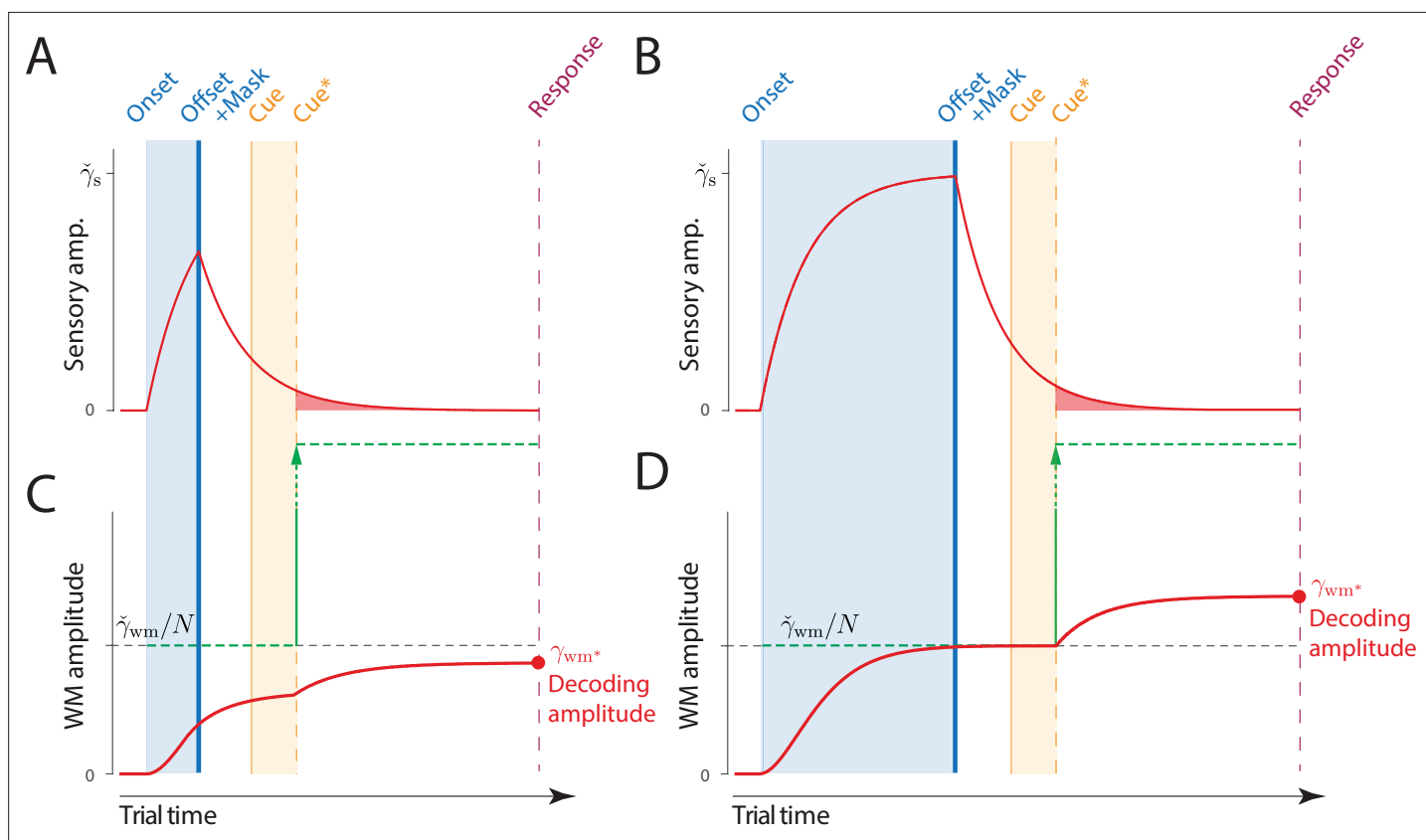

Figure 5. Time course of sensory and working memory (WM) gain with variable exposure duration. (**A, B**) The signal amplitude in the sensory population increases from stimulus onset, exponentially approaching the maximum sensory activity ($\check{\gamma}_s$). For shorter presentation durations (**A**) the attained amplitude at stimulus offset is only a fraction of the maximum (compare B, late offset). Following offset, sensory areas produce a decaying neural response, that is curtailed (faster decay) but not abolished by a backward mask. (**C, D**) Information about the stimulus is accumulated in WM from sensory activity. A shorter presentation (**C**) provides less sensory evidence for the initial accumulation of all items into visual working memory (VWM) (compare D, late offset), and subsequently less decaying sensory activity that can supplement VWM activity for the target item following the cue.
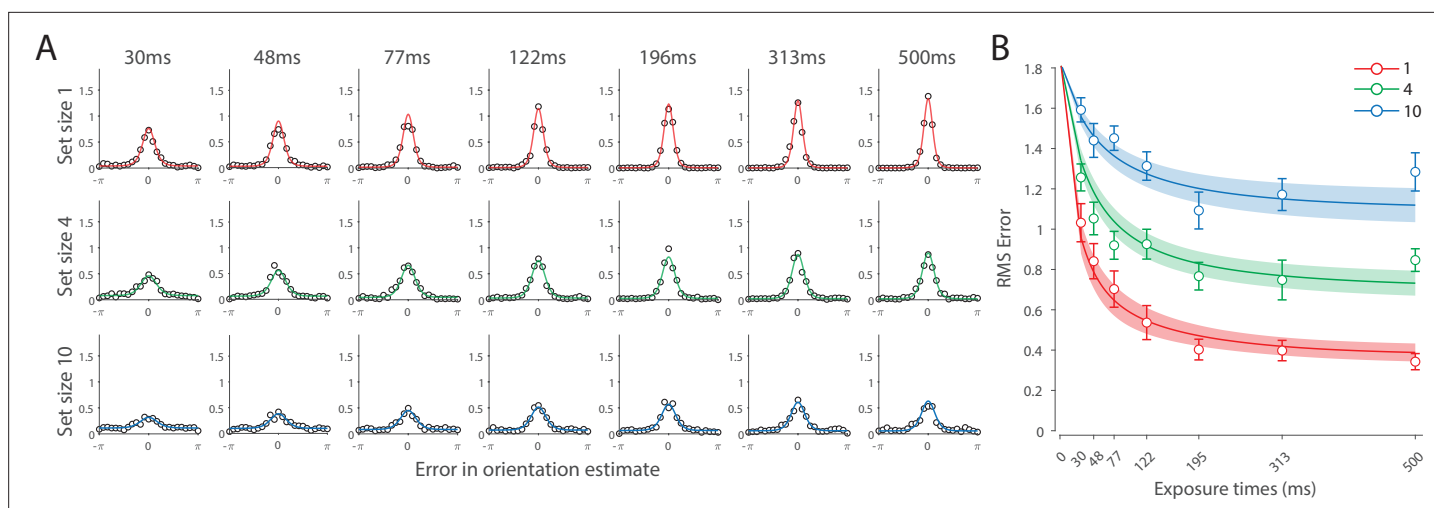
Figure 6. Experiment 2 results and modeling data show the consequences of varying set size and stimulus exposure time on visual working memory (VWM) reproduction error. (**A**) Empirical recall error distributions (black circles) and the dynamic neural resource (DyNR) model fits (colored curves). Different panels correspond to different set sizes (rows) and exposure durations (columns). (**B**) Corresponding RMS errors from experimental data (circles and error bars) and the DyNR model fits (curves and error patches). Error bars and patches indicate ±1 SEM. N = 13.

Experiment 2: Exposure duration

In Experiment 2, we evaluated the encoding phase of VWM, by testing recall of orientation stimuli displayed in arrays of variable size presented for variable durations. In the DyNR model, increasing the sensory evidence by prolonging stimulus presentation has a favorable effect on later recall of stimulus, as more of that evidence can be accumulated into VWM. Importantly, this accumulation is also capped by the VWM resources available to store it (***Figure 5***).

Experimental data

Figure 6 shows the response error for different presentation durations and set sizes. Consistent with previous findings, response error can be seen to decrease with prolonged presentation duration, but increase as the number of items in memory increases. This was confirmed with a significant effect of display duration $(F_{(6,72)} = 29.01, p < 0.001, \eta^2 = 0.21)$, set size $(F_{(2,24)} = 112.51, p < 0.001, \eta^2 = 0.54)$, and their interaction $(F_{(12,144)} = 2.58, p = 0.004, \eta^2 = 0.019)$. We further explored this interaction by first confirming that response error decreased with display duration within each set size $(F_{(6,72)} \geq 10.24, p < 0.001, \eta^2 \geq 0.26)$. A consistent pattern was observed across set sizes, comprising an initial rapid decrease in response error over the briefest presentation times (first 200 ms), followed by saturation at prolonged exposure durations. Next, we calculated the change in recall error between the longest and the shortest display exposure within each set size, revealing that response error decreased more rapidly with display time as the number of items in memory decreased (ANOVA: $F_{(2,24)} = 7.79, p = 0.002, \eta^2 = 0.21$; corrected pairwise comparisons: $t_{1-4} = 3.65, p = 0.016, d = 0.87, t_{4-10} = 0.96, p = 0.72, d = 0.27$).

These results reveal the time course of information accumulation into VWM and forming of stable representations. We again identified several key characteristics of the dynamics of recall fidelity in the data (***Figure 6B***) to test against the DyNR model. Consistent with previous studies, we found recall fidelity changed with both presentation duration and the number of presented stimuli (***Bays et al., 2011***; ***Shibuya and Bundesen, 1988***; ***Vogel et al., 2006***). Specifically, as display duration increased from the shortest exposure, recall error showed an initial rapid decrease followed by a gradual leveling-off. As set size increased, the initial slope became shallower and the plateau occurred at a higher level of error.

DyNR model

Curves in ***Figure 6A and B*** shows fits of the model with ML parameters (mean ± SE: population gain γ = 188.5 ± 109.6, tuning width κ = 10.2 ± 6.08, sensory rise time constant τ_{rise} = 0.33 ± 0.18, sensory

decay time constant τ_{decay} = 0.61 ± 0.19, VWM accumulation time constant τ_{WM} = 0.8 ± 0.34, cue processing constant b = 0.2 s ± 0.09 s, base diffusion σ_{diff}^2 = 0.28 ± 0.08, spatial uncertainty time constant τ_{spatial} = 0.013 ± 0.004, swap probability p = 0.053 ± 0.01). The model provided an excellent quantitative fit to response distributions (*Figure 6A*) and RMSE (*Figure 6B*), successfully reproducing the pattern of changes with set size and presentation duration.

The model predicted that information from a visible stimulus accrues at a high rate immediately after the stimulus onset, consistent with observed changes in human recall error over stimulus durations up to 200 ms (*Figure 6*). This initial high encoding rate emerges naturally in the model due to the joint dynamics of sensory and VWM populations. In the sensory population, a low-pass temporal filter serves as a neural gain control mechanism, attenuating neural response to transient changes in stimuli (*Hess and Snowden, 1992*; *Hawken et al., 1996*). As a consequence, the neural response to stimulus onset increases exponentially (*Figure 5*). The information from sensory areas is accumulated into VWM, such that the accumulation rate is directly proportional to the difference between the current and saturating state (i.e. the rate is faster when accumulated information is far from the saturating state). Therefore, dynamics in the sensory and VWM population jointly account for the initial high rate of information extraction from stimuli, and its dependence on set size.

After the initial steep change, the model predicts that recall fidelity will asymptote. This was again observed in human behavior (*Figure 6*). Extending stimulus presentation beyond 200 ms had negligible impact on recall precision, consistent with previous studies (*Bays et al., 2011*). The model explains this behavior by describing how sensory signal and VWM accumulation independently saturate with time (*Figure 5*). Since the temporal filtering in the sensory population attenuates only high-frequency stimuli (i.e. very short presentations), with sufficient exposure, the sensory signal plateaus, resulting in a stable feedforward input to VWM. Similarly, VWM signal strength is subject to limits determined by normalization. Once the accumulated information reaches the normalized maximum set by the number of objects in memory, further accumulation of sensory evidence is not possible. Following the cue, a portion of the resource is freed, allowing the target representation to be further strengthened. However, because the sensory signal plateaus at longer exposures, the information available for integration after the cue remains constant across the longer exposures, supplementing normalized VWM signal by the same amount. The result is a plateau in fidelity that varies with set size.

Model variants

We investigated whether post-stimulus sensory persistence contributed to the model fits in Experiment 2. We assumed that the signal persisting after stimulus offset would be impaired but not eliminated by the subsequent presentation of a noise mask in this experiment (*Kovács et al., 1995*). An alternative account suggests that the mask immediately terminates any stimulus-related signal. To test for this, we fit a variant of the DyNR model in which the sensory signal was terminated by the onset of the mask, providing a feedforward signal to VWM only for the period of the stimulus presentation. We found that the proposed DyNR model, in which some sensory signal persists after the mask onset, gave a better account of the data than this model variant (ΔAIC = 446.67). Although the alternative model captured the general pattern of changes in memory fidelity with exposure duration, it mispredicted fidelity at shorter exposures, in particular the effect of set size (*Appendix 3—figure 2A*).

