Representation and computation in visual working memory

Paul M Bays[1], **Sebastian Schneegans**[1], **Wei Ji Ma**[2], **and Timothy F Brady**[3,*]

[1]University of Cambridge, Department of Psychology, Downing St, Cambridge CB2 3EB, U.K.
[2]New York University, Center for Neural Science and Department of Psychology, New York, USA
[3]University of California San Diego, Department of Psychology, La Jolla, CA, USA
[*]Corresponding author: tfbrady@ucsd.edu

ABSTRACT

The ability to sustain internal representations of the sensory environment beyond immediate perception is a fundamental requirement of cognitive processing. In recent years, debates regarding the capacity and fidelity of the working memory (WM) system have advanced our understanding of the nature of these representations. In particular, there is growing recognition that WM representations are not merely imperfect copies of a perceived object or event. New experimental tools have revealed that observers possess richer information about the uncertainty in their memories and take advantage of environmental regularities to use limited memory resources optimally. Meanwhile, computational models of visuospatial WM formulated at different levels of implementation have converged on common principles relating capacity to variability and uncertainty. Here we review recent research on human WM from a computational perspective, including the neural mechanisms that support it.

Introduction

Since the dawn of perception research, theoretical frameworks have been built around the notions of representation and computation[1]. A key aspect of internal representations is that they are noisy: they vary even upon repeated presentations of the same physical stimulus. A key aspect of computation is inference: because the brain has no direct access to stimulus properties, it has to build beliefs about them based on the available representations[2]. In perception research, great progress in understanding representation and computation has been made by combining experiments with mathematical process models, which specify precisely how information is received and processed, leading up to a decision. Such models allow the researcher to disentangle representation and computation and to compare theories for each stage.

While this agenda has been pursued for over 150 years in perception research, it has only recently become widespread in the field of visual WM. This field initially[3] used rather simplistic notions of representation and overlooked computation altogether. The dominant notion was that visual WM "holds" internal copies of visual objects or features, which can be directly accessed for judgment or decision making at a later point in time. In the past 20 years, the shortcomings of this metaphor have become clear, in part driven by the "slots-versus-resources" debate (see Box 1). The general conception emerging from this debate is that a combination of visual processing and attention to objects induces a high-dimensional memory state (e.g. a pattern of neural activity) that is informative about the objects' features and can be sustained once they are no longer available to the senses. In this framework, recall can be understood as probabilistic inference, based on the memory state, of the past features and their relationships. This process is illustrated in Fig. 1

for the elementary experimental task of reproducing from memory a colour stimulus, corresponding to a specific point in a space of hues (Fig. 1, left). Due to a combination of factors – including internal noise, limited neural signal, interactions with other stimuli in memory and dynamics during the delay period – the same stimulus can result in many different memory states at the time of the memory test (Fig. 1, middle).

Unlike the stimulus itself, the information that a particular memory state provides about the stimulus cannot in general be captured by a single point in the parameter space. Instead, it is fully described by a likelihood function (Fig. 1, right), which can be interpreted as showing the degree to which the obtained memory state is compatible with different hypothesized stimulus inputs. If the observer is instructed to choose a best estimate of the previously presented hue, they might choose the peak of the likelihood (the "maximum-likelihood estimate") and the experimenter might record the observer's error as the distance between this estimate and the presented hue. The distribution of recall errors over many trials, and in particular the changes in distribution observed when multiple items are held in memory simultaneously, have provided important evidence for discriminating between models of WM (see Models section below). However, unlike the error distribution, a full likelihood function exists on each single trial. For different memory states, the likelihood function could be relatively narrow (compatible with only a small range of possible inputs, top right) or broad (providing little or no information to discriminate between inputs, bottom right). Memory uncertainty can be quantified as the width (e.g. standard deviation) of the likelihood function, but even this description is incomplete, for example when the likelihood is asymmetric (centre right) or multimodal.

Access to memory uncertainty

Just because the memory state provides this richer information does not mean the brain makes use of it or the observer has conscious access to it. In research on human perception, the question of whether perceptual decisions take into account uncertainty is a classic one. The literature on Bayesian integration and Bayesian cue combination[4] has demonstrated convincingly that the mind takes into account uncertainty on a trial-by-trial basis when weighing evidence. In the realm of WM, recent experimental methods have begun to probe in detail the information observers can extract from their memory state (Fig. 2). The familiar sense that we are more certain about some memories than others is experimentally validated by studies that ask observers to report their confidence alongside a point estimate (Fig 2A). As the number of items to remember increases, error becomes more broadly distributed and average reported confidence declines (Fig 2B). Confidence ratings also vary across trials with a fixed set size, and the error distribution is narrower for trials with higher confidence ratings (Fig. 2C;[5]), revealing access to latent information about uncertainty.

Other studies have tried to quantify uncertainty in the stimulus dimension itself rather than using a confidence judgment. Instead of asking subjects for a confidence rating, observers may be instructed to make a secondary, uncertainty-based decision[6–8] (Fig. 2D). For example, the observer could first recall the stimulus, then set an interval around the recalled value, intended to "capture" the true value. Points are awarded for a successful capture, but fewer points when the interval is larger. Thus, a point-maximizing observer would set a larger interval when uncertainty is high and a smaller interval when uncertainty is low. This technique reveals a strong relationship between interval size and error magnitude (Fig. 2E;[6–8]), consistent with the studies that use confidence ratings. Moreover, in parallel to perceptual studies[9], observers combine their memory-based likelihood with prior information about a feature, even if that information varies from trial to trial[7].

Change detection tasks, even when not paired with a confidence report, can serve to establish whether uncertainty is taken into account *implicitly* in WM-based decisions[10,11]. This property makes change detection useful for studying the WM representation of uncertainty in non-human animals[12]. A large difference between the memory representation of the study and the probe provides less evidence for a true change if uncertainty is higher (Fig. 2F). In these studies, variations in uncertainty not only arose spontaneously, but were also experimentally induced by varying the reliability of the stimulus information from trial to trial and from item to item. The studies used formal model comparison to conclude that observers take into account WM uncertainty in their decision.

Taken together, the evidence that uncertainty is maintained in WM, and that uncertainty can be estimated continuously – not just whether the memory is present or absent – is strong at this point. Fundamentally, this means that WM is much richer than previously believed. An open question in perception is whether observers use full probability distributions or only summary statistics such as the width of the distribution[13–15]. WM researchers have started to study the analogous question[8], with initial evidence suggesting use of the likelihood function extends beyond its width.

Models that implement WM uncertainty

Despite variation between models of WM in their levels of implementation and their descriptive language, recent years have seen a notable convergence on a common set of principles required to capture behavioural performance on reproduction tasks. Crucially, the modern models of visual WM described in this section all imply a richer underlying stimulus representation that carries information about memory uncertainty. Other recent models[16–18] have made important advances in understanding how conjunctions of features are stored, and these are reviewed in a separate section (Feature binding) below.

Population coding accounts[19], inspired by similar models of attention, sensory integration and decision-making[20–22], describe WM in terms of encoding and decoding of stimulus information from the noisy activity of large populations of neurons tuned to different features (Fig. 3A). Variability arises in this model as a consequence of the probabilistic generation of spikes. Resource limitations are identified with the allocation of a limited quantity of neural signal or gain between neurons responding to different items. This constraint explains why recall fidelity declines with the number of items held simultaneously in memory, and also accounts for effects of stimulus salience and behavioural priority on recall.

Access to uncertainty in this model is automatic, in the sense that a decoder with knowledge of the population tuning functions can reconstruct a full likelihood function[23,24]. Li and colleagues[25] combined the theory of probabilistic population coding[26,27] with a generative model for fMRI activity[22,28] to decode uncertainty along with the behavioral estimate from the pattern of voxel responses on each trial. The decoded uncertainty correlated with a behavioral read-out of certainty or confidence.

Under specific simplifying assumptions, the decoding of stochastically generated spikes in a neural population response can be viewed as equivalent to averaging of noisy samples of a stimulus feature (Fig. 3B;[29]). This provides a connection to cognitive models that describe resource allocation as distributing a limited (but in some cases arbitrarily large) number of discrete samples between memory items[30–32], a concept that was originally proposed to model selective attention[33] and that was later successfully applied to multiple-object tracking[34,35]. With a fixed, small number of samples, this account has been presented as a variant of the classic slot model (*slots-plus-averaging*;[31]). However, fits to continuous recall data are improved when the number of samples varies randomly and independently between items[29], in analogy to stochastic spiking. Samples are discrete in this account, but moment-to-moment variability in the number

of samples fits less well with the concept of slots, and the resource that is shared among items (the mean number of samples) is a continuous variable.

As an alternative perspective, the TCC model[36] describes WM decisions as based on a noisy familiarity signal whose mean is highest for the shown color (or other stimulus) and lower for stimuli that are less similar to the shown item (Fig. 3C). This model makes an explicit connection to signal detection concepts commonly used in long-term memory measurement, associating WM performance with the discriminability (d') between maximally distant stimuli and confidence with the peak familiarity amplitude. The familiarity function in the TCC model is closely related to the tuning in population coding models, which in turn have a geometric representation in terms of how distinct the representations associated with different stimuli are from each other[37]; A proposed relationship between the familiarity function in the TCC model and empirical measures of psychological similarity is disputed[36,38].

The mathematics of averaging dictate that the dispersion of errors under sampling and population coding models varies with the number of samples or spikes (Fig. 3E), such that their estimates can be succinctly described in terms of particular distributions over precision. Abstracted from a specific implementation, variable-precision models[6,11,39–42] identify WM resource with mean precision, and draw individual precision values from a distribution (Fig. 3D), the key characteristic of which may be a variance that scales with the mean[29].

As noted above, all of these models contain information about uncertainty, not just error. In addition to capturing the changes in error distribution induced by set size (as illustrated in Fig. 2B), both population coding[23,29] and variable-precision models[24] have been shown to account quantitatively for the results of conditioning on confidence in continuous reproduction tasks, shown in Fig. 2C. The relationship between certainty and error in these models predicts that the long-tailed distributions of error commonly observed in WM recall can be decomposed on the basis of subjective certainty into individual distributions that differ in precision. These models also predict the distribution of confidence ratings in continuous report (Fig. 3F), account for performance changes with confidence in change detection tasks[43] and quantitatively reproduce error distributions on whole-report tasks ([44]; Fig. 3G) on the basis that participants choose items to report in decreasing order of confidence[29].

A lesson emerging from these noise-based accounts of WM has been that computation during the retrieval stage is interesting in its own right and requires a non-trivial modeling step. Except in the very simplest tasks, retrieval is not a passive, straightforward recall of features of memorized stimuli. Even in a delayed estimation task with more than one item, computations must be performed to determine which item in memory is indicated by the cue (see Feature binding section below). In other tasks, memory-based likelihood functions associated with individual features need to be combined with a prior (see Introduction), or transformed into a decision about a categorical global variable such as presence of a target[40] or of a change[10–12,41,45]. For example in change detection, if memories are noisy, then *every* item changes in terms of its internal representation, creating a hard decision problem (see Fig. 2F). The brain might make such retrieval-stage decisions in a Bayesian way, that is, by inverting a generative model while minimizing a cost function. Indeed, Bayesian observer models augmented with a resource limitation in the encoding stage have proven successful in capturing WM-based decisions in quantitative detail[7,10–12,40]. The computations during WM retrieval have also been addressed in neural process models[46,47] discussed in later sections.

Resource allocation, rationality and incentives

Most modern models of visual WM allow for flexibility in how resources are allocated. This flexibility is necessary to account for a range of findings in which observers prioritize certain memoranda over others, as a result of differences in their attentional salience or relevance to behavioural goals[48–50]. Control over resource allocation is also critical to many of the sensorimotor functions ascribed to visual WM (Box 2). The assumption that resources are allocated optimally to minimize expected error across trials has been used to quantitatively reproduce the observation that the average precision of an item's representation increases with the probability that the item will be probed for recall[6,19] .

Manipulations of incentives have also been successful in modulating allocation. In a multiple-item delayed-discrimination task of spatial location, items that were marked with a pre-cue as yielding higher reward were remembered better[51]. While in that study, attentional priority and reward coincided, in another study, reward improved performance even when these cues were dissociated[52]. Finally, reward-associated items are remembered better even when task-irrelevant[53].

These results are compatible with a structural constraint on the representational capacity of the WM system. Divisive normalization[54] has been identified as a possible neural basis for such a constraint, whereby inhibition between pools of neurons representing different stimuli in memory implements a limit on combined activity amplitude. Population coding models can account for both effects of set size and flexibility in resource allocation based on this principle[19].

An alternative perspective is based on the theory of resource rationality[55], which proposes that the brain maximizes task performance while at the same time minimizing a biologically relevant cost, such as the cost of neural spiking. Assuming a cost that is linear in encoding precision, this idea can account for effects of set size and probe probability on precision in delayed estimation[56]. In this view, a decrease of precision with set size is not a signature of a structural limitation of WM, but the outcome of a rational cost-benefit analysis – is greater precision "worth" the associated cost?

The resource-rational account can be tested by manipulating the incentives for a task. An increased reward should shift the balance towards higher performance by compensating for the higher associated cost. Delayed estimation performance did not significantly improve when monetary reward was higher[57], nor when the total attainable reward was raised by increasing cue validity[52]. In a change detection experiment, subjects who were asked to try to remember all items performed better than those who were asked to just do their best[58]. However, in another study, "gamification" of a working memory task increased motivation but did not improve recall performance[59].

Taken together, it seems that resource allocation in WM is responsive to reward differences between items or locations, while evidence for effects at the condition or task level is very limited. This might point to different underlying mechanisms: responsivity to inter-item differences might rely on neural circuits dedicated to prioritization, whereas responsivity to overall reward might rely on motivation. Alternatively, it is possible that the differences in reward were too small to elicit an effect.

WM limitations have also been recognized as being an important factor in reward-based instrumental learning[60]. In a task in which subjects had to learn, based on feedback, which of three responses was associated with each of N stimuli, with one stimulus being presented at a time, a pure reinforcement learning model failed to capture the effects of N and delay. A reinforcement learning model augmented with a WM mechanism, consisting of a slot-like limited capacity and forgetting, was able to account for the data[61,62]. Further work should test alternative, resource-based models of WM within this task.

WM in a structured environment

The information we need to hold in WM in real world situations is generally statistically structured and predictable. That is, unlike in typical WM experiments where stimuli tend to be randomly generated and unrelated to each other, when we remember information in a real scene, we have prior knowledge that can help constrain our memories. Knowing we saw a stove on the left of our view is informative about the object that was likely on the right (it is more likely to be a blender than a mailbox[63]; and knowing the object was on a kitchen counter and approximately banana-shaped provides a strong hint it may have been yellow. Thus, a critical aspect of understanding how we use WM in the natural world is understanding how our WM system uses our prior knowledge about what is present and what objects and features generally co-occur to structure our memory representations.

This problem can be recast as one of communication (Fig. 1): to store information successfully in WM, we need to communicate to our future selves only what is unexpected or unknown about the given object or scene. This view focuses on how we could optimally encode information if we know we will later decode it using the same statistical knowledge of the environment. For example, if our environment and body were entirely static, we wouldn't have to encode any information in WM. If they were entirely unpredictable, we would have to encode everything. In theory, if our brain makes use of the learned regularities about what objects are likely to occur and co-occur, then the stronger our prior expectations in a given situation, the less entropy the stimulus has and the less we need to encode about it, and thus the easier it should be to store in memory.

The formal frameworks used to understand the impact of such knowledge on WM thus have often relied on information theoretic principles like compression[64,65] and rate-distortion theory[66,67], which attempt to formalize the entropy of the stimulus and the communication problem faced by our memory system. Another line of work has formalized benefits from prior knowledge by considering that our memory system may encode information with respect to a generative model of the world that constrains the possible scenes we will see[68–70]. Storing information in memory conditioned on such a model reduces the entropy relative to storing it on its own, and so such models also help to provide frameworks for thinking about how our brain makes use of such prior knowledge. Such models also often suggest we preferentially encode objects that are least consistent with our priors, to enhance how much total information we can remember[70].

While these models focus on conjunctions of features and objects, the influence of environmental statistics, and encoding items with respect to these statistics, may also be responsible for anisotropies in the internal representation of individual visual features such as orientation, colour and location[71,72]. These take the form of 'stimulus-specific' variation in precision within a feature dimension (e.g. cardinal orientations are reproduced with less variability than obliques) and systematic biases in reproduction and comparison of features (e.g. reported orientations are on average biased away from the nearest cardinal). It has been proposed that these anisotropies are an adaptation to the unequal distribution of stimulus features in the environment (e.g. cardinal orientations are more prevalent than obliques in natural scenes). According to one expression of the efficient coding principle, encoding resources are preferentially allocated to more frequently encountered stimuli in order to maximize the information transmitted, with consequences for both discriminability and bias[73–75]. This principle can be naturally incorporated into population coding models of WM (Fig. 3A) via an optimal redistribution of tuning functions[76], providing a quantitative account of stimulus-specific effects in memory and their interactions with set size.

More discrete frameworks that have traditionally dominated WM research have often focused on treating WM limits as a limit on how many independent items can be remembered[3,77]. Such frameworks have

generally formalized the usage of prior knowledge via the concept of chunking[3,78]. The most common conception of chunking in WM is entirely discrete, proposing that we learn co-occurrences and use these to create chunks in long-term memory. The content of WM is then often thought to be entirely replaced by a pointer to this information in long-term memory. For example, you could remember the word "cow" as a single pointer to your long-term conception of cows and then, if asked what the 3rd letter was, reconstruct this by decompressing the chunk into the letters by decoding your long-term memory. In this framework, chunks improve performance by replacing to-be-remembered items with compressed representations, which can be decompressed when required from long-term memory[3,78,79]. A similar principle has been invoked to explain anisotropies in recall of individual features, based on supplementing a detailed and continuous memory representation with a coarse categorical one[80,81]. In an information theoretic framework, chunking can be recast as an approximation to more general compression schemes: that is, chunking can be seen as a way of implementing such compression in models where items are treated like discrete units, but many non-discrete compression mechanisms are also possible[65,82,83].

Qualitatively, these theories all make the same basic prediction: that we should be better at holding in mind information if it more strongly matches our prior knowledge. This seems to hold in a wide variety of situations: people are better at remembering stimuli that match real-world co-occurrence statistics[67] or newly learned co-occurrence statistics[65,84]. And they are better at remembering stimuli that are familiar than perceptually-matched stimuli that are scrambled or otherwise do not connect to their prior knowledge[85–87], and better with realistic objects and configurations of objects compared to simple meaningless stimuli or random configurations of objects[88–91]. This is in line with classic work in verbal memory showing semantic coding in working memory[92].

Theories based on chunking or information theoretic principles like rate distortion or compression propose that we change our initial encoding of stimuli based on environmental regularities. However, better recall of stimuli that match prior experience can also arise in many real-world situations from an informed decoding strategy even if encoding is uninformed. For example, even if someone remembered a scene by just randomly sampling a few objects to remember, they would be best served by making informed decisions when tested on their memory: assuming a stove is present in a kitchen will on average improve memory performance even if the stove was not explicitly encoded, since stoves are nearly always present in kitchens. Many studies testing information theoretic accounts of encoding do explicitly test for the coarsest versions of such strategies (for example, Brady and colleagues[65] showed people do not report a priori likely items more often when they are not present), but making precise statements about how much of the benefit of environmental regularities arises at encoding vs. decoding is often impossible. Indeed, the exact predictions for how encoding should vary as a function of environmental regularities will vary with details of the optimization, including the loss function that describes the relative undesirability of different errors[93]. There are also limits to encoding flexibility[94,95], in terms of what adaptation of encoding strategy is possible and how rapidly it can be achieved in response to new information about environmental statistics.

From features to objects

A long-standing question about WM is whether its basic unit is a feature or an object. This question can have different meanings, all of which have recently been recast in the modern noise/resource view of WM. One meaning is whether or not different feature dimensions within an object share the same resource. Using a change localization task and formal comparison of noisy-memory models with an optimal decision stage, it was found that orientation and colour have independent pools of resource[96], broadly consistent

with previous results from delayed estimation[97,98]. Other findings, however, suggest that resource pools are not completely independent. Retrospective cues indicating the feature dimension to be tested in a continuous report task have been found to impact performance, suggesting that resources can to some degree be shifted across feature dimensions[99–101]. Moreover, a modest decline in performance when adding more relevant feature dimensions was observed in delayed comparison[102] and change detection tasks[103,104]. However, it is important to note that in a noisy-memory framework, a decline in accuracy in change detection does not necessarily imply reduced resource; instead, the noise added by the additional features could decrease the overall signal-to-noise ratio in the integration of information across items[96].

A second meaning is whether or not an irrelevant feature of a relevant object is automatically represented in WM. Several studies employing surprise tests for previously irrelevant non-spatial features showed either near-chance performance (when using change detection tasks;[105,106]) or very low precision (in delayed reproduction tasks;[107,108]), and decoding from fMRI or EEG data has shown little evidence for maintenance of task-irrelevant features[109,110]. However, the presence of task-irrelevant features in memory items – and even in items merely inspected in a perceptual task – has been found to degrade recall of other items to the same extent as task-relevant features[111]. Based on formal modeling of change localization performance, Shin & Ma[96] suggested that task-irrelevant features of attended objects are automatically encoded, occupying WM resources, but they are subsequently only weakly maintained under the control of top-down processes, causing their representations to rapidly degrade.

A third meaning is whether an object takes up resources for a feature even when it is neutral with respect to that feature (e.g., a circle is neutral for orientation), as long as the object is task-relevant because of other features. In WM tasks in which N colors and N orientations were divided either over N two-feature objects or over $2N$ one-feature objects, some such "leaking away" of resources to neutral features was observed[96,97]. Two further studies indicate that to prevent this, it is sufficient for different features to share the same location, even if they are not fully integrated into a smaller number of objects[112,113].

Theoretical proposals attempting to unify the different aspects of the feature/object question have included that of a hierarchically structured feature bundle[114] and of partially packaged resource[96]. Further progress will require more systematic investigation of different feature pairs, a reconsideration of older studies in light of the concept of noisy memories, and potentially favoring delayed estimation and delayed comparison over change detection and change localization as paradigms (because the latter require more assumptions about the decision stage).

Feature binding

Beyond memorizing individual feature values, for many tasks both in real life and in experiments it is necessary to maintain the correspondence (binding) between multiple features of a single stimulus. Delayed reproduction tasks in particular require participants to recall the binding between cue and report features in order to make an accurate response when presented with the cue. Failure to accurately retrieve the cued target item leads to swap errors, which are reflected in a specific concentration of responses around the report feature values of non-target items[115–117].