A testable prediction of this alternative model is that the memory fidelity at recall should obey the neural normalization principle because there was no additional signal to supplement the presentation after initial encoding. To test for this, we additionally fitted each exposure condition separately using the original neural resource model with only three parameters (i.e. neural gain, tuning width, and swap probability). This model failed to predict actual fidelity levels at recall (*Appendix 3—figure 2B*), corroborating the findings of the model comparison.

Finally, to investigate the role of the post-stimulus sensory persistence on encoding dynamics, we fit the DyNR model to an additional dataset from *Bays et al., 2011* (for full details, see Appendix 5). This experiment aimed to investigate VWM dynamics during encoding, like our Experiment 2. In contrast to our Experiment 2, *Bays et al., 2011*, used a much longer delay interval (1100 ms vs 100 ms), precluding the possibility of further accumulation of sensory evidence following the cue. We expected that the DyNR model could account for memory dynamics in this study without any post-stimulus sensory activity. This was confirmed by accurately reproducing memory dynamics with

a model in which encoding into VWM relied only on sensory evidence during stimulus presentation (detailed results in Appendix 5).

Alternative accounts

Having demonstrated the need for both post-stimulus sensory persistence and diffusion to account for empirical data, we next considered alternatives to our account of VWM accumulation and information read-out.

Direct read-out of sensory information

In the DyNR model, recall fidelity is enhanced following the cue by integrating remaining sensory activity into capacity-limited VWM. As a consequence, response precision is bounded from above by the memory limit irrespective of the available sensory signal. An alternative possibility is that the decaying sensory representation can be directly read out following the cue to inform a response, bypassing WM limitations. To formalize this alternative model, we assumed that independent sensory and VWM representations would be optimally combined via summation of neural activity to yield population gain

$$\gamma_{\text{sum}}^* = \gamma_{\text{wm}}(t_{\text{cue}*}) + \gamma_{\text{s}}(t_{\text{cue}*}) \tag{14}$$

The model is otherwise identical to the proposed DyNR model. A distinctive prediction of this model is that the precision of recall changes exponentially with delay for every set size, including 1 item (*Appendix 3—figure 3*). This prediction is qualitatively inconsistent with the pattern of results observed in Experiment 1, in contrast with the DyNR model which does not predict any beneficial effect of earlier cues with set size 1. This alternative model provided a worse fit to data from Experiment 1 (ΔAIC = 164) and Experiment 2 (ΔAIC = 84.6), for combined evidence favoring the DyNR model of ΔAIC = 248.6.

Cue processing

In the DyNR model, we assumed that identifying the target stimulus based on the cue is time-consuming, and becomes more so as the number of alternatives increases. Cue processing time encompasses perceptual, attentional, and decision components needed to interpret and act on the cue. We tested the necessity of this component by fitting a model variant in which VWM started accumulating evidence about the cued item at the moment of cue presentation. This model provided a worse fit to empirical data from both Experiment 1 (ΔAIC = 84.5) and Experiment 2 (ΔAIC = 107.5), for total evidence in favor of the DyNR model of ΔAIC = 192 (*Appendix 3—figure 4*). We fit another variant in which cue processing time was constant across set sizes. This alternative provided a worse fit to the data in Experiment 1 (ΔAIC = 191.6) and Experiment 2 (ΔAIC = 105), for combined evidence ΔAIC = 296.6 in favor of the full DyNR model that assumes cue processing time increases with set size. These results corroborate previous findings on the important role of cue processing time in models of attention (*Shih and Sperling, 2002*) and IM (*Sperling, 2018*).

Constant accumulation rate

In the DyNR model, the rate of accumulation into VWM is proportional to the difference between the present VWM amplitude and the maximum normalized amplitude (*Equation 2*). An arguably simpler assumption is that the neural signal approaches saturation at a constant rate (*Boerlin and Denève, 2011*; *Beck et al., 2008*). In particular, the rate at which the signal representing an item is transferred to VWM is constant and depends only on the number of encoded items, i.e.,

$$\dot{\gamma}_{\text{wm}}(t) = \begin{cases} \gamma_{\text{s}}(t)/(M(t)\tau_{\text{wm}}) & \text{if} \quad \gamma_{\text{wm}}(t) < \tilde{\gamma}_{\text{wm}}/M(t) \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

The dependence on $M(t)$ satisfies the constraint that the neural resources in VWM are allocated at a constant rate, irrespective of the number of items. We applied this model to psychophysical data from both experiments (*Appendix 3—figure 5*) and found it provides a worse fit to the data from

Experiment 1 (ΔAIC = 11.5) and Experiment 2 (ΔAIC = 36.2), for combined evidence favoring the DyNR model with exponential saturation (ΔAIC = 47.7).

Discussion

In the present study, we investigated the temporal dynamics of short-term recall fidelity. We conducted two new human psychophysical experiments and analyzed two existing datasets in order to characterize how recall errors are influenced by set size, stimulus duration, and retention interval. We developed a DyNR model to provide a mechanistic explanation of the observed behavior, capturing not only changes in overall fidelity but also the distribution of errors in the stimulus space and frequencies of swaps (intrusion errors). A key finding is that the benefit to recall precision observed at very short delays is due to additional post-cue integration of sensory information into WM, and that direct retrieval from sensory memory is unable to account for the empirical patterns of error.

Sensory and WM dynamics during delay

In the first experiment we investigated the effects of brief unfilled delays on recall fidelity. With multi-item arrays, we observed that memory performance deteriorates precipitously over the first few hundred milliseconds after stimuli disappear, followed by a gradual leveling-off of error with longer delays (*Figure 4*). These results are consistent with previously reported patterns of memory dynamics (*Di Lollo and Dixon, 1988*; *Sperling, 1960*; *Bradley and Pearson, 2012*; *Neisser, 1967*), and estimates of sensory decay ranging between 100 ms and 400 ms (*Loftus et al., 1992*; *Lu et al., 2005*). Here, we shed new light on these results by taking a computational approach in explaining observed temporal dynamics, and asking what this superior recall's neural origin is and its relation with VWM. To answer these questions, we adapted the Neural Resource model of *Bays, 2014*, with a temporal component. The new DyNR model considers dynamics in a sensory neural population registering the stimuli and in a VWM population that stores the stimuli for later recall. Critically, our model assumes that objects encoded with limited precision into VWM can be flexibly supplemented with sensory activity following a recall cue, within a brief temporal window while the sensory population provides a feedforward input post-stimulus. The boost in the representational VWM signal predicts a behavioral benefit of early cues that is consistent with our data and a large corpus of previous experiments (*Coltheart, 1980*).

A common assumption in studies of visual short-term memory is that recall over brief delays is exclusively supported by one of two memory stores, IM or VWM (*Bradley and Pearson, 2012*; *Pratte, 2018*). In this account, a cue presented within the first few hundred milliseconds after stimulus offset allows observers to access high resolution but rapidly deteriorating representations in IM; once the information in IM has decayed, objects must be retrieved from the capacity-limited VWM store. Two pieces of evidence from the current study contradict this view and strongly suggest that recall depends on VWM from the moment objects disappear. First, the recall benefit of short delays was not observed for one item arrays. We propose that this behavior reflects the fact that, during encoding, the entirety of the VWM resource is allocated to a single object, leaving no free capacity for further enhancement based on the available sensory signal post-cue. Second, we found clear evidence that recall fidelity varied with set size even with no delay between stimulus offset and cue (0 ms condition). We argue that this arises from the set size dependence of representational fidelity in VWM, which is only incompletely compensated by integration of the decaying sensory signal post-cue, resulting in lower fidelity for higher set sizes. The DyNR model provides a successful quantitative account for these findings, which are in clear contrast with the traditional view of IM.

The rapid changes in fidelity over short delays can be distinguished from dynamics over longer retention intervals. A number of recent studies have observed a slow deterioration of VWM precision over the course of prolonged retention (*Schneegans and Bays, 2018*; *Pertzov et al., 2017*; *Rademaker et al., 2018*; *Ricker et al., 2014*; *Shin et al., 2017*; *Zhang and Luck, 2009*). The causes of this deterioration are still contested, but growing evidence links this behavior to noise-driven diffusion. At a mechanistic level, diffusion is considered a fundamental property of continuous attractor networks of the kind commonly associated with models of WM (*Brody et al., 2003*; *Khona and Fiete, 2022*). In such networks, memorized features are represented as persistent activity 'bumps' in the network's representational feature space. Over a memory delay, the activity bump is sustained by balanced

excitatory and inhibitory connections, while stochasticity in neural activity causes shifts of the bump along the feature dimension, taking the form of a random walk. Although we did not model the network processes governing stability and diffusion within neural populations, our implementation is consistent with random (Brownian) perturbation, as assumed by attractor models (see also *Schnee-gans and Bays, 2018*).

Our theoretical account of memory dynamics during delay differs from several existing models of forgetting, which emphasize diffusion as the dominant source of error in short-term memory (e.g. *Panichello et al., 2019*; *Koyluoglu et al., 2017*). To solely account for the observed data in Experiment 1, diffusion would need to be strongest early in the retention period, followed by a much weaker diffusion with longer delays. However, it is unclear why the diffusion rate would change, and particularly slow down, with time. Assuming a constant neural signal encoding the stimulus, this would predict greater variability in neural activity initially compared to the later period after stimuli offset. This is inconsistent with electrophysiological data showing relatively stable levels of spiking variability throughout the memory delay period (*Khanna et al., 2019*; *Chang et al., 2012*; *Hussar and Pasternak, 2010*). The results observed here are consistent with the proposal that modulation of neural signal over short memory intervals accounts for an abrupt change in response fidelity, while diffusion accounts for a slower change that grows with time.

In the present study, a model assuming a constant diffusion rate, independent of the stored number of items, was preferred to one in which diffusion rate increases linearly with set size. This is consistent with results of *Shin et al., 2017*, who did not find a significant effect of set size on the rate of memory deterioration. In contrast to that, *Koyluoglu et al., 2017*, recently proposed that the rate of diffusion scales with set size. However, this study did not account for the presence of swap errors, which we found to increase with retention interval as well as set size. To draw strong conclusions about the dependence of diffusion on set size would require a future study to disentangle the different sources of error that could, in principle, increase with delay.

Sensory and WM dynamics during encoding

Having investigated memory degradation during the retention interval, in Experiment 2 we focused on the dynamics arising from accumulation of information during stimulus presentation. Using new psychophysical data, we showed that encoding of information into VWM is contingent on both presentation duration and the number of memorized stimuli. The observed patterns of data indicate that VWM encoding of elementary stimuli is mostly completed within the first 200 ms of presentation even at the largest set sizes, with minimal benefit of longer exposures, extending previous work (*Bays et al., 2011*; *Shibuya and Bundesen, 1988*; *Vogel et al., 2006*). This fast encoding process may have an adaptive role: with a key function of VWM to store and accumulate information across saccadic eye movements, an efficient system should deploy its resources within the duration of a typical gaze fixation (*Aagten-Murphy and Bays, 2018*; *Rolfs and Schweitzer, 2022*).

Our aim was again to move beyond the description of the encoding dynamics and to provide a biologically plausible neurocomputational account of these dynamics. To achieve that, we applied the same VWM accumulation process that operates post-cue to the sensory information during stimulus presentation. Using previously published and newly collected data, we show that a model in which VWM accumulates dynamical sensory input up to a fidelity limit can successfully account for patterns of human recall errors with variable set size and stimulus presentation. An important result of our modeling is that the accumulated information in VWM increases with a rate proportional to unfilled capacity. In particular, the model with such exponential accumulation provided a better fit than a model assuming a constant encoding rate. This parallels previous observations that models based on exponential-like extraction of information successfully characterize attention (*Bundesen, 1990*; *Sperling and Weichselgartner, 1995*), WM encoding (*Bays et al., 2011*; *Smith and Ratcliff, 2009*), memory updating (*Oberauer and Kliegl, 2006*), and broader cognitive processes (*Usher and McClelland, 2001*; *McClelland, 1979*). We hypothesize that this pattern represents an approach to an equilibrium state of balanced excitation from the sensory input and lateral inhibition within the VWM population, which is the basis for capacity of the memory system.

In Experiment 2, the longest presentation duration shows an upward trend in error at set sizes 4 and 10. While this falls within the range of measurement error, it is also possible that this is a meaningful pattern arising from visual adaptation of the sensory signal, whereby neural populations reduce

their activity after prolonged stimulation. This would mean less residual sensory signal would be available after the cue to supplement VWM activity, predicting a decline in fidelity at higher set sizes. Visual adaptation has previously been successfully accounted for by a type of delayed normalization model in which the sensory signal undergoes a series of linear and nonlinear transformations (*Zhou et al., 2019*). Such a model could in future be incorporated into DyNR and validated against psychophysical and neural data.

Our computational account of VWM encoding dynamics differs from several existing modeling frameworks aiming to explain similar data. For example, the theory of visual attention (TVA; *Bundesen, 1990*) assumes that visual stimuli participate in a parallel exponential race toward limited VWM. Like the DyNR model, TVA assumes a form of normalization in the sense that the speed with which items race toward VWM depends on the number of items in the visual field. Unlike our dynamic model, TVA is not a theory of VWM, and it considers VWM only as a storage for categorizations of visual objects. In particular, TVA takes into account the limits of VWM but does not specify why or how these limitations arise. Finally, TVA considers whether an object was selected for entry into VWM in an all-or-none fashion; our dynamic model is mostly concerned with the fidelity of representations. A somewhat alternative account of VWM encoding is provided by the competitive interaction theory (CIT; *Sewell et al., 2014*), which is similarly based on the signal detection theory and principles of normalization (*Reynolds and Heeger, 2009*). Like TVA, CIT is mostly focused on item selection and merely incorporates a concept of VWM capacity derived from object-based models of VWM. Although CIT had success in accounting for behavioral data from a two-alternative orthogonal discrimination task using up to four items and a limited range of encoding times, it remains an open question whether this model can account for error distributions as measured in a continuous report task, and a larger range of set sizes and stimulus exposures. Importantly, compared to both TVA and CIT, the DyNR model is strongly rooted in and inspired by findings from neuroscience. This not only adds to the biological plausibility of our model but also allows future studies to test the model's predictions using physiological methods.

Neural mechanisms

The theory presented here generalizes the Neural Resource model of *Bays, 2014*, a simple encoding-decoding model in which visual features are represented in the noisy spiking activity of neural populations (*Pouget et al., 2000*), and where the activity representing each feature scales inversely with the total number of representations, consistent with the prevalence of normalization mechanisms in the brain and observations from single-neuron recording (*Buschman et al., 2011*) and fMRI decoding (*Sprague et al., 2014*) studies. The population coding in the model is based on an abstract idealization of neural response functions. Nevertheless, it has recently been shown that more realistic population coding schemes that allow for heterogeneity in neural tuning curves and correlated spiking activity as observed in visual cortex maintain the key predictions of the idealized model (*Taylor and Bays, 2020*; *Schneegans et al., 2020*). This may be seen as a consequence of the different population codes inducing a common representational geometry (*Kriegeskorte and Wei, 2021*).

We adapted the stationary VWM model by first incorporating a sensory population that provides an input drive to the VWM population. In parallel with neurophysiological observations, a common approach is to model these dynamics with a low-pass filter which acts like a neural gain modulation mechanism (*Hawken et al., 1996*). As a consequence, the sensory response to stimulus onset and offset is an exponential rise and decay in activity, respectively. The decaying component of the response has been recognized as a neural substrate of visual persistence and IM (*van Kerkoerle et al., 2017*; *Teeuwen et al., 2021*). Here, we modeled sensory decay with an exponential function (*Zylberberg et al., 2009*), although other forms of decay have been proposed. For example, *Loftus et al., 1992*, showed that iconic decay could be better captured using a gamma survival function, a generalization of exponential decay that could simply be implemented in our neural model by replacing a single filter with a cascade of exponential low-pass filters.

In addition to the dynamics in the sensory population, two features of VWM introduce additional dynamics in representation fidelity: the accumulation of information (discussed above) and the diffusion of representations owing to accumulated noise. Although we did not aim to model the neural processes behind diffusion, our implementation is consistent with the consequences of neural variability in attractor networks (*Burak and Fiete, 2012*; *Khona and Fiete, 2022*). Converging neural

evidence demonstrating such diffusion has been observed using single-unit neural recording in monkeys (*Wimmer et al., 2014*), as well as EEG (*Wolff et al., 2020*) and fMRI (*Lim et al., 2019*; *Yu et al., 2020*) studies in humans.

As well as being implicated in higher cognitive processes including VWM (*Buschman et al., 2011*; *Sprague et al., 2014*), divisive normalization has been shown to be widespread in basic sensory processing (*Bonin et al., 2005*; *Busse et al., 2009*; *Ni and Maunsell, 2017*). The DyNR model presently incorporates the former but not the latter type of normalization. While the data observed in our experiments do not provide evidence for normalization of sensory signals (note comparable recall errors across set size in the simultaneous cue condition of Experiment 1), this may be because sensory suppressive effects are localized and our stimuli were relatively widely separated in the visual field: future research could explore the consequences of sensory normalization for recall from VWM using, e.g., center-surround stimuli (*Bloem et al., 2018*).

Following onset of a stimulus, the visual signal ascends through visual areas via a cascade of feedforward connections. This feedforward sweep conveys sensory information that persists during stimulus presentation and briefly after it disappears (*Lamme et al., 1998*). Simultaneously, reciprocal feedback connections carry higher-order information back toward antecedent cortical areas (*Lamme and Roelfsema, 2000*). In our psychophysical task, feedback connections likely play a critical role in orienting attention toward the cued item, facilitating the extraction of persisting sensory signals, and potentially signaling continuous information on the available resources for VWM encoding. While our computational study does not address the nature of these feedforward and feedback signals, a challenge for future research is to describe the relative contributions of these signals in mediating transmission of information between sensory memory and WM (*Semedo et al., 2022*).

Our model makes a clear distinction between dynamics in sensory and VWM populations, however, it remains agnostic as to whether the populations have the same or different anatomical locus (*Rademaker et al., 2019*). Albeit inspired by the properties of orientation-selective neurons in area V1, population tuning of this kind is a common coding motif across the brain (*Pouget et al., 2000*). While it could be considered efficient to use already specialized circuits to maintain as well as process visual information, it is still debated whether sensory areas are a feasible candidate for memory storage (*Serences, 2016*; *Xu, 2017*). While some studies have focused on prefrontal (*Goldman-Rakic, 1995*), parietal (*Bettencourt and Xu, 2016*), or occipital (*Harrison and Tong, 2009*) cortices as the primary locus of VWM, others argue for distributed storage by demonstrating that VWM contents can be decoded from imaging signals originating in multiple brain areas (*Christophel et al., 2018*).

Representational dynamics of cue-dimension features

Memory retrieval failures in which a non-cued item is reported in place of the intended target represent an important source of error in VWM recall. These swap errors occur more often at higher set sizes and when spatial confusability is high (*Bays et al., 2009*; *Emrich and Ferber, 2012*; *Rerko et al., 2014*; *Bays, 2016b*), as predicted by models in which they arise from uncertainty in the recall of cue-dimension features leading to incorrect selection of an item in memory (*Schneegans and Bays, 2017*; *McMaster et al., 2022*). In the current study, we assumed memory for spatial location (the cue feature) undergoes similar dynamics to memory for orientation (the report feature), and in particular that spatial information degrades with retention time (*Schneegans and Bays, 2018*), leading to changes in swap error frequency with delay interval. Similarly, during encoding the fidelity of spatial representation increases with the accumulation of sensory evidence (*Zimmermann et al., 2013*), reducing the uncertainty at retrieval and consequently swap errors at longer stimulus exposure. Although we did not explicitly model the neural signals representing location, the modeled dynamics in the probability of swap errors were consistent with those of the primary memory feature. We provided a more detailed neural account of swap errors in our earlier works that is theoretically compatible with the DyNR model (*Schneegans and Bays, 2017*; *McMaster et al., 2022*).

The DyNR model successfully captured the observed pattern of swap frequencies (intrusion errors). The only notable discrepancy between DyNR and the three-component mixture model (*Appendix 2—figure 1*) arises with the largest set size and longest delay, although with considerable interindividual variability. As the variability in report dimension increases, the estimates of swap frequency become more variable due to the growing overlap between the probability distributions of swap and non-swap

responses. This may explain apparent deviations from the modeled swap frequencies with the highest set size and longest delay where orientation response variability was greatest.

Removal of information from WM

In the DyNR model, taking advantage of early cues requires rapid removal of the VWM signal associated with uncued items, to admit further accumulation of activity encoding the cued item. To achieve this, an active process of selective content elimination may be required (*Oberauer, 2018*), as opposed to a passive decay of uncued representations during the post-cue interval. Mounting evidence for such active removal has been provided at the behavioral (*Williams et al., 2013*) and neural (*LaRocque et al., 2013*) level. Importantly, studies show that a functional role of such active removal is to release resources allocated to the uncued representations, facilitating the encoding of new information (*Taylor et al., 2023*; *Souza et al., 2014*). The fast reallocation of neural resources assumed by the DyNR model is consistent with such a description of active removal.

Methods

Participants

A total of 23 naive observers (12 females, 11 males; aged 18–34) took part in the study after giving informed consent in accordance with the Declaration of Helsinki. Ten observers participated in Experiment 1 and 13 observers participated in Experiment 2. Volunteers were recruited through the Cambridge Psychology research sign-up system. All observers reported normal color vision and normal or corrected-to-normal visual acuity, and were remunerated £10/hr for their participation. Procedures were approved by the University of Cambridge Psychology Research Ethics Committee (approval number PRE.2015.099).

General methods

Experimental setup

Stimuli were presented on a 69 cm gamma-corrected LCD monitor with a refresh rate of 144 Hz. Participants were seated in a dark room and viewed the monitor at a distance of 60 cm, with their head supported by a forehead and chin rest. Responses were collected using Magic Trackpad 2, a pointing device (16×11.5 cm^2) with a tactile sensor operating at ~90 Hz (Apple Inc). Eye position was monitored online at 1000 Hz using an infrared eye tracker (SR Research). Stimulus presentation and response registration were controlled by a script written in Psychtoolbox and run using Matlab (The Mathworks Inc).

Stimuli

Memory stimuli consisted of randomly oriented Gabor patches (wavelength of the sinusoid, 0.65° of visual angle; s.d. of Gaussian envelope, 0.5°) presented on a uniform mid-gray background. The contrast of Gabor patches varied between experiments (see below). Memory stimulus positions were randomly chosen from a set of 10 equidistant locations on the perimeter of an invisible circle with radius 6° centered at fixation. At the start of each trial, a black fixation annulus was shown ($r = 0.15°$ and $R = 0.25°$) in the display center. Once steady fixation was registered, the size of the inner radius increased ($r = 0.2°$). Observers perceived this change as the annulus becoming thinner. The fixation annulus then stayed visible throughout the trial. Items were cued for recall by displaying a black arrow (2° length) extending from the center of the display and pointing to one of the previously occupied locations without overlapping with it.

Procedure

Each trial started with presentation of the central fixation annulus. Observers were required to maintain gaze fixation for 500 ms within a radius of 2° around the central annulus in order for a trial to proceed. Following stable fixation, the appearance of the fixation annulus changed, indicating that the memory array would appear in 500 ms. The memory sample array consisting of 1, 4, or 10 randomly oriented Gabor patches was then presented. This was followed by a delay period and finally a cue

display, indicating to observers to report the memorized orientation of an item previously displayed at the indicated location.

Observers were instructed to reproduce the remembered orientation as accurately and as quickly as possible by executing a single movement of their index fingertip over the surface of the touchpad located centrally in front of them. Simultaneously with the observer's movement, a blue line appeared on the screen, extending from the center of the screen and mimicking the observer's response in real time. The response was terminated if one of the following conditions was satisfied: the observer stopped movement for 500 ms; the observer lifted their finger from the touchpad; or the response line reached the edge of the display. This was followed by a feedback display, consisting of the actual orientation (shown with a white line) and reported orientation (shown with a blue line) overlaid at the location of the cued item. The recalled orientation was calculated as the angle of the line connecting a starting point and an endpoint of hand movement on the touchpad.

Observers were required to maintain central fixation during the stimulus presentation and delay phase. If gaze position deviated by more than 2° a message appeared on the screen, and the trial was aborted and restarted with newly randomized orientations. Participants completed the task in blocks of 50 trials, and each block corresponded to one experimental condition. The order of blocks was randomized for every observer. At the beginning of the testing session observers familiarized themselves with the task and experimental setup by doing at most 50 practice trials.

Experiment 1

In Experiment 1 we investigated the temporal dynamics of VWM fidelity over short delays by presenting observers with sets of stimuli of variable size and then cueing one of them for recall after a variable delay relative to the stimuli offset. A typical trial sequence is shown in *Figure 3A*. The memory sample array (Michelson contrast = 0.5) was presented for 200 ms. In 50% of trials, the stimuli changed phase (by 180°) and contrast (Michelson contrast = 1) for the last 50 ms of presentation, while remaining at the same orientation. This manipulation was intended to minimize retinal after-effects (see, e.g., *Kelly and Martinez-Uriegas, 1993*, for similar techniques and Appendix 1 for validation). The stimuli offset was followed by a variable blank delay of 0, 100, 200, 400, or 1000 ms, after which one item was cued for recall. In one additional condition, the cue was instead presented simultaneously with the memory sample array, indicating an item while it was still visible on the screen (*Figure 3B*).

Each observer completed a total of 1800 trials, split into 36 blocks. The experiment was organized such that half of the observers first completed 18 blocks with phase shift (see above), and the other half first completed blocks without phase shift. Except for this constraint, block order was randomized for every observer. The testing was divided into four equal testing sessions, each lasting approximately 1.5 hr, with a separation of at least 1 day between sessions.

Experiment 2

In Experiment 2 we investigated the temporal dynamics of VWM fidelity during encoding. To this end, we displayed oriented stimuli for a variable duration and in sets of variable size. The experiment was similar to the previous experiment with a few exceptions (*Figure 3C*). Each trial started with a presentation of a fixation annulus, followed by a memory array (Michelson contrast = 0.3). The stimuli stayed on the screen for a variable duration of 30, 48, 77, 122, 196, 313, or 500 ms, and were then replaced by noise masks (100 ms). Mask stimuli consisted of white noise at full contrast, windowed with a Gaussian envelope (0.5° s.d.) and flickering at 35 Hz. At the offset of the masking stimuli, one memory item was cued for recall. Each observer completed 21 blocks, for a total of 1050 trials. Blocks were spread over two testing sessions, each lasting approximately 1.5 hr, and taking place on different days. Observers completed 10 blocks in the first, and the remaining 11 blocks in the second session.