Our understanding of this type of error has substantially improved in recent years. The frequency of swap errors depends on the feature (or features) used as a cue[118,119], and they occur most often between a target and a non-target item that are similar in their cue feature[120–124]. This would not be predicted if swap errors arose from a failure of a separate memory system for storing the binding between features, as employed in some traditional models[125]. The observations are instead consistent with a view that

emphasises uncertainty in memory representations, which applies not only to the reported feature, but also to the cue feature. This uncertainty can lead to a non-target item in memory being judged as matching the given cue, especially if the non-target item is similar to the target in its cue feature. Figure 4A&B illustrates how this mechanism can give rise to swap errors, even if the underlying (noisy) memory representation explicitly encodes feature conjunctions. Recent findings suggest that such an account based on variability in memory for cue features is sufficient to fully explain swap errors in analogue report tasks[126].

Consistent with this mechanism, most current models of WM assume that binding between features is inherently encoded in the memory representation. This is either implemented through activity in conjunctive neural population codes, in which each neuron's activity is modulated by multiple stimulus features[17,18,47], or through rapidly formed synaptic connections between neurons sensitive for a single feature[16,127]. For instance, the interference model[16] employs a two-dimensional binding space as its central working memory substrate, which encodes feature conjunctions with limited precision and gives rise to swap errors dependent on cue feature similarity, as outlined above. This mechanism is combined with separate single-feature representations and set-size dependent background noise to give rise to different forms of recall errors, and can quantitatively fit experimental data from continuous reproduction as well as change detection tasks[128]. Models based on conjunctive coding have likewise been successful at fitting behavioral results[17,18], with an interesting recent extension additionally describing feature binding across multiple levels of visual processing[129].

Among visual features, location has long been considered to have a special role in both perception and WM[130,131]. Unlike other features, location is robustly recalled even when task-irrelevant[132–135], albeit with reduced precision[136]. Location is a particularly effective retrieval cue[119], and spatial congruency between stimuli affects recall performance[137,138].

It has sometimes been argued that binding of objects to locations constitutes a weaker (relational or extrinsic) binding than that between an object's features such as shape and color[139,140]. In contrast, Schneegans and Bays[18] proposed that binding in WM, as in visual perception, is achieved through feature maps over visual space, with different non-spatial features of an object bound to each other only indirectly via their shared location. This account allows for independent resource pools for different non-spatial features while still employing inherently conjunctive memory representations, and it explains patterns of error correlations in dual-report paradigms[98,141–143]. Recent work further indicates that for sequentially presented stimuli, presentation time may take a similar role as location in binding visual features ([144–146]; see also[17,147]). However, Son and colleagues[148] criticized the dual-report methodology employed in several of these studies[18,141,144], arguing that it may underestimate error correlations for different visual features of an object by using sequential reports. The authors found that a simultaneous report method revealed reliable correlations of memory quality for color and orientation (though still weaker than those between location and other features,[149]), which they interpreted as evidence that features in WM are organized at least partly in an object-based manner.

Feature binding in WM has also been investigated in clinical populations and older adults. Recent work shows no specific decline in binding performance associated with healthy aging[150–152], nor with most other clinical conditions[153,154]. However, a specific binding impairment has been observed in association with Alzheimer's disease[153,155] and has been proposed as a diagnostic tool to differentiate Alzheimer's from other forms of dementia[156].

Multiple competing sources of bias in WM

In addition to swap errors, where one feature is inadvertently reported in place of another, a diverse range of influences have been identified that produce graded shifts in target feature estimates towards or away from other points in the feature space. For example, Golomb and colleagues[137] found that shifting attention between memory items increased the frequency of swap errors, whereas attending to items simultaneously tended to result in reports being shifted slightly towards each other (Fig 4C).

One important source of biases is the history of previously observed stimuli with similar features. Attempts to characterize these influences have identified multiple competing sources of bias, some attracting current representations toward preceding stimuli and some repulsing them away, with systematic differences in strength, time course, and specificity[157–159].

Classical adaptation effects[160], exemplified by the tilt after-effect (Fig 4D) and the waterfall illusion, are typically repulsive, tightly spatially localized, and have their effects in immediate perception of stimuli, feeding through to WM representations. Such short-term adaptation may co-exist with or contribute to efficient encoding strategies based on long-term environmental statistics (see above). In contrast, more recently identified biases associated with the term "serial dependence"[161,162] are primarily attractive and appear to generalize across a broader range of spatial locations while specifically affecting stimulus features similar to those of preceding stimuli (Fig 4E). These attractive effects are typically observed only for stimulus features maintained in WM, and grow in strength with delay interval[163–165]. One possibility is that this reflects a greater reliance on stimulus history when the representation of the current stimulus becomes less precise, following Bayesian principles[166–168]; in perceptual tasks, where uncertainty is less, smaller attractive biases may be masked or cancelled out by repulsive biases associated with classical adaptation.

The attractive biases to preceding stimuli described as serial dependence are typically observed experimentally as influences of items presented on previous trials, which have therefore ceased to be relevant to the instructed task. In contrast, previously-presented stimuli within the same trial, which remain relevant to the current task and are presumably actively maintained in WM, have been found to have a repulsive influence on subsequent stimuli (Fig 4F;[169–171]). It is currently unclear whether the mechanisms that attract recall estimates towards previous stimuli are inactive while those stimuli remain relevant, or are active but overwhelmed by stronger repulsive biases between items held simultaneously in memory.

Repulsion is also commonly observed between two similar stimuli when they are presented simultaneously ([169,172,173]). This bias causes the stimuli to be reported as more distinct from each other than they really were, and it has has been suggested that implicitly differentiating memory representations in this way could serve to reduce inter-item confusion ([174]). By contrast, when many items are held in mind, or when memories are weak for another reason ([175]), items tend to be reported as more similar to each other than they really were (Fig 4G;[69,169,172,173,176,177]). This has been explained in terms of memories being 'compressed' (see above).

Finally, there are biases that variously attract or repel stimulus estimates relative to fixed points or landmarks in the stimulus space, some evident in immediate perception (e.g. cardinal repulsion, discussed above; Fig 4H), some that develop during a memory delay (e.g. compressive biases in spatial memory;[178]), and others that may arise at the decision stage (e.g. reference repulsion;[179]). A unifying theory of such biases has not yet been found.

Changes in WM over delay

The maintenance of information in WM over delays is imperfect, and the results from analogue report tasks confirm that the precision of individual memory representations deteriorates over time[180,181]. However, this effect is relatively subtle and variable[182] in comparison to the strong and robust effects of set size.

The gradual deterioration of WM representations has been addressed in continuous attractor models (Figure 5A). This type of model employs an idealized population of neurons whose tuning functions cover the space of possible feature values. A memorized feature is then represented by activity in a group of neurons with similar preferred feature values, sustained over time by recurrent excitation. Delay effects can be explained in these models by random drift, i.e. gradual shifts in the subset of active neurons due to noise in neural activity[46,183,184].

Several memory decoding studies have observed gradual changes in encoded feature values over time that correlate with response errors, consistent with this theoretical account[185–187]. This account is further supported by behavioural results comparing response errors and latencies across different set size and delay conditions[181], and is consistent with findings from signal detection analyses of behavioural data indicating that deterioration of memory is driven by accumulation of internal noise[188]. Another study observed that drift over delays is not entirely random, but rather shows biases towards specific feature values[189].

While attractor models of WM have typically been designed to maintain only a point estimate of a stimulus, recent work aims to incorporate uncertainty as well, e.g. represented in the amplitude of the population activity[190,191]. In future work, neural models of WM could focus on how this richer representation is used in decision-making; trained recurrent networks have already proven useful to yield mechanistic insights in tandem with accounts of behavioural data[192].

Deterioration of memory over time may also be driven by interference between multiple memory items[193]. One proposed model explains this effect by a combination of sharing representational resources in an attractor model with efficient encoding[194]. Another model combines separate continuous attractor networks, each storing a single feature, with a randomly connected neural network in which different feature representations interfere with each other to explain both set size and delay effects[195].

Directed interactions between items as described in the previous section also evolve over time. In particular, repulsion between memorized feature values has been observed to increase with longer retention intervals[173,174]. Such interactions also occur in continuous attractor models as a result of mutual excitation and inhibition between active sub-populations[46,184,196,197], although it is not clear whether these effects can fully account for the behavioural observations.

Dynamic neural representations

The continuous attractor models addressed in the previous section reflect a traditional view on the neural mechanism underlying WM, in which information is maintained through persistent activity in feature-sensitive neurons, driven by some form of recurrent excitation. This yields stable representations in the state space of neural activities (Figure 5B, left panel). Support for such a mechanism comes from electrophysiological studies in monkeys, in particular in delayed oculomotor response tasks[185,198–200]. Persistent activity has also been observed in rare electrophysiology studies in humans[201,202].

However, a number of recent works have challenged various aspects of this view, primarily based on studies that decode memory content from fMRI or EEG recordings using techniques such as inverted

encoding models[203,204]. In this type of study, it has often been found that there is little generalization in decoder efficacy between sample and delay period[205,206], or between different phases of the delay period[207–209]. While changes in neural representations immediately after stimulus presentation may reflect transitions from perceptual and iconic memory[210] to WM, qualitative changes in representational format during maintenance are inconsistent with traditional conceptualizations of WM as implemented in attractor models. This has lead to postulates that WM activity is substantially more dynamic than previously recognized[211,212] (Figure 5B, middle panel). This view is also supported by a number of electrophysiological studies in rodents and monkeys that found a reproducible sequence of activation states during the memory delay, rather than a single stable state[205,213–215]. In neural network models, it has been shown that both stable persistent activity and reproducible sequences of activation states can arise as WM mechanisms dependent on task demands and network parameters[192].

The conflicting findings may at least in part be reconciled by recent studies analyzing the neural coding of WM content in macaque monkeys. These confirmed the presence of strong temporal dynamics, allowing for instance the decoding of time passed since stimulus presentation, but also found stable subspaces in the neural code (Figure 5B, right panel) within which time-invariant decoding of memory content is possible[216–219]. This would in particular allow the read-out of memory via fixed synaptic weights despite changing activation states. Consistent results have also been obtained in an EEG experiment in humans[187].

Activity-silent WM and the focus of attention

Beyond the debate on stable vs dynamic representations, it has also been questioned in recent years whether continuous neural activity is necessary at all for WM maintenance. An alternative proposal is that at any time only a small portion of working memory content that is currently behaviorally relevant is represented through neural activity, often just a single item. This active memory is sometimes equated with the "focus of attention" in previous models[220,221]. Other items are proposed to be held in an activity-silent state[211] realized through mechanisms classically associated with long term memory, such as rapid synaptic plasticity or short-term changes in neural excitability[222,223].

The primary motivation for this idea is findings from the dual retro-cue paradigm, in which participants view two sample stimuli, and then perform two sequential memory tests for which one sample item is cued. LaRocque and colleagues[224,225] observed that the identity of the currently attended (cued) item could be decoded from neural activity using either EEG or fMRI recordings, but the currently unattended item could not (Figure 5C). Critically, a previously unattended item became decodable again if it was cued for the second test, demonstrating that it was still held in memory. A similar restoration in the decodability of memory items has also been observed following an informative retrospective cue[226], and transiently following a transcranial magnetic stimulation pulse[227] or a salient, but task-irrelevant visual stimulus[206]. The latter result has been explained by interactions of the stimulus with activity-silent WM states, e.g. in the form of altered synaptic connectivity, that elicit an identifiable impulse response in the neural activity (see also[205], for a similar account of findings in monkey electrophysiology).