Acknowledgements

Additional information

Author contributions

Ivan Tomić, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing; Paul M Bays, Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – review and editing

Author ORCIDs

Ivan Tomić ⬡ https://orcid.org/0000-0001-6046-9642
Paul M Bays ⬡ https://orcid.org/0000-0003-4684-4893

Ethics

Human subjects: Procedures were approved by the University of Cambridge Psychology Research Ethics Committee (approval number PRE.2015.099). All subjects gave informed consent in accordance with the Declaration of Helsinki.

Peer review material

Reviewer #2 (Public Review): https://doi.org/10.7554/eLife.91034.3.sa1
Reviewer #3 (Public Review): https://doi.org/10.7554/eLife.91034.3.sa2
Author response https://doi.org/10.7554/eLife.91034.3.sa3

Additional files

Supplementary files
• MDAR checklist

Data availability

Data and code related to this study will be made available at https://doi.org/10.17863/CAM.95223.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Tomić I, Bays PM	2024	Research data supporting 'A dynamic neural resource model bridges sensory and working memory'	https://doi.org/10.17863/CAM.95223	Apollo - University of Cambridge Repository, 10.17863/CAM.95223

References

Aagten-Murphy D, Bays PM. 2018. Functions of memory across saccadic eye movements. Hodgson T (Ed). *Processes of Visuospatial Attention and Working Memory*. Springer. p. 155–183. DOI: https://doi.org/10.1007/7854_2018_66

Aksay E, Gamkrelidze G, Seung HS, Baker R, Tank DW. 2001. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature Neuroscience* **4**:184–193. DOI: https://doi.org/10.1038/84023, PMID: 11175880

Barlow HB. 1981. The Ferrier Lecture, 1980. Critical limiting factors in the design of the eye and visual cortex. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **212**:1–34. DOI: https://doi.org/10.1098/rspb.1981.0022, PMID: 6115386

Bays PM, Catalao RFG, Husain M. 2009. The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision* **9**:7. DOI: https://doi.org/10.1167/9.10.7, PMID: 19810788

Bays PM, Gorgoraptis N, Wee N, Marshall L, Husain M. 2011. Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision* **11**:6. DOI: https://doi.org/10.1167/11.10.6, PMID: 21911739

Bays PM. 2014. Noise in neural populations accounts for errors in working memory. *The Journal of Neuroscience* **34**:3632–3645. DOI: https://doi.org/10.1523/JNEUROSCI.3204-13.2014, PMID: 24599462

Bays PM. 2015. Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences* **19**:431–438. DOI: https://doi.org/10.1016/j.tics.2015.06.004

Bays PM. 2016a. A signature of neural coding at human perceptual limits. *Journal of Vision* **16**:4. DOI: https://doi.org/10.1167/16.11.4, PMID: 27604067

Bays PM. 2016b. Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports* **6**:19203. DOI: https://doi.org/10.1038/srep19203, PMID: 26758902

Bays PM, Taylor R. 2018. A neural model of retrospective attention in visual working memory. *Cognitive Psychology* **100**:43–52. DOI: https://doi.org/10.1016/j.cogpsych.2017.12.001, PMID: 29272732

Bays P, Schneegans S, Ma WJ, Brady TF. 2024. Representation and computation in working memory. *Nature Human Behaviour*. In press.

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A. 2008. Probabilistic population codes for Bayesian decision making. *Neuron* **60**:1142–1152. DOI: https://doi.org/10.1016/j.neuron.2008.09.021, PMID: 19109917

Bettencourt KC, Xu Y. 2016. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience* **19**:150–157. DOI: https://doi.org/10.1038/nn.4174

Bloem IM, Watanabe YL, Kibbe MM, Ling S. 2018. Visual memories bypass normalization. *Psychological Science* **29**:845–856. DOI: https://doi.org/10.1177/0956797617747091, PMID: 29596038

Boerlin M, Denève S. 2011. Spike-based population coding and working memory. *PLOS Computational Biology* **7**:e1001080. DOI: https://doi.org/10.1371/journal.pcbi.1001080, PMID: 21379319

Bonin V, Mante V, Carandini M. 2005. The suppressive field of neurons in lateral geniculate nucleus. *The Journal of Neuroscience* **25**:10844–10856. DOI: https://doi.org/10.1523/JNEUROSCI.3562-05.2005

Bouchacourt F, Buschman TJ. 2019. A flexible model of working memory. *Neuron* **103**:147–160.. DOI: https://doi.org/10.1016/j.neuron.2019.04.020, PMID: 31103359

Braak CJFT. 2006. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing* **16**:239–249. DOI: https://doi.org/10.1007/s11222-006-8769-1

Bradley C, Pearson J. 2012. The sensory components of high-capacity iconic memory and visual working memory. *Frontiers in Psychology* **3**:355. DOI: https://doi.org/10.3389/fpsyg.2012.00355, PMID: 23055993

Brody CD, Romo R, Kepecs A. 2003. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology* **13**:204–211. DOI: https://doi.org/10.1016/s0959-4388(03)00050-3, PMID: 12744975

Brunton BW, Botvinick MM, Brody CD. 2013. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**:95–98. DOI: https://doi.org/10.1126/science.1233912, PMID: 23559254

Bundesen C. 1990. A theory of visual attention. *Psychological Review* **97**:523–547. DOI: https://doi.org/10.1037/0033-295x.97.4.523, PMID: 2247540

Burak Y, Fiete IR. 2012. Fundamental limits on persistent activity in networks of noisy neurons. *PNAS* **109**:17645–17650. DOI: https://doi.org/10.1073/pnas.1117386109

Buschman TJ, Siegel M, Roy JE, Miller EK. 2011. Neural substrates of cognitive capacity limitations. *PNAS* **108**:11252–11255. DOI: https://doi.org/10.1073/pnas.1104666108

Busse L, Wade AR, Carandini M. 2009. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**:931–942. DOI: https://doi.org/10.1016/j.neuron.2009.11.004

Carandini M, Heeger DJ. 2011. Normalization as a canonical neural computation. *Nature Reviews. Neuroscience* **13**:51–62. DOI: https://doi.org/10.1038/nrn3136, PMID: 22108672

Chang MH, Armstrong KM, Moore T. 2012. Dissociation of response variability from firing rate effects in frontal eye field neurons during visual stimulation, working memory, and attention. *The Journal of Neuroscience* **32**:2204–2216. DOI: https://doi.org/10.1523/JNEUROSCI.2967-11.2012, PMID: 22323732

Christophel TB, Iamshchinina P, Yan C, Allefeld C, Haynes JD. 2018. Cortical specialization for attended versus unattended working memory. *Nature Neuroscience* **21**:494–496. DOI: https://doi.org/10.1038/s41593-018-0094-4, PMID: 29507410

Coltheart M. 1980. Iconic memory and visible persistence. *Perception & Psychophysics* **27**:183–228. DOI: https://doi.org/10.3758/BF03204258

Compte A, Brunel N, Goldman-Rakic PS, Wang XJ. 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex* **10**:910–923. DOI: https://doi.org/10.1093/cercor/10.9.910

Derrington AM, Lennie P. 1984. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology* **357**:219–240. DOI: https://doi.org/10.1113/jphysiol.1984.sp015498

D'Esposito M, Postle BR. 2015. The cognitive neuroscience of working memory. *Annual Review of Psychology* **66**:115–142. DOI: https://doi.org/10.1146/annurev-psych-010814-015031, PMID: 25251486

Di Lollo V, Dixon P. 1988. Two forms of persistence in visual information processing. *Journal of Experimental Psychology. Human Perception and Performance* **14**:671–681. DOI: https://doi.org/10.1037/0096-1523.14.4. 671, PMID: 2974875

Doost R, Turvey MT. 1971. Iconic memory and central processing capacity. *Perception & Psychophysics* **9**:269–274. DOI: https://doi.org/10.3758/BF03212646

Emrich SM, Ferber S. 2012. Competition increases binding errors in visual working memory. *Journal of Vision* **12**:12. DOI: https://doi.org/10.1167/12.4.12, PMID: 22523399

Gold JI, Shadlen MN. 2007. The neural basis of decision making. *Annual Review of Neuroscience* **30**:535–574. DOI: https://doi.org/10.1146/annurev.neuro.29.051605.113038, PMID: 17600525

Goldman-Rakic PS. 1995. Cellular basis of working memory. *Neuron* **14**:477–485. DOI: https://doi.org/10.1016/0896-6273(95)90304-6

Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature* **458**:632–635. DOI: https://doi.org/10.1038/nature07832, PMID: 19225460

Hawken MJ, Shapley RM, Grosof DH. 1996. Temporal-frequency selectivity in monkey visual cortex. *Visual Neuroscience* **13**:477–492. DOI: https://doi.org/10.1017/s0952523800008154, PMID: 8782375

Hess RF, Snowden RJ. 1992. Temporal properties of human visual filters: number, shapes and spatial covariation. *Vision Research* **32**:47–59. DOI: https://doi.org/10.1016/0042-6989(92)90112-V

Hick WE. 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology* **4**:11–26. DOI: https://doi.org/10.1080/17470215208416600

Hussar C, Pasternak T. 2010. Trial-to-trial variability of the prefrontal neurons reveals the nature of their engagement in a motion discrimination task. *PNAS* **107**:21842–21847. DOI: https://doi.org/10.1073/pnas. 1009956107

Jazayeri M, Movshon JA. 2006. Optimal representation of sensory information by neural populations. *Nature Neuroscience* **9**:690–696. DOI: https://doi.org/10.1038/nn1691, PMID: 16617339

Kelly DH, Martinez-Uriegas E. 1993. Measurements of chromatic and achromatic afterimages. *Journal of the Optical Society of America. A, Optics and Image Science* **10**:29–37. DOI: https://doi.org/10.1364/josaa.10. 000029, PMID: 8478743

Khanna SB, Snyder AC, Smith MA. 2019. Distinct sources of variability affect eye movement preparation. *The Journal of Neuroscience* **39**:4511–4526. DOI: https://doi.org/10.1523/JNEUROSCI.2329-18.2019, PMID: 30914447

Khona M, Fiete IR. 2022. Attractor and integrator networks in the brain. *Nature Reviews. Neuroscience* **23**:744–766. DOI: https://doi.org/10.1038/s41583-022-00642-0, PMID: 36329249

Kovács G, Vogels R, Orban GA. 1995. Cortical correlate of pattern backward masking. *PNAS* **92**:5587–5591. DOI: https://doi.org/10.1073/pnas.92.12.5587

Koyluoglu OO, Pertzov Y, Manohar S, Husain M, Fiete IR. 2017. Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *eLife* **6**:e22225. DOI: https://doi.org/10. 7554/eLife.22225, PMID: 28879851

Kriegeskorte N, Wei XX. 2021. Neural tuning and representational geometry. *Nature Reviews. Neuroscience* **22**:703–718. DOI: https://doi.org/10.1038/s41583-021-00502-3, PMID: 34522043

Lamme VA, Supèr H, Spekreijse H. 1998. Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology* **8**:529–535. DOI: https://doi.org/10.1016/s0959-4388(98)80042-1, PMID: 9751656

Lamme VA, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences* **23**:571–579. DOI: https://doi.org/10.1016/s0166-2236(00)01657-x, PMID: 11074267

LaRocque JJ, Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. 2013. Decoding attended information in short-term memory: an EEG study. *Journal of Cognitive Neuroscience* **25**:127–142. DOI: https://doi.org/10. 1162/jocn_a_00305, PMID: 23198894

Lee BB, Martin PR, Valberg A. 1989. Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *The Journal of Physiology* **414**:223–243. DOI: https://doi.org/10.1113/jphysiol.1989.sp017685

Lim PC, Ward EJ, Vickery TJ, Johnson MR. 2019. Not-so-working memory: drift in functional magnetic resonance imaging pattern representations during maintenance predicts errors in a visual working memory task. *Journal of Cognitive Neuroscience* **31**:1520–1534. DOI: https://doi.org/10.1162/jocn_a_01427, PMID: 31112474

Loftus GR, Duncan J, Gehrig P. 1992. On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology* **18**:530–549. DOI: https://doi.org/10.1037/0096-1523. 18.2.530

Lu ZL, Neuse J, Madigan S, Dosher BA. 2005. Fast decay of iconic memory in observers with mild cognitive impairments. *PNAS* **102**:1797–1802. DOI: https://doi.org/10.1073/pnas.0408402102

Ma WJ, Beck JM, Latham PE, Pouget A. 2006. Bayesian inference with probabilistic population codes. *Nature Neuroscience* **9**:1432–1438. DOI: https://doi.org/10.1038/nn1790, PMID: 17057707

Ma WJ, Husain M, Bays PM. 2014. Changing concepts of working memory. *Nature Neuroscience* **17**:347–356. DOI: https://doi.org/10.1038/nn.3655, PMID: 24569831

McClelland JL. 1979. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review* **86**:287–330. DOI: https://doi.org/10.1037/0033-295X.86.4.287

McMaster JMV, Tomić I, Schneegans S, Bays PM. 2022. Swap errors in visual working memory are fully explained by cue-feature variability. *Cognitive Psychology* **137**:101493. DOI: https://doi.org/10.1016/j.cogpsych.2022. 101493, PMID: 35777189

Müller JR, Metha AB, Krauskopf J, Lennie P. 2001. Information conveyed by onset transients in responses of striate cortical neurons. *The Journal of Neuroscience* **21**:6978–6990. DOI: https://doi.org/10.1523/JNEUROSCI.21-17-06978.2001, PMID: 11517285

Neisser U. 1967. *Cognitive Psychology. Number 1966 in Century Psychology Series Award*. Appleton-Century-Crofts.

Ni AM, Maunsell JHR. 2017. Spatially tuned normalization explains attention modulation variance within neurons. *Journal of Neurophysiology* **118**:1903–1913. DOI: https://doi.org/10.1152/jn.00218.2017, PMID: 28701536

Oberauer K, Kliegl R. 2006. A formal model of capacity limits in working memory. *Journal of Memory and Language* **55**:601–626. DOI: https://doi.org/10.1016/j.jml.2006.08.009

Oberauer K. 2018. Removal of irrelevant information from working memory: sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Sciences* **1424**:239–255. DOI: https://doi.org/10.1111/nyas.13603, PMID: 29532484

Ohshiro T, Angelaki DE, DeAngelis GC. 2011. A normalization model of multisensory integration. *Nature Neuroscience* **14**:775–782. DOI: https://doi.org/10.1038/nn.2815, PMID: 21552274

Oram MW, Perrett DI. 1992. Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology* **68**:70–84. DOI: https://doi.org/10.1152/jn.1992.68.1.70, PMID: 1517829

Panichello MF, DePasquale B, Pillow JW, Buschman TJ. 2019. Error-correcting dynamics in visual working memory. *Nature Communications* **10**:3366. DOI: https://doi.org/10.1038/s41467-019-11298-3, PMID: 31358740

Pasternak T, Greenlee MW. 2005. Working memory in primate sensory systems. *Nature Reviews. Neuroscience* **6**:97–107. DOI: https://doi.org/10.1038/nrn1603, PMID: 15654324

Pertzov Y, Manohar S, Husain M. 2017. Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition* **43**:528–536. DOI: https://doi.org/10.1037/xlm0000328, PMID: 27668485

Pouget A, Dayan P, Zemel R. 2000. Information processing with population codes. *Nature Reviews. Neuroscience* **1**:125–132. DOI: https://doi.org/10.1038/35039062, PMID: 11252775

Pratte MS. 2018. Iconic memories die a sudden death. *Psychological Science* **29**:877–887. DOI: https://doi.org/10.1177/0956797617747118, PMID: 29671682

Priebe NJ, Churchland MM, Lisberger SG. 2002. Constraints on the source of short-term motion adaptation in macaque area MT. I. the role of input and intrinsic mechanisms. *Journal of Neurophysiology* **88**:354–369. DOI: https://doi.org/10.1152/jn.00852.2001, PMID: 12091560

Rademaker RL, Park YE, Sack AT, Tong F. 2018. Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance* **44**:925–940. DOI: https://doi.org/10.1037/xhp0000491, PMID: 29494191

Rademaker RL, Chunharas C, Serences JT. 2019. Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience* **22**:1336–1344. DOI: https://doi.org/10.1038/s41593-019-0428-x, PMID: 31263205

Rerko L, Oberauer K, Lin HY. 2014. Spatial transposition gradients in visual working memory. *Quarterly Journal of Experimental Psychology* **67**:3–15. DOI: https://doi.org/10.1080/17470218.2013.789543

Reynolds JH, Heeger DJ. 2009. The normalization model of attention. *Neuron* **61**:168–185. DOI: https://doi.org/10.1016/j.neuron.2009.01.002, PMID: 19186161

Ricker TJ, Spiegel LR, Cowan N. 2014. Time-based loss in visual short-term memory is from trace decay, not temporal distinctiveness. *Journal of Experimental Psychology* **40**:1510–1523. DOI: https://doi.org/10.1037/xlm0000018

Ringach DL, Hawken MJ, Shapley R. 2003. Dynamics of orientation tuning in macaque V1: the role of global and tuned suppression. *Journal of Neurophysiology* **90**:342–352. DOI: https://doi.org/10.1152/jn.01018.2002, PMID: 12611936

Rolfs M, Schweitzer R. 2022. Coupling perception to action through incidental sensory consequences of motor behaviour. *Nature Reviews Psychology* **1**:112–123. DOI: https://doi.org/10.1038/s44159-021-00015-x

Rolls ET, Tovee M. 1994. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London. Series B* **257**:9–15. DOI: https://doi.org/10.1098/rspb.1994.0087

Schneegans S, Bays PM. 2017. Neural architecture for feature binding in visual working memory. *The Journal of Neuroscience* **37**:3913–3925. DOI: https://doi.org/10.1523/JNEUROSCI.3493-16.2017, PMID: 28270569

Schneegans S, Bays PM. 2018. Drift in neural population activity causes working memory to deteriorate over time. *The Journal of Neuroscience* **38**:4859–4869. DOI: https://doi.org/10.1523/JNEUROSCI.3440-17.2018, PMID: 29703786

Schneegans S, Taylor R, Bays PM. 2020. Stochastic sampling provides a unifying account of visual working memory limits. *PNAS* **117**:20959–20968. DOI: https://doi.org/10.1073/pnas.2004306117

Semedo JD, Jasper AI, Zandvakili A, Krishna A, Aschner A, Machens CK, Kohn A, Yu BM. 2022. Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nature Communications* **13**:1099. DOI: https://doi.org/10.1038/s41467-022-28552-w, PMID: 35232956

Serences JT. 2016. Neural mechanisms of information storage in visual short-term memory. *Vision Research* **128**:53–67. DOI: https://doi.org/10.1016/j.visres.2016.09.010

Sewell DK, Lilburn SD, Smith PL. 2014. An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology. Human Perception and Performance* **40**:2214–2242. DOI: https://doi.org/10.1037/a0037744, PMID: 25222469

Shibuya H, Bundesen C. 1988. Visual selection from multielement displays: measuring and modeling effects of exposure duration. *Journal of Experimental Psychology. Human Perception and Performance* **14**:591–600. DOI: https://doi.org/10.1037//0096-1523.14.4.591, PMID: 2974870

Shih SI, Sperling G. 2002. Measuring and modeling the trajectory of visual spatial attention. *Psychological Review* **109**:260–305. DOI: https://doi.org/10.1037/0033-295x.109.2.260, PMID: 11990319

Shin H, Zou Q, Ma WJ. 2017. The effects of delay duration on visual working memory for orientation. *Journal of Vision* **17**:10. DOI: https://doi.org/10.1167/17.14.10

Sit YF, Chen Y, Geisler WS, Miikkulainen R, Seidemann E. 2009. Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron* **64**:943–956. DOI: https://doi.org/10.1016/j.neuron.2009.08.041, PMID: 20064399

Smith PL, Ratcliff R. 2009. An integrated theory of attention and decision making in visual signal detection. *Psychological Review* **116**:283–317. DOI: https://doi.org/10.1037/a0015156, PMID: 19348543

Souza AS, Rerko L, Oberauer K. 2014. Unloading and reloading working memory: attending to one item frees capacity. *Journal of Experimental Psychology. Human Perception and Performance* **40**:1237–1256. DOI: https://doi.org/10.1037/a0036331, PMID: 24730737

Sperling G. 1960. The information available in brief visual presentations. *Psychological Monographs* **74**:1–29. DOI: https://doi.org/10.1037/h0093759

Sperling G, Weichselgartner E. 1995. Episodic theory of the dynamics of spatial attention. *Psychological Review* **102**:503–532. DOI: https://doi.org/10.1037/0033-295X.102.3.503

Sperling G. 2018. A brief overview of computational models of spatial, temporal, and feature visual attention. Lachmann T, Weis T (Eds). *Invariances in Human Information Processing*. Routledge. p. 143–182. DOI: https://doi.org/10.4324/9781315169903

Sprague TC, Ester EF, Serences JT. 2014. Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology* **24**:2174–2180. DOI: https://doi.org/10.1016/j.cub.2014.07.066, PMID: 25201683

Taylor R, Bays PM. 2020. Theory of neural coding predicts an upper bound on estimates of memory variability. *Psychological Review* **127**:700–718. DOI: https://doi.org/10.1037/rev0000189, PMID: 32191074

Taylor R, Tomić I, Aagten-Murphy D, Bays PM. 2023. Working memory is updated by reallocation of resources from obsolete to new items. *Attention, Perception, & Psychophysics* **85**:1437–1451. DOI: https://doi.org/10.3758/s13414-022-02584-2

Teeuwen RRM, Wacongne C, Schnabel UH, Self MW, Roelfsema PR. 2021. A neuronal basis of iconic memory in macaque primary visual cortex. *Current Biology* **31**:5401–5414.. DOI: https://doi.org/10.1016/j.cub.2021.09.052, PMID: 34653360

Tomić I, Bays PM. 2018. Internal but not external noise frees working memory resources. *PLOS Computational Biology* **14**:e1006488. DOI: https://doi.org/10.1371/journal.pcbi.1006488, PMID: 30321172

Tomić I, Bays PM. 2024. Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition* **50**:535–549. DOI: https://doi.org/10.1037/xlm0001172, PMID: 36442045

Tomić I, Girones Z, Lengyel M, Bays PM. 2024. Natural statistics and stimulus representations in visual working memory. Computational and Systems Neuroscience (CoSyNe). .

Tsuchiya N, Koch C. 2005. Continuous flash suppression reduces negative afterimages. *Nature Neuroscience* **8**:1096–1101. DOI: https://doi.org/10.1038/nn1500

Turvey MT. 1973. On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review* **80**:1–52. DOI: https://doi.org/10.1037/h0033872, PMID: 4689202

Usher M, McClelland JL. 2001. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review* **108**:550–592. DOI: https://doi.org/10.1037/0033-295x.108.3.550, PMID: 11488378

van Ede F, Chekroud SR, Stokes MG, Nobre AC. 2019. Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience* **22**:477–483. DOI: https://doi.org/10.1038/s41593-018-0335-6, PMID: 30718904

Van Essen DC, Anderson CH, Felleman DJ. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**:419–423. DOI: https://doi.org/10.1126/science.1734518, PMID: 1734518

van Kerkoerle T, Self MW, Roelfsema PR. 2017. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications* **8**:13804. DOI: https://doi.org/10.1038/ncomms13804

Vogel EK, Woodman GF, Luck SJ. 2006. The time course of consolidation in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance* **32**:1436–1451. DOI: https://doi.org/10.1037/0096-1523.32.6.1436, PMID: 17154783

Williams M, Hong SW, Kang MS, Carlisle NB, Woodman GF. 2013. The benefit of forgetting. *Psychonomic Bulletin & Review* **20**:348–355. DOI: https://doi.org/10.3758/s13423-012-0354-3

Wimmer K, Nykamp DQ, Constantinidis C, Compte A. 2014. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience* **17**:431–439. DOI: https://doi.org/10.1038/nn.3645, PMID: 24487232

**Wolff MJ**, Jochim J, Akyürek EG, Buschman TJ, Stokes MG. 2020. Drifting codes within a stable coding scheme for working memory. *PLOS Biology* **18**:e3000625. DOI: https://doi.org/10.1371/journal.pbio.3000625, PMID: 32119658

Xu Y. 2017. Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences* **21**:794–815. DOI: https://doi.org/10.1016/j.tics.2017.06.013

Yu Q, Panichello MF, Cai Y, Postle BR, Buschman TJ. 2020. Delay-period activity in frontal, parietal, and occipital cortex tracks noise and biases in visual working memory. *PLOS Biology* **18**:e3000854. DOI: https://doi.org/10.1371/journal.pbio.3000854, PMID: 32898172

Zhang W, Luck SJ. 2009. Sudden death and gradual decay in visual working memory. *Psychological Science* **20**:423–428. DOI: https://doi.org/10.1111/j.1467-9280.2009.02322.x, PMID: 19320861

Zhou J, Benson NC, Kay K, Winawer J. 2019 Predicting neuronal dynamics with a delayed gain control model. *PLOS Computational Biology* **15**:e1007484. DOI: https://doi.org/10.1371/journal.pcbi.1007484

Zimmermann E, Morrone MC, Burr DC. 2013. Spatial position information accumulates steadily over time. *The Journal of Neuroscience* **33**:18396–18401. DOI: https://doi.org/10.1523/JNEUROSCI.1864-13.2013, PMID: 24259564

Zylberberg A, Dehaene S, Mindlin GB, Sigman M. 2009. Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Frontiers in Computational Neuroscience* **3**:4. DOI: https://doi.org/10.3389/neuro.10.004.2009, PMID: 19325713

Appendix 1

Minimizing retinal after-effects

We assessed the method of minimizing retinal afterimages by repeating all measurements, with the exception of not using phase shift of stimuli (*Figure 3A*). We predicted retinal afterimages could serve as an additional source of information, but only for a brief period after stimuli offset. Therefore, here we expected to see better performance for brief delays compared to conditions with phase shift. *Appendix 1—figure 1A* shows recall error increased with both set size and delay. Both of these effects were statistically significant, as well as their interaction (set size: $F_{(2,18)} = 47.3, p < 0.001, \eta^2 = 0.31$; delay time: $F_{(5,45)} = 48.4, p < 0.001, \eta^2 = 0.26$; interaction: $F_{(10,90)} = 21.3, p < 0.001, \eta^2 = 0.14$), reminiscent of findings for data with phase shift.