Computational models based on activity-silent memory mechanisms have accounted for neurophysiological data from working memory tasks, including noise correlations[228] and trial-to-trial variations in neural responses[127,229]. Moreover, a recurrent neural network endowed with short-term synaptic plasticity developed predominantly activity-silent mechanisms to solve working memory tasks as long as these required no active manipulation of information[230]. An alternative to the standard activity-silent account proposes that unattended memory items are maintained actively, but in an altered representational format[231,232]. This has also been explored in computational models[233].

Several other papers have countered the claims of activity-silent working memory. The results of the decoding studies, which relied primarily on null results, have been called into questions by work successfully decoding the identity of unattended items[234,235] (Figure 5D), even in data that had previously been used as support for activity-silent states[236]. Schneegans and Bays[237] further demonstrated in a neural network model that restoration of decodability following an informative cue can also arise in a system with purely active WM states, and is no evidence for activity-silent memory states.

Activity-silent memory states are in conflict with assumptions underlying commonly used methods of estimating the number of items held in memory from neural activity. In particular, the strength of contralateral delay activity in EEG data increases with memory load[238,239], saturating at higher set sizes[240], and memory load can also be estimated through classification methods applied to multivariate EEG[241] or fMRI data[242]. It is possible that these measures arise despite the presence of activity-silent states, e.g. due to switching of the active state between multiple memory items. Sutterer and colleagues[243] tested this by comparing the strength of reconstructions for memorized locations from EEG data across different set sizes, and concluded that multiple locations are maintained concurrently in neural activity. In view of these results, some authors have argued that the findings supporting activity-silent memory simply reflect contributions of classical long-term memory in WM tasks, without demonstrating a specific neural WM mechanism[244,245]. The debate about the activity state of WM representations is also linked to the ongoing question of their anatomical localization[246–249], although the latter has generally been studied without the possibility of activity-silent memory in mind.

The debate on different neural WM states has parallels in the debate over different functional states in cognitive models, although caution must be taken when equating the two[250]. Models that assume that only a single item can be in the focus of attention[16,220], giving it a privileged role in influencing visual attention, contrast with alternative conceptualizations in which the focus of attention can encompass multiple items[251]. This debate takes a more concrete form in the question of whether only one[252,253] or multiple WM representations[254,255] can serve simultaneously as templates for visual search. A possible resolution to this question may be provided by recent findings indicating that multiple search templates may be prepared in parallel with little cost, but a bottleneck arises when these templates are engaged to select multiple targets[256]. Alternatively, due to variations in noise across items, it may be that it is rare for more than a single item to be represented accurately enough to successfully guide attention[257].

Another proposal is that WM is maintained by intermittent bursts of activity[258–260], bridged by mechanisms such as synaptic plasticity[222,223]. Proponents of this model point out that the appearance of persistent firing is often an artifact of averaging across trials, which hides trial-to-trial variability in neural activity[261]. The debate on the degree of persistence in neural firing during WM maintenance is still ongoing[262,263]. Unlike the proposal of activity-silent memory, the intermittent activity account does not imply different neural mechanisms for different functional memory states (e.g. attended vs. unattended items), but it may explain observations of rhythmic fluctuations in the strength of attentional guidance between multiple memory items[264].

WM versus perception and future directions

The past decade of research has brought into focus similarities and differences between visual WM and visual perception, two strongly overlapping psychological constructs studied using similar experimental methods but to a large extent by separate researchers in independent literatures. Many theoretical and experimental findings conceived of in terms of perception have counterparts in WM and vice versa, e.g. prioritization based on stimulus salience and goal relevance, probabilistic inference and use of uncertainty,

efficient coding and influences of environmental statistics. Whereas the limited capacity of visual WM was once considered fundamentally different in nature to the factors limiting visual perception, it is increasingly clear that both can be described in terms of the relative amplitude of signal to noise (SNR), with increasing WM load decreasing SNR for each stimulus in memory similarly to how decreasing visual contrast affects a discrimination judgement. Indeed, introducing perceptual or attentional bottlenecks on performance seems to change error distributions in a similar way to increasing set size[23,265,266].

Despite these areas of similarity, it is clear that WM is much more than a passive persistence of sensory-invoked activity. There are unique challenges associated with maintaining selected elements of sensory information over time independently of subsequent input, and controlling what information is added, removed, replaced and updated in memory. Key questions for further research include: How is sensory information selected for maintenance in WM – is the mechanism of selection distinct from the operation of selective visual attention (e.g.,[267])? What mechanisms allow sensory input to be segregated from existing WM representations, or integrated with it, according to behavioural requirements (e.g.,[268])? Are errors in long-term memory representations fundamentally different from those in WM and perception[269], or can they all be unified in a single model?

In answering these questions it will be critical to move beyond lab-based studies using sparse, static displays and single responses to consider richer, uncertainty-based representations, as well as how WM is deployed during natural behaviour in everyday environments. While initial steps have been taken in this direction experimentally[270–272], most computational models of WM aim only to capture recall of visual stimuli with low dimensionality. The rapidly advancing capability of artificial neural networks (ANNs) to perform dimensionality reduction on complex images may represent an opportunity to extend WM models into the real world (e.g.,[273]).

References

1. Wade, N. & Swanston, M. *Visual perception: An introduction* (Psychology Press, 2013).

2. Knill, D. C. & Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719, DOI: 10.1016/j.tins.2004.10.007 (2004).

3. Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. brain sciences* **24**, 87–114 (2001).

4. Trommershauser, J., Kording, K. & Landy, M. S. *Sensory cue integration* (Computational Neuroscience, 2011).

5. Rademaker, R. L., Tredway, C. H. & Tong, F. Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *J. vision* **12**, 21–21 (2012).

6. Yoo, A. H., Klyszejko, Z., Curtis, C. E. & Ma, W. J. Strategic allocation of working memory resource. *Sci. reports* **8**, 1–8 (2018).

7. Honig, M., Ma, W. J. & Fougnie, D. Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proc. Natl. Acad. Sci.* **117**, 8391–8397 (2020).

8. Jabar, S. B. *et al.* Probabilistic and rich individual working memories revealed by a betting game. *Sci. Reports* DOI: 10.1038/s41598-023-48242-x (in press).

9. Acerbi, L., Vijayakumar, S. & Wolpert, D. M. On the origins of suboptimality in human probabilistic inference. *PLoS computational biology* **10**, e1003661 (2014).

10. Keshvari, S., Van den Berg, R. & Ma, W. J. Probabilistic computation in human perception under variability in encoding precision. *PLoS One* **7**, e40216 (2012).

11. Yoo, A. H., Acerbi, L. & Ma, W. J. Uncertainty is maintained and used in working memory. *J. vision* **21**, 13–13 (2021).

12. Devkar, D., Wright, A. A. & Ma, W. J. Monkeys and humans take local uncertainty into account when localizing a change. *J. Vis.* **17**, 4–4 (2017).

13. Meyniel, F., Sigman, M. & Mainen, Z. F. Confidence as bayesian probability: From neural origins to behavior. *Neuron* **88**, 78–92 (2015).

14. Fleming, S. M. & Daw, N. D. Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychol. review* **124**, 91 (2017).

15. Yeon, J. & Rahnev, D. The suboptimality of perceptual decision making with multiple alternatives. *Nat. communications* **11**, 1–12 (2020).

16. Oberauer, K. & Lin, H.-Y. An interference model of visual working memory. *Psychol. Rev.* **124**, 21–59, DOI: 10.1037/rev0000044 (2017).

17. Swan, G. & Wyble, B. The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophys.* **76**, 2136–2157, DOI: 10.3758/s13414-014-0633-3 (2014).

18. Schneegans, S. & Bays, P. M. Neural Architecture for Feature Binding in Visual Working Memory. *The J. Neurosci.* **37**, 3913–3925, DOI: 10.1523/JNEUROSCI.3493-16.2017 (2017).

19. Bays, P. M. Noise in Neural Populations Accounts for Errors in Working Memory. *J. Neurosci.* **34**, 3632–3645 (2014).

20. Pouget, A., Dayan, P. & Zemel, R. Information processing with population codes. *Nat. Rev. Neurosci.* **1**, 125–32 (2000).

21. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. A normalization model of multisensory integration. *Nat. Neurosci.* **14**, 775–782, DOI: 10.1038/nn.2815 (2011).

22. Reynolds, J. H. & Heeger, D. J. The Normalization Model of Attention. *Neuron* **61**, 168–185, DOI: 10.1016/j.neuron.2009.01.002 (2009).

23. Bays, P. M. A signature of neural coding at human perceptual limits. *J. Vis.* **16**, 4, DOI: 10.1167/16.11.4 (2016).

24. Van den Berg, R., Yoo, A. H. & Ma, W. J. Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychol. review* **124**, 197 (2017).

25. Li, H.-H., Sprague, T. C., Yoo, A. H., Ma, W. J. & Curtis, C. E. Joint representation of working memory and uncertainty in human cortex. *Neuron* **109**, 3699–3712.e6, DOI: 10.1016/j.neuron.2021.08.022 (2021).

26. Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. neuroscience* **9**, 1432–1438 (2006).

27. Jazayeri, M. & Movshon, J. A. Optimal representation of sensory information by neural populations. *Nat. neuroscience* **9**, 690–696 (2006).

28. Van Bergen, R. & Jehee, J. Tafkap: An improved method for probabilistic decoding of cortical activity. *BioRxiv* 2021–03 (2021).

29. Schneegans, S., Taylor, R. & Bays, P. M. Stochastic sampling provides a unifying account of visual working memory limits. *Proc. Natl. Acad. Sci.* DOI: 10.1073/pnas.2004306117 (2020).

30. Palmer, J. Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 332–350, DOI: 10.1037/0096-1523.16.2.332 (1990).

31. Zhang, W. & Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233–235, DOI: nature06860 (2008).

32. Sewell, D. K., Lilburn, S. D. & Smith, P. L. An information capacity limitation of visual short-term memory. *J. experimental psychology: human perception performance* **40**, 2214, DOI: 10.1037/a0037744 (2014).

33. Shaw, M. L. Identifying attentional and decision-making components in information processing. *Atten. performance VIII* **8**, 277–295 (1980).

34. Ma, W. J. & Huang, W. No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *J. Vis.* **9**, 3–3 (2009).

35. Vul, E., Alvarez, G., Tenenbaum, J. & Black, M. Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Adv. neural information processing systems* **22** (2009).

36. Schurgin, M. W., Wixted, J. T. & Brady, T. F. Psychophysical scaling reveals a unified theory of visual memory strength. *Nat. Hum. Behav.* 1–17, DOI: 10.1038/s41562-020-00938-0 (2020).

37. Kriegeskorte, N. & Wei, X.-X. Neural tuning and representational geometry. *Nat. Rev. Neurosci.* **22**, 703–718, DOI: 10.1038/s41583-021-00502-3 (2021).

38. Tomić, I. & Bays, P. M. Perceptual similarity judgments do not predict the distribution of errors in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* DOI: 10.1037/xlm0001172 (2022).