Next, we focused on the comparison of conditions with and without phase shift of stimuli (*Appendix 1—figure 1B*). We illustrate the difference in performance by subtracting RMSE obtained in the condition without phase shift (*Figure 4B*) from RMSE shown in *Appendix 1—figure 1A*. Negative values indicate better performance in a condition without phase shift. As predicted, the overall pattern of data suggested performance was comparable for 1 item across all delays, and for all set sizes for extreme delays (simultaneous presentation and 1000 ms), indicated by the difference values around 0. We confirmed the difference in recall error for 1 item across all delays did not differ consistently with and without phase shift, as neither phase shift $(F_{(1,9)} = 0.03, p = 0.86, \eta^2 < 0.001, BF_{\text{incl}} = 0.143)$ nor the interaction of phase shift and delay $(F_{(5,45)} = 0.41, p = 0.89, \eta^2 = 0.00, BF_{\text{incl}} = 0.042)$ reached significance. Based on this result, we conducted all remaining analyses using only the remaining two set sizes. We ran separate repeated measures ANOVAs for each delay using phase shift and set size as factors. The pattern of results we observed was clear: performance was comparable with and without phase shift with the simultaneous presentation and 1000 ms delay (phase shift, $F_{(1,9)} \leq 1.08, p \geq 0.33, \eta^2 \leq 0.002, BF_{\text{excl}} \geq 3.62$; interaction, $F_{(2,18)} \leq 0.8, p \geq 0.44, \eta^2 \leq 0.02, BF_{\text{excl}} \geq 3.39$), while for the remaining intermediate delays recall error was consistently lower when phase shift was omitted (phase shift, $F_{(1,9)} \geq 5.8, p \leq 0.039, \eta^2 \geq 0.06$; interaction, $F_{(1,9)} \leq 2.8, p \geq 0.13, \eta^2 \leq 0.001$).
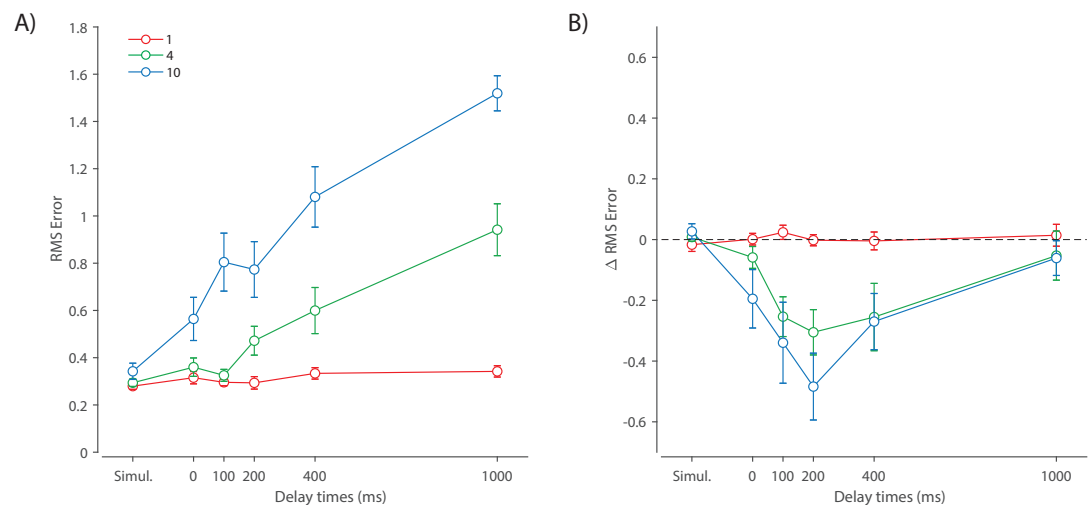

Appendix 1—figure 1. Minimizing retinal after-effects. Cancelling retinal afterimages. (**A**) Experiment 1 RMSE for trials without phase shift. (**B**) Differences in RMSE between trials with and without phase shift across set size and delay conditions. Negative values indicate better performance in the condition without phase shift. Error bars indicate ±1 SEM. N = 10.

Taken together, performance with and without phase shift of stimuli was comparable in perceptual condition (simultaneous presentation) and with the longest delay, suggesting phase shift did not change visibility or encoding of information into VWM. In contrast, we found strong evidence that observers had access to an additional source of information over intermediate delays when phase shift was not used, demonstrated by a better recall performance from 0 ms to 400 ms delay. Specifically, this source of information was available immediately after stimuli offset and was

short-lived, consistent with the theoretical description of retinal afterimages (*Tsuchiya and Koch, 2005*).
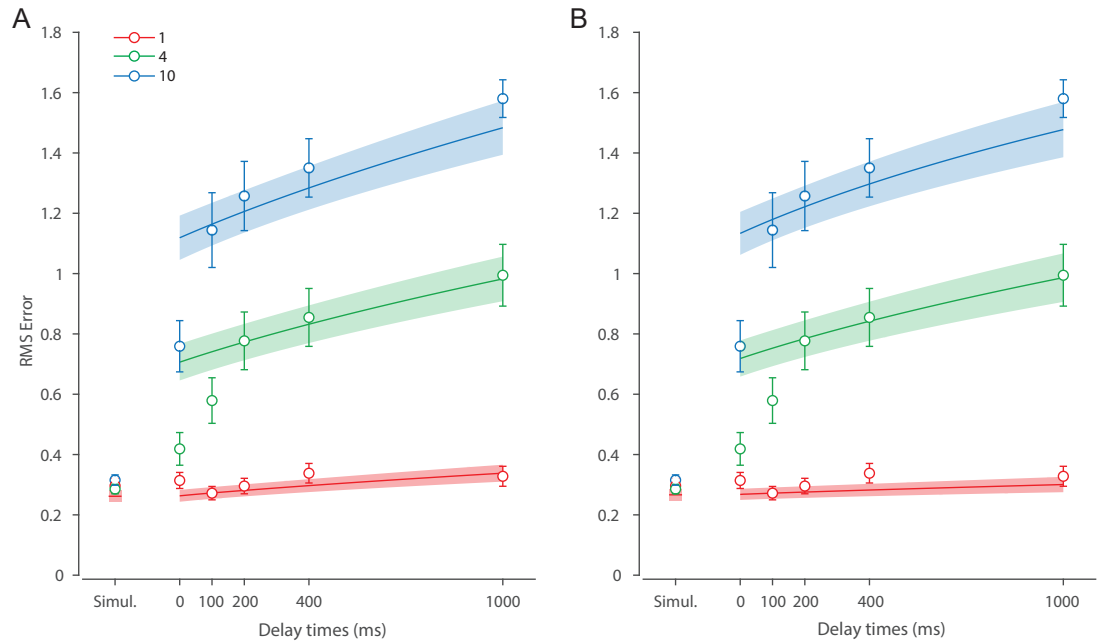
Appendix 2

Swap error estimates
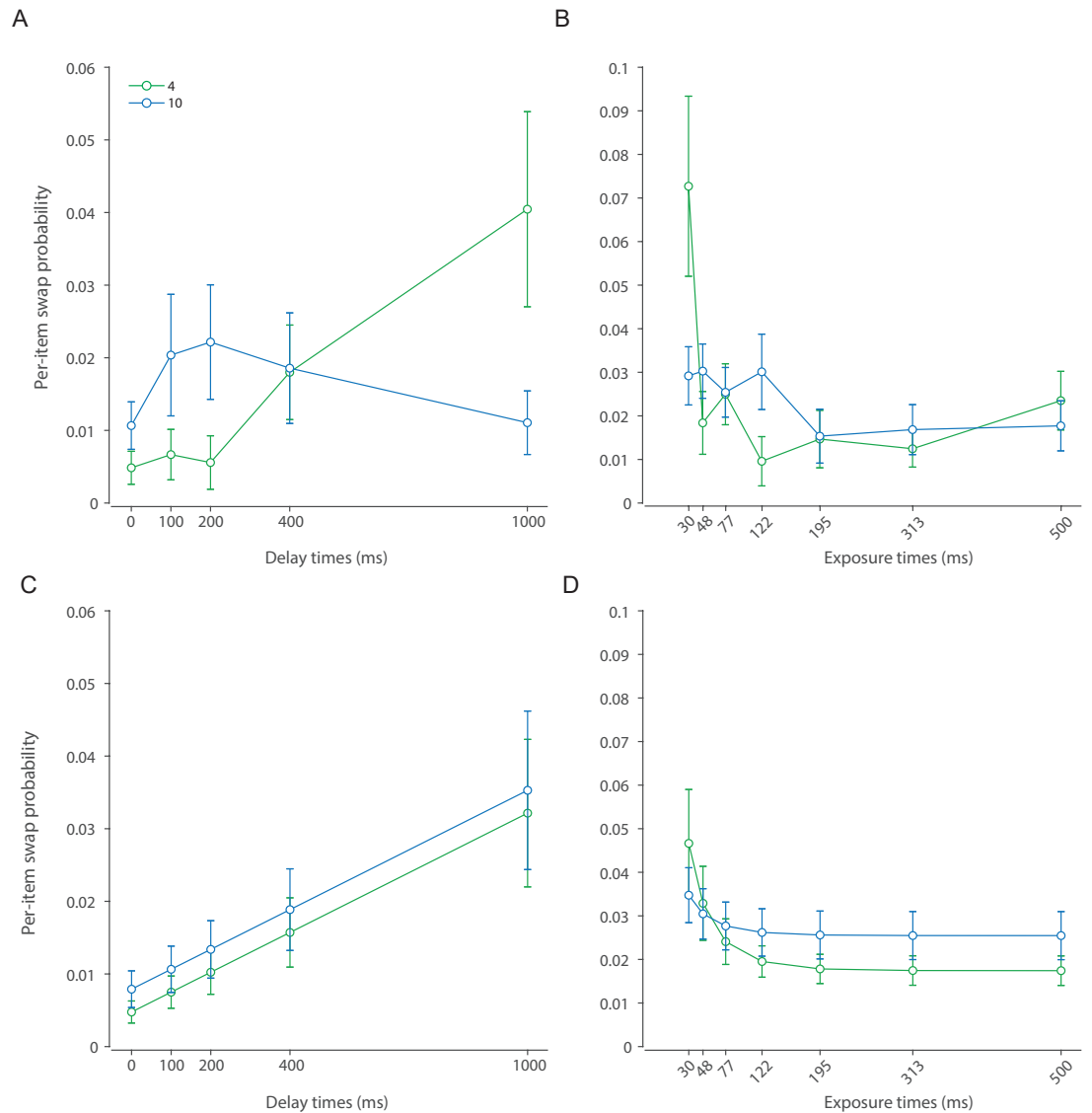
Appendix 2—figure 1. Swap error estimates. (**A and B**) Probability of swap errors estimated from empirical data using the three-component mixture model (*Bays et al., 2009*) in Experiment 1 (**A**) and Experiment 2 (**B**). (**C and D**) Probability of swap errors in best-fitting dynamic neural resource (DyNR) model in Experiment 1 (N = 10) (**C**) and Experiment 2 (N = 13) (**D**). Error bars indicate ±1 SEM.

Appendix 3

Alternative models' fits

Appendix 3—figure 1. Experiment 1 behavioral data and model fit for the dynamic neural resource (DyNR) model without sensory persistence after stimulus offset. (**A**) A version of the DyNR model with equal diffusion across set sizes. (**B**) A version of the DyNR model with diffusion that scales with set size. Error bars and patches indicate ±1 SEM. N = 10.
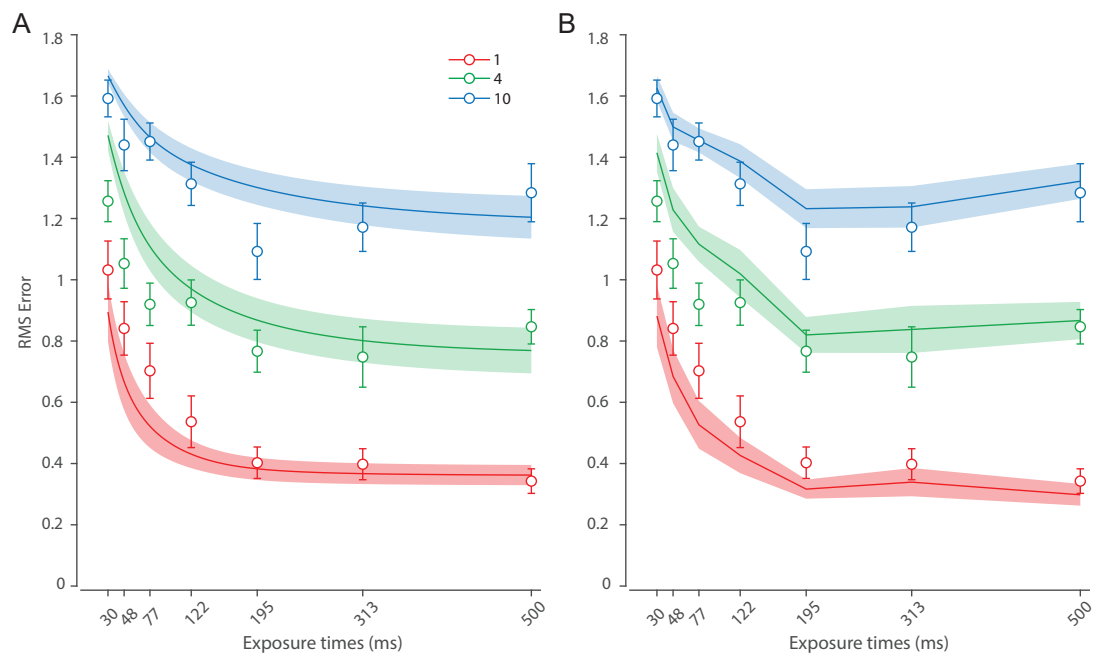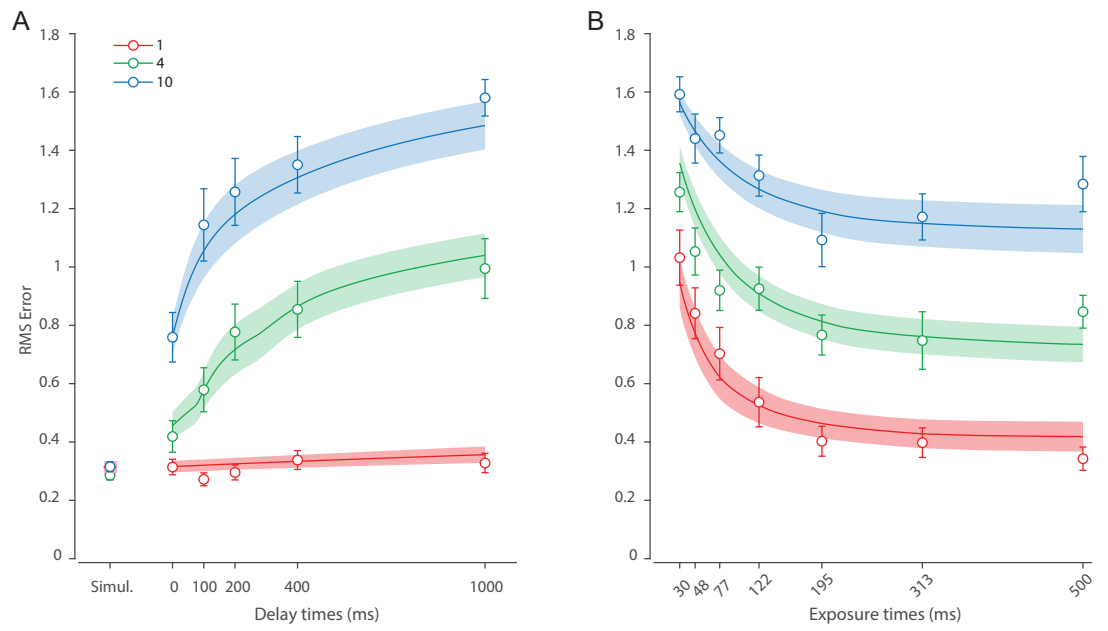

Appendix 3—figure 2. Experiment 2 behavioral data and model fit for the neural model without sensory persistence after stimulus offset. (**A**) A version of the dynamic neural resource (DyNR) model without sensory persistence. (**B**) Separate fits of the simplified neural model to each exposure time. Error bars and patches indicate ±1 SEM. N = 13.
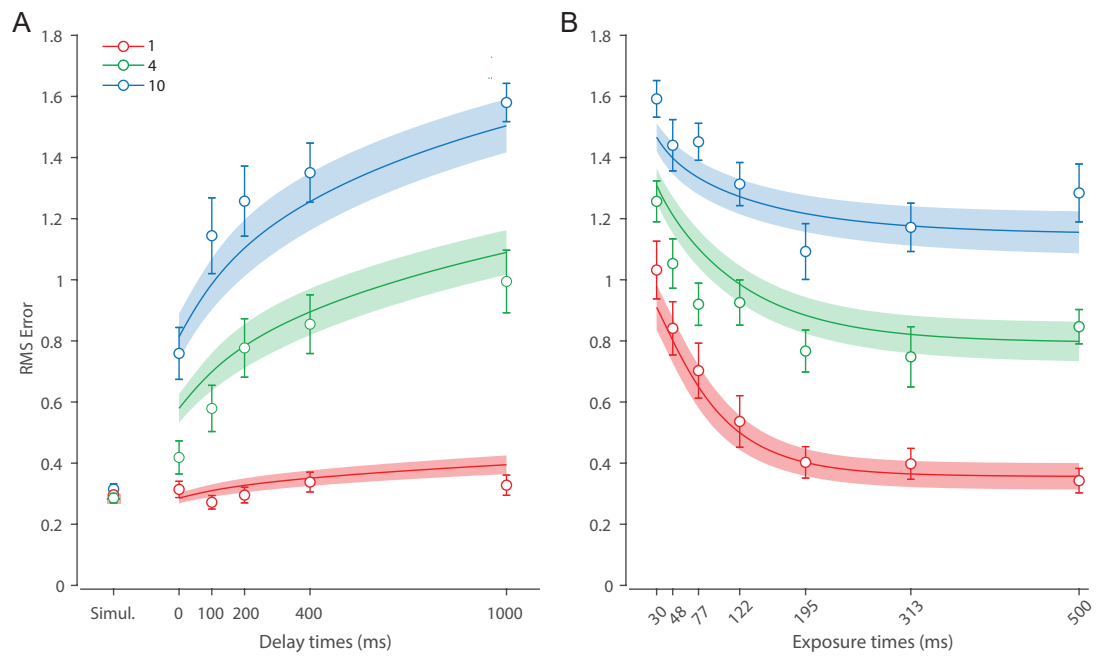
Appendix 3—figure 3. Behavioral data and model fit for a neural model with the direct read-out of information from sensory memory for (**A**) Experiment 1 (N = 10) and (**B**) Experiment 2 (N = 13). Error bars and patches indicate ±1 SEM.
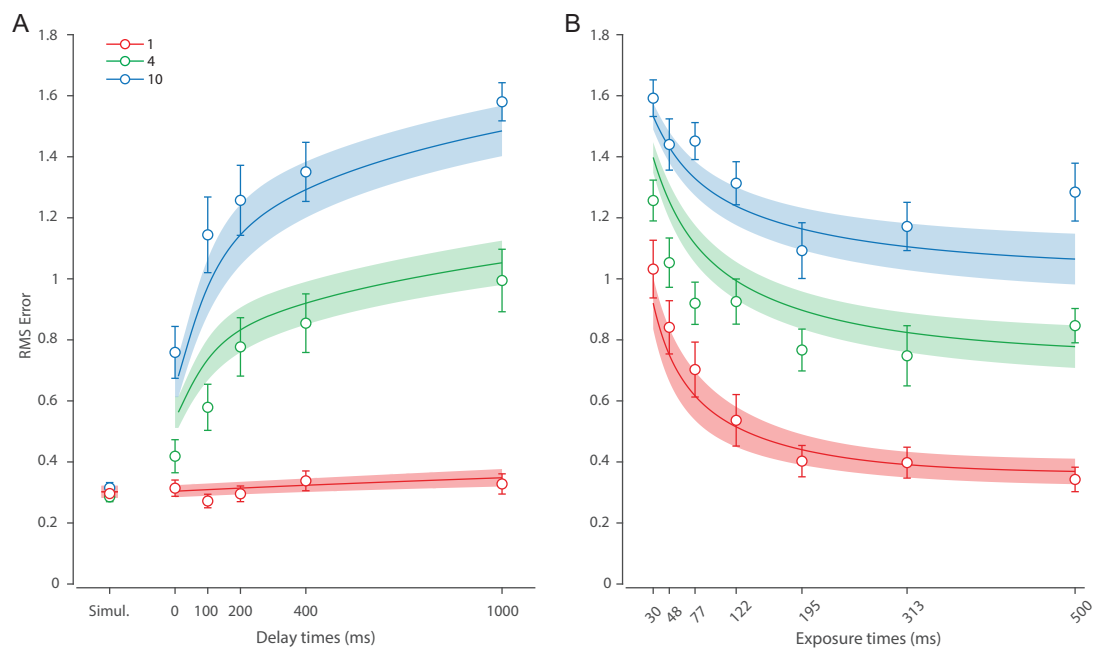
Appendix 3—figure 4. Behavioral data and model fit for the dynamic neural resource (DyNR) model without the cue processing time for (**A**) Experiment 1 (N = 10) and (**B**) Experiment 2 (N = 13). Error bars and patches indicate ±1 SEM.

Appendix 3—figure 5. Behavioral data and model fit for a neural model with constant accumulation of information into working memory (WM) for (**A**) Experiment 1 (N = 10) and (**B**) Experiment 2 (N = 13). Error bars and patches indicate ±1 SEM.
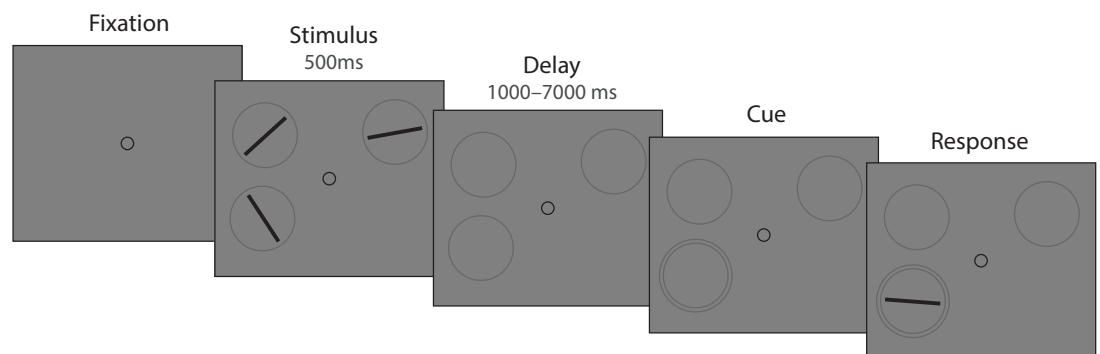
Appendix 4

Additional dataset 1

To further investigate the role of diffusion in memory dynamics, we analyzed an additional dataset collected in our lab (*Tomić et al., 2024*). In this experiment we varied the set size and delay duration similar to Experiment 1. In contrast to Experiment 1, we used longer memory delays, which allowed us to examine the diffusion mechanism on a more suitable time scale. Moreover, memory delays used in this study are out of reach of the decaying sensory information, enabling us to investigate the diffusion without changes in the neural signal strength post-cue.

Methods

Ten observers (six females, four males, aged 18–34) took part in this experiment. The data for this experiment was collected using the same equipment and the testing setting as described for the main experiments. A typical trial sequence is illustrated in *Appendix 4—figure 1*. Each trial began with the presentation of a central annulus which served as a fixation point. Once a stable fixation was achieved, the inner annulus radius changed indicating that stimuli would appear in 500 ms. The memory sample array was then presented for a duration of 500 ms. The array consisted of one or three randomly oriented black bars (length 2.8°). Each bar was positioned in one of six predetermined locations equally distributed around the circle with a radius of 5° around center of the screen. Each bar was presented along with a placeholder circle (radius 1.5°).

Appendix 4—figure 1. Experimental procedure. Stimuli are not drawn to scale.

Memory array presentation was followed by a memory delay during which fixation circle and placeholders stayed visible. The retention interval was either 1 s or 7 s long. After that, one stimulus was randomly cued for recall. The cue consisted of a second, larger circle drawn around one of the placeholders. Observers were instructed to start rotating a response dial (Griffin Technology PowerMate USB) once they were ready to respond. After the rotation of the response dial was detected, a randomly oriented black bar was displayed within the placeholder. Observers were instructed to rotate the dial until the displayed bar matched the remembered orientation of the cued item. Observers confirmed their response by pressing the dial. Trials with different set sizes and delay durations were randomly interleaved.

Eye movements were monitored from the beginning of the trial until stimuli offset, and observers were required to hold steady fixation during that period. If the gaze position deviated by more than 2° a message appeared on the screen and the trial was aborted and restarted with new orientations. Each observer completed 700 trials, divided into two sessions and each consisting of seven equal blocks. Two sessions were separated by at least 1 day, and each lasted approximately 1 hr. At the beginning of each session observers familiarized themselves with the task and experimental setup by doing at most 50 practice trials.