39. Van den Berg, R., Shin, H., Chou, W.-C., George, R. & Ma, W. J. Variability in Encoding Precision Accounts for Visual Short-Term Memory Limitations. *Proc. Natl. Acad. Sci.* **109**, 8780–8785, DOI: 10.1073/pnas.1117465109 (2012).

40. Mazyar, H., Van den Berg, R. & Ma, W. J. Does precision decrease with set size? *J. vision* **12**, 10–10 (2012).

41. Keshvari, S., Van den Berg, R. & Ma, W. J. No evidence for an item limit in change detection. *PLoS computational biology* **9**, e1002927 (2013).

42. Van den Berg, R., Awh, E. & Ma, W. J. Factorial comparison of working memory models. *Psychol. review* **121**, 124 (2014).

43. Williams, J. R., Robinson, M. M., Schurgin, M., Wixted, J. & Brady, T. You can't "count" how many items people remember in working memory: The importance of signal detection-based measures for understanding change detection performance. *J. Exp. Psychol. Hum. Percept. Perform.* (2022).

44. Adam, K. C., Vogel, E. K. & Awh, E. Clear evidence for item limits in visual working memory. *Cogn. psychology* **97**, 79–97 (2017).

45. Wilken, P. & Ma, W. J. A detection theory account of change detection. *J. vision* **4**, 11–11 (2004).

46. Johnson, J. S., Spencer, J. P., Luck, S. J. & Schöner, G. A Dynamic Neural Field Model of Visual Working Memory and Change Detection. *Psychol. Sci.* **20**, 568–577 (2009).

47. Schneegans, S., Spencer, J. P. & Schöner, G. Integrating "what" and "where": Visual working memory for objects in a scene. In *Dynamic Thinking: A Primer on Dynamic Field Theory* (Oxford University Press, 2015).

48. Emrich, S. M., Lockhart, H. A. & Al-Aidroos, N. Attention Mediates the Flexible Allocation of Visual Working Memory Resources. *J. Exp. Psychol. Hum. Percept. Perform.* DOI: 10.1037/xhp0000398 (2017).

49. Gorgoraptis, N., Catalao, R. F. G., Bays, P. M. & Husain, M. Dynamic Updating of Working Memory Resources for Visual Objects. *J. Neurosci.* **31**, 8502–8511, DOI: 10.1523/JNEUROSCI.0208-11.2011 (2011).

50. Rajsic, J., Sun, S. Z., Huxtable, L., Pratt, J. & Ferber, S. Pop-out and pop-in: Visual working memory advantages for unique items. *Psychon. Bull. & Rev.* **23**, 1787–1793, DOI: 10.3758/s13423-016-1034-5 (2016).

51. Klyszejko, Z., Rahmati, M. & Curtis, C. E. Attentional priority determines working memory precision. *Vis. research* **105**, 70–76 (2014).

52. Brissenden, J. A., Adkins, T. J., Hsu, Y. T. & Lee, T. G. Reward influences the allocation but not the availability of resources in visual working memory. *J. Exp. Psychol. Gen.* (2023).

53. Gong, M. & Li, S. Learned reward association improves visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 841 (2014).

54. Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62, DOI: 10.1038/nrn3136 (2012).

55. Lieder, F. & Griffiths, T. L. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43** (2020).

56. Van den Berg, R. & Ma, W. J. A resource-rational theory of set size effects in human visual working memory. *ELife* **7**, e34963 (2018).

57. van den Berg, R., Zou, Q., Li, Y. & Ma, W. J. No effect of monetary reward in a visual working memory task. *PloS one* **18**, e0280257 (2023).

58. Bengson, J. J. & Luck, S. J. Effects of strategy on visual working memory capacity. *Psychon. Bull. & Rev.* **23**, 265–270 (2016).

59. Mystakidou, M. & van den Berg, R. More motivated but equally good: No effect of gamification on visual working memory performance, DOI: 10.1101/2020.01.12.903203 (2020).

60. Yoo, A. H. & Collins, A. G. How working memory and reinforcement learning are intertwined: a cognitive, neural, and computational perspective. *J. cognitive neuroscience* **34**, 551–568 (2022).

61. Collins, A. G. & Frank, M. J. How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *Eur. J. Neurosci.* **35**, 1024–1035 (2012).

62. Collins, A. G. The tortoise and the hare: Interactions between reinforcement learning and working memory. *J. cognitive neuroscience* **30**, 1422–1432 (2018).

63. Brewer, W. F. & Treyens, J. C. Role of schemata in memory for places. *Cogn. psychology* **13**, 207–230 (1981).

64. Bates, C. J. & Jacobs, R. A. Efficient data compression in perception and perceptual memory. *Psychol. review* **127**, 891 (2020).

65. Brady, T. F., Konkle, T. & Alvarez, G. A. Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. Gen.* **138**, 487 (2009).

66. Orhan, A. E., Sims, C. R., Jacobs, R. A. & Knill, D. C. The adaptive nature of visual working memory. *Curr. directions psychological science* **23**, 164–170 (2014).

67. Sims, C. R., Jacobs, R. A. & Knill, D. C. An ideal observer analysis of visual working memory. *Psychol. review* **119**, 807 (2012).

68. Lew, T. F. & Vul, E. Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *J. Vis.* **15**, 10, DOI: 10.1167/15.4.10 (2015).

69. Orhan, A. E. & Jacobs, R. A. A probabilistic clustering theory of the organization of visual short-term memory. *Psychol. review* **120**, 297 (2013).

70. Brady, T. F. & Tenenbaum, J. B. A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* **120**, 85–109, DOI: 10.1037/a0030779 (2013).

71. Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. neuroscience* **14**, 926–932 (2011).

72. Huttenlocher, J., Hedges, L. V., Corrigan, B. & Crawford, L. E. Spatial categories and the estimation of location. *Cognition* **93**, 75–97 (2004).

73. Ganguli, D. & Simoncelli, E. P. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural computation* DOI: 10.1162/NECO_a_00638 (2014).

74. Wei, X.-X. & Stocker, A. A. A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nat. neuroscience* **18**, 1509–1517 (2015).

75. Morais, M. & Pillow, J. W. Power-law efficient neural codes provide general link between perceptual bias and discriminability. *Adv. Neural Inf. Process. Syst.* **31** (2018).

76. Taylor, R. & Bays, P. M. Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *J. Neurosci.* 1018–18, DOI: 10.1523/JNEUROSCI.1018-18.2018 (2018).

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

78. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. review* **63**, 81 (1956).

79. Simon, H. A. How big is a chunk? by combining data from several experiments, a basic human memory unit can be identified and measured. *Science* **183**, 482–488 (1974).

80. Bae, G.-Y., Olkkonen, M., Allred, S. R. & Flombaum, J. I. Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *J. Exp. Psychol. Gen.* **144**, 744 (2015).

81. Hardman, K. O., Vergauwe, E. & Ricker, T. J. Categorical working memory representations are used in delayed estimation of continuous colors. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 30 (2017).

82. Mathy, F. & Feldman, J. What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* **122**, 346–362, DOI: 10.1016/j.cognition.2011.11.003 (2012).

83. Norris, D., Kalm, K. & Hall, J. Chunking and redintegration in verbal short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **46**, 872 (2020).

84. Ngiam, W. X., Brissenden, J. A. & Awh, E. "memory compression" effects in visual working memory are contingent on explicit long-term memory. *J. Exp. Psychol. Gen.* **148**, 1373 (2019).

85. Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. science* **15**, 106–111 (2004).

86. Asp, I. E., Störmer, V. S. & Brady, T. F. Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful. *J. cognitive neuroscience* **33**, 902–918 (2021).

87. Starr, A., Srinivasan, M. & Bunge, S. A. Semantic knowledge influences visual working memory in adults and children. *PloS one* **15**, e0241110 (2020).

88. Brady, T. F. & Störmer, V. S. The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. *J. Exp. Psychol. Learn. Mem. Cogn.* (2021).

89. Kaiser, D., Stein, T. & Peelen, M. V. Real-world spatial regularities affect visual working memory for objects. *Psychon. Bull. & Rev.* **22**, 1784–1790 (2015).

90. Hu, R. & Jacobs, R. A. Semantic influence on visual working memory of object identity and location. *Cognition* **217**, 104891 (2021).

91. O'Donnell, R. E., Clement, A. & Brockmole, J. R. Semantic and functional relationships among objects increase the capacity of visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **44**, 1151 (2018).

92. Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychol. Rev.* **77**, 1 (1970).

93. Park, I. M. & Pillow, J. W. Bayesian efficient coding. *BioRxiv* 178418 (2020).

94. Weber, A. I., Krishnamurthy, K. & Fairhall, A. L. Coding principles in adaptation. *Annu. review vision science* **5**, 427–449 (2019).

95. Benucci, A., Saleem, A. B. & Carandini, M. Adaptation maintains population homeostasis in primary visual cortex. *Nat. neuroscience* **16**, 724–729 (2013).

96. Shin, H. & Ma, W. J. Visual short-term memory for oriented, colored objects. *J. Vis.* **17**, 12, DOI: 10.1167/17.9.12 (2017).

97. Fougnie, D., Asplund, C. L. & Marois, R. What are the units of storage in visual working memory? *J. vision* **10**, 27–27 (2010).

98. Bays, P. M., Wu, E. Y. & Husain, M. Storage and binding of object features in visual working memory. *Neuropsychologia* **49**, 1622–1631, DOI: 10.1016/j.neuropsychologia.2010.12.023 (2011).

99. Ye, C., Hu, Z., Ristaniemi, T., Gendron, M. & Liu, Q. Retro-dimension-cue benefit in visual working memory. *Sci. Reports* **6**, 35573, DOI: 10.1038/srep35573 (2016).

100. Park, Y. E., Sy, J. L., Hong, S. W. & Tong, F. Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychol. Sci.* **28**, 1773–1785, DOI: doi.org/10.1177/0956797617719949 (2017).

101. Hajonides, J. E., van Ede, F., Stokes, M. G. & Nobre, A. C. Comparing the prioritization of items and feature-dimensions in visual working memory. *J. Vis.* **20**, 25, DOI: 10.1167/jov.20.8.25 (2020).

102. Palmer, J., Boston, B. & Moore, C. M. Limited capacity for memory tasks with multiple features within a single object. *Attention, Perception, & Psychophys.* **77**, 1488–1499 (2015).

103. Oberauer, K. & Eichenberger, S. Visual working memory declines when more features must be remembered for each object. *Mem. & cognition* **41**, 1212–1227 (2013).

104. Hardman, K. O. & Cowan, N. Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *J. Exp. Psychol. Learn. Mem. Cogn.* **41**, 325 (2015).

105. Chen, H. & Wyble, B. Attribute amnesia reflects a lack of memory consolidation for attended information. *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 225–234, DOI: 10.1037/xhp0000133 (2016).

106. Wyble, B., Hess, M., O'Donnell, R. E., Chen, H. & Eitam, B. Learning how to exploit sources of information. *Mem. & Cogn.* **47**, 696–705, DOI: 10.3758/s13421-018-0881-x (2019).

107. Shin, H. & Ma, W. J. Crowdsourced single-trial probes of visual working memory for irrelevant features. *J. Vis.* **16**, 10, DOI: 10.1167/16.5.10 (2016).