Results

Behavioral data

Recall performance is shown in *Appendix 4—figure 2*. As predicted, response error increased with set size and memory delay. A repeated measures ANOVA revealed a significant effect of set size ($F_{(1,9)} = 111.17, p < 0.001, \eta^2 = 0.76$) and memory interval ($F_{(1,9)} = 58.14, p < 0.001, \eta^2 = 0.12$), and their

interaction ($F_{(1,9)} = 10.66, p = 0.01, \eta^2 = 0.02$) on response error. Moreover, conducting paired t-tests within each set size revealed recall error increased with the delay with set size 1 ($t_{(9)} = 5.83, p < .001, d = 1.84$) and set size 3 ($t_{(9)} = 5.78, p < 0.001, d = 1.83$). The interaction effect was a consequence of a larger increase in error with delay for set size 3 compared to set size 1 ($\Delta \mathrm{RMSE} = \mathrm{RMSE}_{7000ms} - \mathrm{RMSE}_{1000ms}$). These results are consistent with Experiment 1, corroborating our finding that increasing the set size and delay time have a disadvantageous effect on memory fidelity.
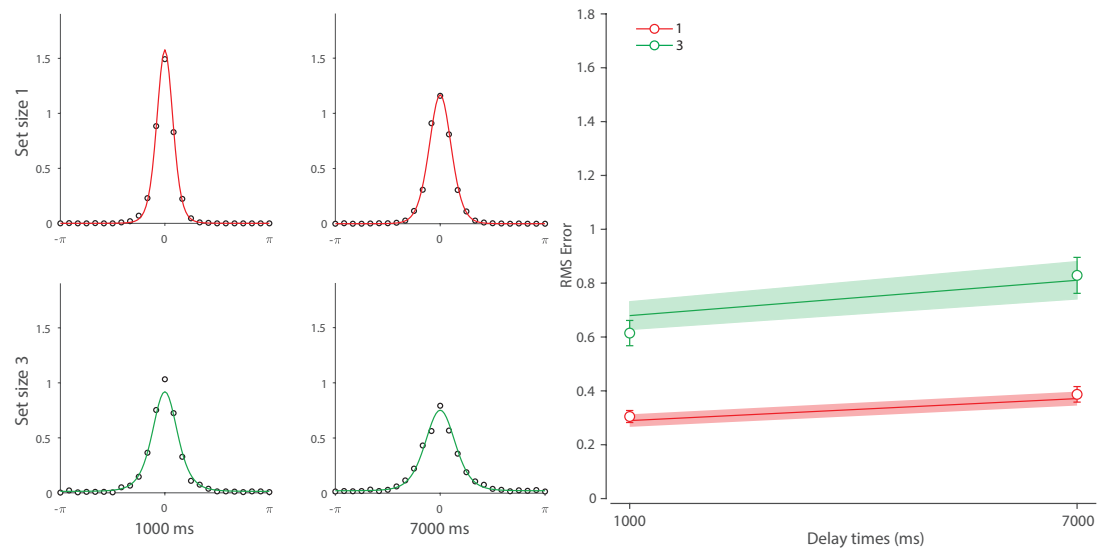
Appendix 4—figure 2. Behavioral data and model fit for Experiment 1a. Error bars and patches indicate ±1 SEM. N = 10.

Neural model

We fitted the DyNR model to the data to test whether noise-driven diffusion is sufficient to account for changes in recall fidelity with longer memory intervals. We applied a simplified version of the model without sensory decay and VWM accumulation components. This was justified given that estimate of sensory decay from Experiment 1 was shorter (mean life $\tau = 0.21$) than the shortest interval used in this experiment (1 s). Moreover, based on our findings in Experiment 2, we argue that a display duration of 500 ms is sufficient to fully encode objects into VWM.

Curves in *Appendix 4—figure 2* show fits of the model with ML parameters (mean ± SE: population gain $\gamma = 385.02 \pm 208.3$, tuning width $\kappa = 2.67 \pm 0.43$, cue processing constant $b = 0.68 \pm 0.67$, base diffusion $\sigma^2_{\mathrm{diff}} = 0.009 \pm 0.001$, swap probability p = 0.005 ± 0.002). The model provided an excellent quantitative fit to response distributions and summary statistics (*Appendix 4—figure 2*), successfully explaining the adverse effects of set size and memory interval on recall fidelity. Critically, and consistent with results from Experiment 1, the proposed DyNR model provided a better fit to human response error compared to the matching model without diffusion (ΔAIC = 144.75) or the model in which diffusion rate increases with set size (ΔAIC = 42.3). In conclusion, this result shows that variability in representations over longer memory intervals can be fully accounted for by noise-driven accumulation without changes in the representational signal (*Schneegans and Bays, 2018*; *Panichello et al., 2019*; *Wolff et al., 2020*).

Appendix 5

Additional dataset 2

To further validate predictions of the DyNR model we fitted it to an existing WM study (Experiment 1 in *Bays et al., 2011*). This study focused on the role of temporal dynamics during WM encoding, thereby addressing the same question as our Experiment 2. In contrast to our Experiment 2, *Bays et al., 2011*, used a longer delay period (1100 ms), precluding the strengthening influence of decaying sensory information on recall. This dataset therefore isolates the initial information accumulation process during stimuli presentation.

Methods

The observers (*N*=32) performed a continuous report task in which a variable number of oriented bars was presented for a variable duration, followed by a pattern mask (100 ms) and a 1 s delay period after which one of the items was probed for recall. Set size was manipulated between observers and exposure duration was manipulated within observers. Each observer performed 100 trials per exposure duration, for a total of 25,600 trials in the study. A more detailed description of the experiment is provided in *Bays et al., 2011*.

Analysis

Considering only exposure duration in this experiment was manipulated at the observer level, we decided to expand our modeling approach by employing a Bayesian hierarchical method as a compromise between fitting the data for each observer (i.e. set size) independently and pooling the data across all observers. Using a Bayesian hierarchical modeling, individual-observer parameters are considered samples from population distributions, whose means and variances are estimated based on all available data. In general, this approach has a desirable characteristic of constraining individual-level parameters with the population-level distribution and producing meaningful parameter estimates when a model is fitted across separate groups. The dynamic neural model fitted to the data is identical to the model fitted in Experiment 2, with the exception that here we assumed any existing post-stimulus sensory activity completely diminished by the time of the cue (1100 ms post-stimulus offset), and therefore we did not model sensory decay here. To obtain the hierarchical fit, we used the differential evolution Markov chain algorithm (*Braak, 2006*). All individual-level parameters were samples drawn from normal (i.e. Gaussian) distributions, with corresponding mean and standard deviation being constrained by uniform hyperprior distributions. We collected 240,000 post-warmup samples across 12 chains and computed median and 95% equal-tailed intervals (ETI) of posterior distributions to obtain the group and individual-level parameter estimates. Prior specifications and empirical data for all analyses can be found along with the published code.

Results

Appendix 5—figure 1 and *Appendix 5—figure 2* show empirical distributions and summary statistics across all conditions. Similar to Experiment 2, increasing the exposure duration ($F_{(7,196)} = 110.9, p < 0.001, \eta^2 = 0.188$) and decreasing the set size ($F_{(3,28)} = 22.83, p < 0.001, \eta^2 = 0.53$) had beneficial effect on response error. Interaction of exposure duration and set size was significant ($F_{(21,196)} = 3.13, p < 0.001, \eta^2 = 0.02$). Critically, the pattern of memory fidelity dynamics largely matches the pattern observed in Experiment 2, with response errors decreasing rapidly as presentation duration was increased from the minimum duration, saturating at longer durations. This pattern was consistent across all set sizes, which only differed in the absolute error.

These dynamics were accurately predicted by the DyNR model, both at the level of response distributions (curves in *Appendix 5—figure 1*) and summary statistics (curves in *Appendix 5—figure 2*). The parameters used to generate model predictions were obtained by taking the individual observer's posterior medians. We observed the following hyperparameters (median and 95% ETI of hyperposterior): population gain γ = 109.47 (88.1–133.57), tuning width κ = 3.23 (2.6–4.03), sensory rise time constant τ_{rise} = 0.0049 (0.0019–0.0091), VWM accumulation time constant τ_{WM} = 0.067 ±(0.051–0.087), cue processing constant b = 0.423 (0.093–0.8436), base diffusion σ^2_{diff} = 0.095 (0.057–0.149), spatial uncertainty time constant τ_{spatial} = 0.031 (0.022–0.041), swap probability p = 0.02 (0.011–0.034).
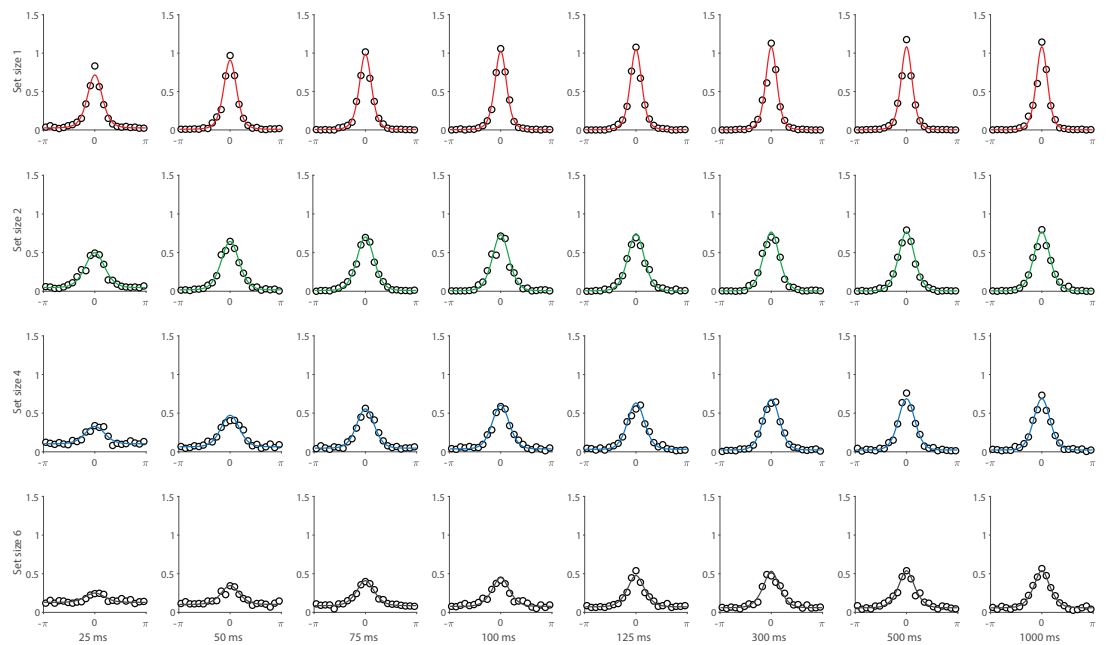
Appendix 5—figure 1. Empirical recall error distributions (black circles) from Experiment 1 in *Bays et al., 2011*, and the dynamic neural resource (DyNR) model fits to the data (colored curves).
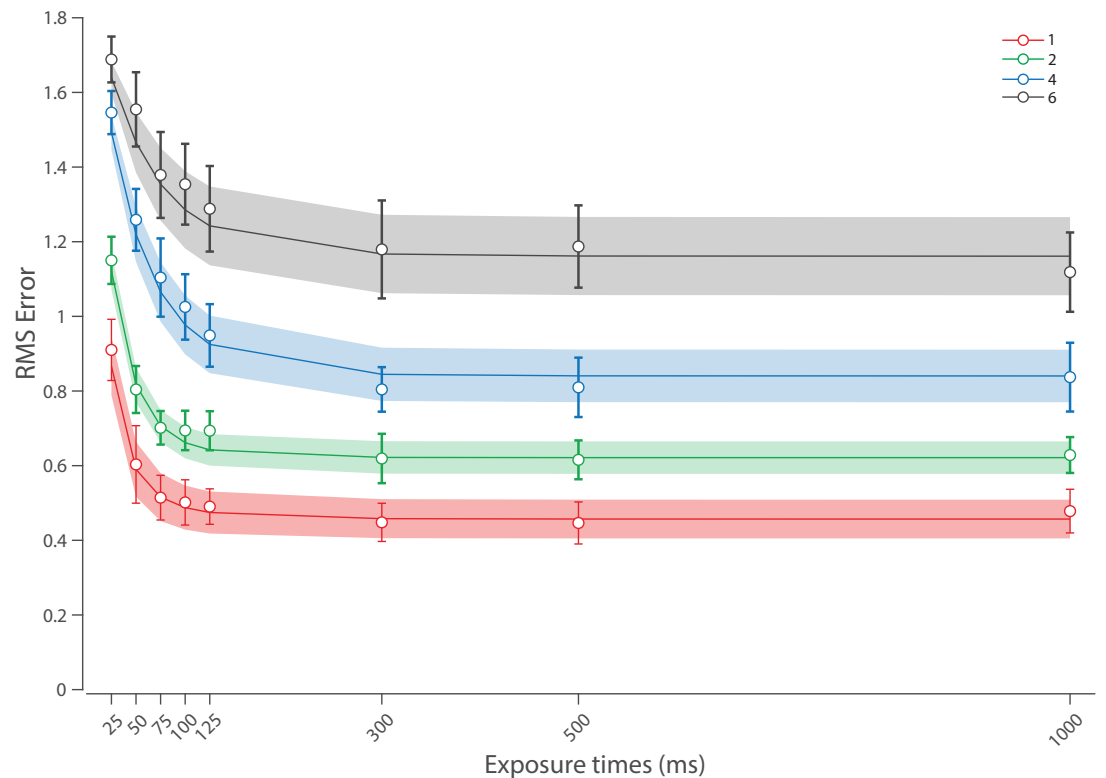
Appendix 5—figure 2. Summary statistics (black circles) from Experiment 1 in *Bays et al., 2011* and the dynamic neural resource (DyNR) model fits to the data (colored curves). The DyNR model was fit to the distributions of recall errors shown in *Appendix 5—figure 1*. Error bars and patches indicate ±1 SEM. N = 32.