108. Swan, G., Collins, J. & Wyble, B. Memory for a single object has differently variable precisions for relevant and irrelevant features. *J. Vis.* **16**, 32, DOI: 10.1167/16.3.32 (2016).

109. Yu, Q. & Shim, W. M. Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage* **157**, 97–107, DOI: 10.1016/j.neuroimage.2017.05.055 (2017).

110. Bocincova, A. & Johnson, J. S. The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. *Cortex* **111**, 196–209, DOI: 10.1016/j.cortex.2018.10.013 (2019).

111. Marshall, L. & Bays, P. M. Obligatory encoding of task-irrelevant features depletes working memory resources. *J. Vis.* **13**, 21–21, DOI: 10.1167/13.2.21 (2013).

112. Wang, B., Cao, X., Theeuwes, J., Olivers, C. N. L. & Wang, Z. Location-based effects underlie feature conjunction benefits in visual working memory. *J. Vis.* **16**, 12, DOI: 10.1167/16.11.12 (2016).

113. Markov, Y. A., Tiurina, N. A. & Utochkin, I. S. Different features are stored independently in visual working memory but mediated by object-based representations. *Acta Psychol.* **197**, 52–63, DOI: 10.1016/j.actpsy.2019.05.003 (2019).

114. Brady, T. F., Konkle, T. & Alvarez, G. A. A review of visual memory capacity: Beyond individual items and toward structured representations. *J. vision* **11**, 4–4 (2011).

115. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working memory is set by allocation of a shared resource. *J. Vis.* **9**, 7–7, DOI: 10.1167/9.10.7 (2009).

116. Huang, L. Distinguishing target biases and strategic guesses in visual working memory. *Attention, Perception, & Psychophys.* **82**, 1258–1270, DOI: 10.3758/s13414-019-01913-2 (2020).

117. Pratte, M. S. Swap errors in spatial working memory are guesses. *Psychon. Bull. & Rev.* **26**, 958–966, DOI: 10.3758/s13423-018-1524-8 (2019).

118. Rajsic, J. & Wilson, D. E. Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophys.* **76**, 1902–1913, DOI: 10.3758/s13414-014-0723-2 (2014).

119. Rajsic, J., Swan, G., Wilson, D. E. & Pratt, J. Accessibility limits recall from visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 1415–1431, DOI: 10.1037/xlm0000387 (2017).

120. Bays, P. M. Evaluating and excluding swap errors in analogue tests of working memory. *Sci. Reports* **6**, 19203, DOI: 10.1038/srep19203 (2016).

121. Emrich, S. M. & Ferber, S. Competition increases binding errors in visual working memory. *J. Vis.* **12**, 12–12, DOI: 10.1167/12.4.12 (2012).

122. Rerko, L., Oberauer, K. & Lin, H.-Y. Spatial Transposition Gradients in Visual Working Memory. *Q. J. Exp. Psychol.* **67**, 3–15, DOI: 10.1080/17470218.2013.789543 (2014).

123. Souza, A. S., Rerko, L., Lin, H.-Y. & Oberauer, K. Focused attention improves working memory: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophys.* **76**, 2080–2102, DOI: 10.3758/s13414-014-0687-2 (2014).

124. Sahan, M. I., Dalmaijer, E. S., Verguts, T., Husain, M. & Fias, W. The Graded Fate of Unattended Stimulus Representations in Visuospatial Working Memory. *Front. Psychol.* **10**, 374, DOI: 10.3389/fpsyg.2019.00374 (2019).

125. Wheeler, M. E. & Treisman, A. M. Binding in short-term visual memory. *J. Exp. Psychol. Gen.* **131**, 48–64, DOI: 10.1037/0096-3445.131.1.48 (2002).

126. McMaster, J. M. V., Tomić, I., Schneegans, S. & Bays, P. M. Swap errors in visual working memory are fully explained by cue-feature variability. *Cogn. Psychol.* (2022).

127. Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P. & Husain, M. Neural mechanisms of attending to items in working memory. *Neurosci. & Biobehav. Rev.* **101**, 1–12, DOI: 10.1016/j.neubiorev.2019.03.017 (2019).

128. Lin, H.-Y. & Oberauer, K. An interference model for visual working memory: Applications to the change detection task. *Cogn. Psychol.* **133**, 101463, DOI: 10.1016/j.cogpsych.2022.101463 (2022).

129. Hedayati, S., O′Donnell, R. & Wyble, B. Memory for Latent Representations: An Account of Working Memory that Builds on Visual Knowledge for Efficient and Detailed Visual Representations. Preprint, Neuroscience (2021). DOI: 10.1101/2021.02.07.430171.

130. Treisman, A. & Zhang, W. Location and binding in visual working memory. *Mem. & Cogn.* **34**, 1704–1719, DOI: 10.3758/BF03195932 (2006).

131. Huang, L. Unit of visual working memory: A Boolean map provides a better account than an object does. *J. Exp. Psychol. Gen.* **149**, 1–30, DOI: 10.1037/xge0000616 (2020).

132. Chen, H. & Wyble, B. The location but not the attributes of visual cues are automatically encoded into working memory. *Vis. Res.* **107**, 76–85, DOI: 10.1016/j.visres.2014.11.010 (2015).

133. Kondo, A. & Saiki, J. Feature-Specific Encoding Flexibility in Visual Working Memory. *PLoS ONE* **7**, e50962, DOI: 10.1371/journal.pone.0050962 (2012).

134. Foster, J. J., Bsales, E. M., Jaffe, R. J. & Awh, E. Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Curr. Biol.* **27**, 3216–3223.e6, DOI: 10.1016/j.cub.2017.09.031 (2017).

135. Cai, Y., Sheldon, A. D., Yu, Q. & Postle, B. R. Overlapping and distinct contributions of stimulus location and of spatial context to nonspatial visual short-term memory. *J. Neurophysiol.* **121**, 1222–1231, DOI: 10.1152/jn.00062.2019 (2019).

136. Tam, J. & Wyble, B. Location has a privilege, but it is limited: Evidence from probing task-irrelevant location. *J. Exp. Psychol. Learn. Mem. Cogn.* DOI: 10.1037/xlm0001147 (2022).

137. Golomb, J. D., Kupitz, C. N. & Thiemann, C. T. The influence of object location on identity: A "spatial congruency bias". *J. Exp. Psychol. Gen.* **143**, 2262–2278, DOI: 10.1037/xge0000017 (2014).

138. Teng, C. & Postle, B. R. Spatial specificity of feature-based interaction between working memory and visual processing. *J. Exp. Psychol. Hum. Percept. Perform.* **47**, 495–507, DOI: 10.1037/xhp0000899 (2021).

139. Parra, M. A. *et al.* Relational and conjunctive binding functions dissociate in short-term memory. *Neurocase* **21**, 56–66 (2015).

140. Piekema, C., Rijpkema, M., Fernández, G. & Kessels, R. P. Dissociating the neural correlates of intra-item and inter-item working-memory binding. *PloS one* **5**, e10214 (2010).

141. Fougnie, D. & Alvarez, G. A. Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *J. Vis.* **11**, 3–3, DOI: 10.1167/11.12.3 (2011).

142. Kovacs, O. & Harris, I. M. The role of location in visual feature binding. *Attention, Perception, & Psychophys.* **81**, 1551–1563, DOI: 10.3758/s13414-018-01638-8 (2019).

143. Markov, Y. A., Utochkin, I. S. & Brady, T. F. Real-world objects are not stored in holistic representations in visual working memory. *J. Vis.* **21**, 18, DOI: 10.1167/jov.21.3.18 (2021).

144. Schneegans, S., McMaster, J. M. V. & Bays, P. M. Role of time in binding features in visual working memory. *Psychol. Rev.* DOI: 10.1037/rev0000331 (2022).

145. Heuer, A. & Rolfs, M. Incidental encoding of visual information in temporal reference frames in working memory. *Cognition* **207**, 104526, DOI: 10.1016/j.cognition.2020.104526 (2021).

146. Heuer, A. & Rolfs, M. Temporal and spatial reference frames in visual working memory are defined by ordinal and relational properties. *J. Exp. Psychol. Learn. Mem. Cogn.* (2022).

147. Bowman, H. & Wyble, B. The simultaneous type, serial token model of temporal attention and working memory. *Psychol. review* **114**, 38 (2007).

148. Sone, H., Kang, M.-S., Li, A. Y., Tsubomi, H. & Fukuda, K. Simultaneous estimation procedure reveals the object-based, but not space-based, dependence of visual working memory representations. *Cognition* **209**, 104579, DOI: 10.1016/j.cognition.2020.104579 (2021).

149. Brown, G., Kasem, I., Bays, P. M. & Schneegans, S. Mechanisms of feature binding in visual working memory are stable over long delays. *J. Vis.* **21**, 7–7 (2021).

150. Read, C. A., Rogers, J. M. & Wilson, P. H. Working memory binding of visual object features in older adults. *Aging, Neuropsychol. Cogn.* **23**, 263–281, DOI: 10.1080/13825585.2015.1083937 (2016).

151. Rhodes, S., Parra, M. A., Cowan, N. & Logie, R. H. Healthy aging and visual working memory: The effect of mixing feature and conjunction changes. *Psychol. Aging* **32**, 354–366, DOI: 10.1037/pag0000152 (2017).

152. Pertzov, Y., Heider, M., Liang, Y. & Husain, M. Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychol. Aging* **30**, 26–35, DOI: 10.1037/a0038396 (2015).

153. Della Sala, S., Parra, M. A., Fabi, K., Luzzi, S. & Abrahams, S. Short-term memory binding is impaired in AD but not in non-AD dementias. *Neuropsychologia* **50**, 833–840, DOI: 10.1016/j.neuropsychologia.2012.01.018 (2012).

154. Lugtmeijer, S. *et al.* Consequence of stroke for feature recall and binding in visual working memory. *Neurobiol. Learn. Mem.* **179**, 107387, DOI: 10.1016/j.nlm.2021.107387 (2021).

155. Liang, Y. *et al.* Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex* **78**, 150–164, DOI: 10.1016/j.cortex.2016.01.015 (2016).

156. Martínez, J. F., Trujillo, C., Arévalo, A., Ibáñez, A. & Cardona, J. F. Assessment of Conjunctive Binding in Aging: A Promising Approach for Alzheimer's Disease Detection. *J. Alzheimer's Dis.* **69**, 71–81, DOI: 10.3233/JAD-181154 (2019).

157. Fornaciai, M. & Park, J. Attractive serial dependence between memorized stimuli. *Cognition* **200**, 104250, DOI: 10.1016/j.cognition.2020.104250 (2020). Publisher: Elsevier.

158. Czoschke, S., Peters, B., Rahm, B., Kaiser, J. & Bledowski, C. Visual objects interact differently during encoding and memory maintenance. *Attention, Perception, & Psychophys.* **82**, 1241–1257, DOI: 10.3758/s13414-019-01861-x (2020).

159. Teng, C., Fulvio, J. M., Jiang, J. & Postle, B. R. Flexible top-down control in the interaction between working memory and perception. *J. Vis.* **22**, 3–3 (2022).

160. Webster, M. A. Visual Adaptation. *Annu. Rev. Vis. Sci.* **1**, 547–567, DOI: 10.1146/annurev-vision-082114-035509 (2015). _eprint: https://doi.org/10.1146/annurev-vision-082114-035509.

161. Cicchini, G. M., Benedetto, A. & Burr, D. C. Perceptual history propagates down to early levels of sensory analysis. *Curr. Biol.* **31**, 1245–1250.e2, DOI: 10.1016/j.cub.2020.12.004 (2021).

162. Kiyonaga, A., Scimeca, J. M., Bliss, D. P. & Whitney, D. Serial Dependence across Perception, Attention, and Memory. *Trends Cogn. Sci.* **21**, 493–497, DOI: 10.1016/j.tics.2017.04.011 (2017).

163. Bliss, D. P., Sun, J. J. & D'Esposito, M. Serial dependence is absent at the time of perception but increases in visual working memory. *Sci. Reports* **7**, 14739, DOI: 10.1038/s41598-017-15199-7 (2017). Tex.ids= BlissEtAl2017a number: 1 publisher: Nature Publishing Group.

164. Barbosa, J. & Compte, A. Build-up of serial dependence in color working memory. *Sci. Reports* **10**, 10959, DOI: 10.1038/s41598-020-67861-2 (2020). Number: 1 Publisher: Nature Publishing Group.

165. Fritsche, M., Mostert, P. & de Lange, F. P. Opposite Effects of Recent History on Perception and Decision. *Curr. Biol.* **27**, 590–595, DOI: 10.1016/j.cub.2017.01.006 (2017).

166. Bergen, R. S. v. & Jehee, J. F. M. Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *J. Neurosci.* **39**, 8164–8176, DOI: 10.1523/JNEUROSCI.3212-18.2019 (2019). Publisher: Society for Neuroscience Section: Research Articles.

167. Fritsche, M., Spaak, E. & de Lange, F. P. A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife* **9**, e55389, DOI: 10.7554/eLife.55389 (2020).

168. Cicchini, G. M., Mikellidou, K. & Burr, D. C. The functional role of serial dependence. *Proc. Royal Soc. B* **285**, 20181722 (2018). Publisher: The Royal Society.

169. Bae, G.-Y. & Luck, S. J. Interactions between visual working memory representations. *Attention, Perception, & Psychophys.* **79**, 2376–2395, DOI: 10.3758/s13414-017-1404-8 (2017).

170. Czoschke, S., Fischer, C., Beitner, J., Kaiser, J. & Bledowski, C. Two types of serial dependence in visual working memory. *Br. J. Psychol.* **110**, 256–267, DOI: 10.1111/bjop.12349 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12349.

171. Kang, M.-S. & Choi, J. Retrieval-induced inhibition in short-term memory. *Psychol. Sci.* **26**, 1014–1025, DOI: 10.1177/0956797615577358 (2015). Publisher: Sage Publications Sage CA: Los Angeles, CA.

172. Lively, Z., Robinson, M. M. & Benjamin, A. S. Memory Fidelity Reveals Qualitative Changes in Interactions Between Items in Visual Working Memory. *Psychol. Sci.* **32**, 1426–1441, DOI: 10.1177/0956797621997367 (2021).

173. Chunharas, C., Rademaker, R. L., Brady, T. F. & Serences, J. T. An adaptive perspective on visual working memory distortions. *J. Exp. Psychol. Gen.* (2022).

174. Scotti, P. S., Hong, Y., Golomb, J. D. & Leber, A. B. Statistical learning as a reference point for memory distortions: Swap and shift errors. *Attention, Perception, & Psychophys.* **83**, 1652–1672, DOI: 10.3758/s13414-020-02236-3 (2021).

175. Dubé, C., Zhou, F., Kahana, M. J. & Sekuler, R. Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vis. Res.* **96**, 8–16, DOI: 10.1016/j.visres.2013.12.016 (2014).

176. Brady, T. F. & Alvarez, G. A. Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychol. Sci.* **22**, 384–392, DOI: 10.1177/0956797610397956 (2011).

177. Papenmeier, F. & Timm, J. D. Do group ensemble statistics bias visual working memory for individual items? A registered replication of Brady and Alvarez (2011). *Attention, Perception, & Psychophys.* **83**, 1329–1336, DOI: 10.3758/s13414-020-02209-6 (2021).

178. Sheth, B. R. & Shimojo, S. Compression of space in visual memory. *Vis. research* **41**, 329–341, DOI: 10.1016/S0042-6989(00)00230-3 (2001).

179. Luu, L. & Stocker, A. A. Categorical judgments do not modify sensory representations in working memory. *PLOS Comput. Biol.* **17**, e1008968, DOI: 10.1371/journal.pcbi.1008968 (2021).

180. Rademaker, R. L., Park, Y. E., Sack, A. T. & Tong, F. Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 925–940, DOI: 10.1037/xhp0000491 (2018).

181. Schneegans, S. & Bays, P. M. Drift in Neural Population Activity Causes Working Memory to Deteriorate Over Time. *The J. Neurosci.* **38**, 4859–4869, DOI: 10.1523/JNEUROSCI.3440-17.2018 (2018).

182. Shin, H., Zou, Q. & Ma, W. J. The effects of delay duration on visual working memory for orientation. *J. Vis.* **17**, 10, DOI: 10.1167/17.14.10 (2017).

183. Compte, A., Brunel, N., Goldman-Rakic, P. S. & Wang, X.-J. Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cereb. Cortex* **10**, 910–923, DOI: 10.1093/cercor/10.9.910 (2000).

184. Wei, Z., Wang, X.-J. & Wang, D.-H. From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. *J. Neurosci.* **32**, 11228–11240, DOI: 10.1523/JNEUROSCI.0735-12.2012 (2012).

185. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439, DOI: 10.1038/nn.3645 (2014).

186. Lim, P. C., Ward, E. J., Vickery, T. J. & Johnson, M. R. Not-so-working Memory: Drift in Functional Magnetic Resonance Imaging Pattern Representations during Maintenance Predicts Errors in a Visual Working Memory Task. *J. Cogn. Neurosci.* **31**, 1520–1534, DOI: 10.1162/jocn_a_01427 (2019).

187. Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J. & Stokes, M. G. Drifting codes within a stable coding scheme for working memory. *PLOS Biol.* **18**, e3000625, DOI: 10.1371/journal.pbio.3000625 (2020).

188. Kuuramo, C., Saarinen, J. & Kurki, I. Forgetting in visual working memory: Internal noise explains decay of feature representations. *J. Vis.* **22**, 8–8 (2022).

189. Panichello, M. F., DePasquale, B., Pillow, J. W. & Buschman, T. J. Error-correcting dynamics in visual working memory. *Nat. Commun.* **10**, 3366, DOI: 10.1038/s41467-019-11298-3 (2019).

190. Carroll, S., Josić, K. & Kilpatrick, Z. P. Encoding certainty in bump attractors. *J. computational neuroscience* **37**, 29–48 (2014).

191. Kutschireiter, A., Basnak, M. A., Wilson, R. I. & Drugowitsch, J. Bayesian inference in ring attractor networks. *Proc. Natl. Acad. Sci.* **120**, e2210622120, DOI: 10.1073/pnas.2210622120 (2023).

192. Orhan, A. E. & Ma, W. J. A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nat. Neurosci.* **22**, 275–283, DOI: 10.1038/s41593-018-0314-y (2019).

193. Pertzov, Y., Manohar, S. & Husain, M. Rapid forgetting results from competition over time between items in visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **43**, 528–536, DOI: 10.1037/xlm0000328 (2017).

194. Koyluoglu, O. O., Pertzov, Y., Manohar, S., Husain, M. & Fiete, I. R. Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *eLife* **6**, e22225, DOI: 10.7554/eLife.22225 (2017).

195. Bouchacourt, F. & Buschman, T. J. A Flexible Model of Working Memory. *Neuron* **103**, 147–160.e8, DOI: 10.1016/j.neuron.2019.04.020 (2019).

196. Almeida, R., Barbosa, J. & Compte, A. Neural circuit basis of visuo-spatial working memory precision: A computational and behavioral study. *J. Neurophysiol.* **114**, 1806–1818, DOI: 10.1152/jn.00362.2015 (2015).

197. Johnson, J. S., van Lamsweerde, A. E., Dineva, E. & Spencer, J. P. Neural interactions in working memory explain decreased recall precision and similarity-based feature repulsion. *Sci. Reports* **12**, 17756 (2022).

198. Fuster, J. M. & Alexander, G. E. Neuron Activity Related to Short-Term Memory. *Science* **173**, 652–654, DOI: 10.1126/science.173.3997.652 (1971).

199. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349, DOI: 10.1152/jn.1989.61.2.331 (1989).

200. Hart, E. & Huk, A. C. Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *eLife* **9**, e52460, DOI: 10.7554/eLife.52460 (2020).

201. Kamiński, J. *et al.* Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nat. Neurosci.* **20**, 590–601, DOI: 10.1038/nn.4509 (2017).

202. Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I. & Mormann, F. Persistent Single-Neuron Activity during Working Memory in the Human Medial Temporal Lobe. *Curr. Biol.* **27**, 1026–1032, DOI: 10.1016/j.cub.2017.02.013 (2017).

203. Brouwer, G. J. & Heeger, D. J. Decoding and Reconstructing Color from Responses in Human Visual Cortex. *J. Neurosci.* **29**, 13992–14003, DOI: 10.1523/JNEUROSCI.3577-09.2009 (2009).

204. Ester, E. F., Anderson, D. E., Serences, J. T. & Awh, E. A Neural Measure of Precision in Visual Working Memory. *J. Cogn. Neurosci.* **25**, 754–761, DOI: 10.1162/jocn_a_00357 (2013).

205. Stokes, M. G. *et al.* Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron* **78**, 364–375, DOI: 10.1016/j.neuron.2013.01.039 (2013).

206. Wolff, M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* **20**, 864–871, DOI: 10.1038/nn.4546 (2017).

207. Sreenivasan, K. K., Vytlacil, J. & D'Esposito, M. Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex. *J. Cogn. Neurosci.* **26**, 1141–1153, DOI: 10.1162/jocn_a_00556 (2014).

208. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *J. Neurophysiol.* **100**, 1407–1419, DOI: 10.1152/jn.90248.2008 (2008).

209. Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T. & Kennerley, S. W. Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nat. Commun.* **9**, 3498, DOI: 10.1038/s41467-018-05873-3 (2018).

210. Coltheart, M. Iconic memory and visible persistence. *Percept. & Psychophys.* **27**, 183–228, DOI: 10.3758/BF03204258 (1980).

211. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405, DOI: 10.1016/j.tics.2015.05.004 (2015).

212. Postle, B. R. Neural Bases of the Short-term Retention of Visual Information. In *Mechanisms of Sensory Working Memory*, 43–58, DOI: 10.1016/B978-0-12-801371-7.00005-3 (Elsevier, 2015).

213. Baeg, E. *et al.* Dynamics of Population Code for Working Memory in the Prefrontal Cortex. *Neuron* **40**, 177–188, DOI: 10.1016/S0896-6273(03)00597-X (2003).

214. MacDonald, C. J., Lepage, K. Q., Eden, U. T. & Eichenbaum, H. Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. *Neuron* **71**, 737–749, DOI: 10.1016/j.neuron.2011.07.012 (2011).

215. Scott, B. B. *et al.* Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron* **95**, 385–398.e5, DOI: 10.1016/j.neuron.2017.06.013 (2017).

216. Murray, J. D. *et al.* Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci.* **114**, 394–399, DOI: 10.1073/pnas.1619449114 (2017).

217. Parthasarathy, A. *et al.* Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nat. Commun.* **10**, 4995, DOI: 10.1038/s41467-019-12841-y (2019).

218. Spaak, E., Watanabe, K., Funahashi, S. & Stokes, M. G. Stable and Dynamic Coding for Working Memory in Primate Prefrontal Cortex. *The J. Neurosci.* **37**, 6503–6516, DOI: 10.1523/JNEUROSCI.3364-16.2017 (2017).

219. Cueva, C. J. *et al.* Low-dimensional dynamics for working memory and time encoding. *Proc. Natl. Acad. Sci.* **117**, 23021–23032, DOI: 10.1073/pnas.1915984117 (2020).

220. Oberauer, K. Access to information in working memory: Exploring the focus of attention. *J. Exp. Psychol. Learn. Mem. Cogn.* **28**, 411–421, DOI: 10.1037/0278-7393.28.3.411 (2002).

221. Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *J. Cogn. Neurosci.* **24**, 61–79, DOI: 10.1162/jocn_a_00140 (2012).

222. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic Theory of Working Memory. *Science* **319**, 1543–1546, DOI: 10.1126/science.1150769 (2008).

223. Barak, O. & Tsodyks, M. Working models of working memory. *Curr. Opin. Neurobiol.* **25**, 20–24, DOI: 10.1016/j.conb.2013.10.008 (2014).

224. LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K. & Postle, B. R. Decoding Attended Information in Short-term Memory: An EEG Study. *J. Cogn. Neurosci.* **25**, 127–142, DOI: 10.1162/jocn_a_00305 (2013).

225. LaRocque, J. J., Riggall, A. C., Emrich, S. M. & Postle, B. R. Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cereb. Cortex* cercor;bhw283v1, DOI: 10.1093/cercor/bhw283 (2017).

226. Sprague, T. C., Ester, E. F. & Serences, J. T. Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron* **91**, 694–707, DOI: 10.1016/j.neuron.2016.07.006 (2016).

227. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354**, 1136–1139, DOI: 10.1126/science.aah7011 (2016).

228. Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M. & Richmond, B. J. Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS computational biology* **4**, e1000073 (2008).

229. Bocincova, A., Buschman, T. J., Stokes, M. G. & Manohar, S. G. Neural signature of flexible coding in prefrontal cortex. *Proc. Natl. Acad. Sci.* **119**, e2200400119 (2022).

230. Masse, N. Y., Yang, G. R., Song, H. F., Wang, X.-J. & Freedman, D. J. Circuit mechanisms for the maintenance and manipulation of information in working memory. *Nat. neuroscience* **22**, 1159–1167 (2019).

231. van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J. & Olivers, C. N. Current and future goals are represented in opposite patterns in object-selective cortex. *ELife* **7**, e38677 (2018).

232. Yu, Q., Teng, C. & Postle, B. R. Different states of priority recruit different neural representations in visual working memory. *PLoS biology* **18**, e3000769 (2020).

233. Wan, Q., Menendez, J. A. & Postle, B. R. Priority-based transformations of stimulus representation in visual working memory. *PLOS Comput. Biol.* **18**, e1009062 (2022).

234. Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J.-D. Cortical specialization for attended versus unattended working memory. *Nat. Neurosci.* **21**, 494–496, DOI: 10.1038/s41593-018-0094-4 (2018).

235. Iamshchinina, P., Christophel, T. B., Gayet, S. & Rademaker, R. L. Essential considerations for exploring visual working memory storage in the human brain. *Vis. Cogn.* **29**, 425–436, DOI: 10.1080/13506285.2021.1915902 (2021).

236. Barbosa, J., Lozano-Soldevilla, D. & Compte, A. Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. *PLOS Biol.* **19**, e3001436, DOI: 10.1371/journal.pbio.3001436 (2021).

237. Schneegans, S. & Bays, P. M. Restoration of fMRI Decodability Does Not Imply Latent Working Memory States. *J. Cogn. Neurosci.* **29**, 1977–1994, DOI: 10.1162/jocn_a_01180 (2017).

238. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual differences in visual working memory capacity. *Nature* **428**, 748–751, DOI: 10.1038/nature02447 (2004).

239. Luria, R., Balaban, H., Awh, E. & Vogel, E. K. The contralateral delay activity as a neural measure of visual working memory. *Neurosci. & Biobehav. Rev.* **62**, 100–108, DOI: 10.1016/j.neubiorev.2016.01.003 (2016).

240. Bays, P. M. Reassessing the Evidence for Capacity Limits in Neural Signals Related to Working Memory. *Cereb. Cortex* **28**, 1432–1438, DOI: 10.1093/cercor/bhx351 (2018).

241. Adam, K. C. S., Vogel, E. K. & Awh, E. Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology* **57**, DOI: 10.1111/psyp.13691 (2020).

242. Emrich, S. M., Riggall, A. C., LaRocque, J. J. & Postle, B. R. Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *J. Neurosci.* **33**, 6516–6523, DOI: 10.1523/JNEUROSCI.5732-12.2013 (2013).

243. Sutterer, D. W., Foster, J. J., Adam, K. C. S., Vogel, E. K. & Awh, E. Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLOS Biol.* **17**, e3000239, DOI: 10.1371/journal.pbio.3000239 (2019).

244. Beukers, A. O., Buschman, T. J., Cohen, J. D. & Norman, K. A. Is Activity Silent Working Memory Simply Episodic Memory? *Trends Cogn. Sci.* **25**, 284–293, DOI: 10.1016/j.tics.2021.01.003 (2021).

245. Foster, J. J., Vogel, E. K. & Awh, E. Working memory as persistent neural activity. Preprint, PsyArXiv (2019). DOI: 10.31234/osf.io/jh6e3.

246. Riley, M. R. & Constantinidis, C. Role of Prefrontal Persistent Activity in Working Memory. *Front. Syst. Neurosci.* **9**, DOI: 10.3389/fnsys.2015.00181 (2016).

247. D'Esposito, M. & Postle, B. R. The Cognitive Neuroscience of Working Memory. *Annu. Rev. Psychol.* **66**, 115–142, DOI: 10.1146/annurev-psych-010814-015031 (2015).

248. Xu, Y. Revisit once more the sensory storage account of visual working memory. *Vis. Cogn.* **28**, 433–446, DOI: 10.1080/13506285.2020.1818659 (2020).

249. Serences, J. T. Neural mechanisms of information storage in visual short-term memory. *Vis. research* **128**, 53–67 (2016).

250. Stokes, M. G., Muhle-Karbe, P. S. & Myers, N. E. Theoretical distinction between functional states in working memory and their corresponding neural states. *Vis. Cogn.* **28**, 420–432, DOI: 10.1080/13506285.2020.1825141 (2020).

251. Cowan, N. The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia* **49**, 1401–1406, DOI: 10.1016/j.neuropsychologia.2011.01.035 (2011).

252. Olivers, C. N., Peters, J., Houtkamp, R. & Roelfsema, P. R. Different states in visual working memory: When it guides attention and when it does not. *Trends Cogn. Sci.* S1364661311000854, DOI: 10.1016/j.tics.2011.05.004 (2011).

253. Ort, E., Fahrenfort, J. J. & Olivers, C. N. L. Lack of free choice reveals the cost of multiple-target search within and across feature dimensions. *Attention, Perception, & Psychophys.* **80**, 1904–1917, DOI: 10.3758/s13414-018-1579-7 (2018).

254. Beck, V. M., Hollingworth, A. & Luck, S. J. Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychol. Sci.* **23**, 887–898, DOI: 10.1177/0956797612439068 (2012).

255. Bahle, B., Thayer, D. D., Mordkoff, J. T. & Hollingworth, A. The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *J. Exp. Psychol. Gen.* **149**, 967–983, DOI: 10.1037/xge0000694 (2020).

256. Ort, E., Fahrenfort, J. J., ten Cate, T., Eimer, M. & Olivers, C. N. Humans can efficiently look for but not select multiple visual objects. *eLife* **8**, e49130, DOI: 10.7554/eLife.49130 (2019).

257. Williams, J. R., Brady, T. F. & Störmer, V. S. Guidance of attention by working memory is a matter of representational fidelity. *J. Exp. Psychol. Hum. Percept. Perform.* (2022).

258. Lundqvist, M., Compte, A. & Lansner, A. Bistable, irregular firing and population oscillations in a modular attractor memory network. *PLoS Comput. Biol.* **6**, e1000803 (2010).

259. Lundqvist, M. *et al.* Gamma and Beta Bursts Underlie Working Memory. *Neuron* **90**, 152–164, DOI: 10.1016/j.neuron.2016.02.028 (2016).

260. Fiebig, F. & Lansner, A. A spiking working memory model based on hebbian short-term potentiation. *J. Neurosci.* **37**, 83–96 (2017).

261. Shafi, M. *et al.* Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* **146**, 1082–1108, DOI: 10.1016/j.neuroscience.2006.12.072 (2007).

262. Lundqvist, M., Herman, P. & Miller, E. K. Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *The J. Neurosci.* **38**, 7013–7019, DOI: 10.1523/JNEUROSCI.2485-17.2018 (2018).

263. Constantinidis, C. *et al.* Persistent Spiking Activity Underlies Working Memory. *The J. Neurosci.* **38**, 7020–7028, DOI: 10.1523/JNEUROSCI.2486-17.2018 (2018).

264. Pomper, U. & Ansorge, U. Theta-Rhythmic Oscillation of Working Memory Performance. *Psychol. Sci.* **32**, 1801–1810, DOI: 10.1177/09567976211013045 (2021).

265. Cohen, M., Keefe, J. M. & Brady, T. Perceptual awareness occurs along a graded continuum: Evidence from psychophysical scaling. *Psychol. Sci.* (2023).

266. Taylor, R. & Bays, P. M. Theory of neural coding predicts an upper bound on estimates of memory variability. *Psychol. review* **127**, 700 (2020).

267. Zhou, Y., Curtis, C. E., Sreenivasan, K. & Fougnie, D. Common neural mechanisms control attention and working memory. *J. Neurosci.* (2022).

268. Rademaker, R. L., Chunharas, C. & Serences, J. T. Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. neuroscience* **22**, 1336–1344 (2019).

269. Miner, A. E., Schurgin, M. W. & Brady, T. F. Is working memory inherently more "precise" than long-term memory? extremely high fidelity visual long-term memories for frequently encountered objects. *J. Exp. Psychol. Hum. Percept. Perform.* **46**, 813 (2020).

270. Draschkow, D., Kallmayer, M. & Nobre, A. C. When Natural Behavior Engages Working Memory. *Curr. Biol.* **31**, 869–874.e5, DOI: 10.1016/j.cub.2020.11.013 (2021).

271. Kristjánsson, Á. & Draschkow, D. Keeping it real: Looking beyond capacity limits in visual cognition. *Attention, Perception, & Psychophys.* **83**, 1375–1390, DOI: 10.3758/s13414-021-02256-7 (2021).

272. Issen, L. A. & Knill, D. C. Decoupling eye and hand movement control: Visual short-term memory influences reach planning more than saccade planning. *J. Vis.* **12**, DOI: 10.1167/12.1.3 (2012).

273. Hedayati, S., O'Donnell, R. E. & Wyble, B. A model of working memory for latent representations. *Nat. Hum. Behav.* **6**, 709–719 (2022).

274. Bays, P. M. & Husain, M. Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science* **321**, 851–854, DOI: 10.1126/science.1158023 (2008).

275. Awh, E., Barton, B. & Vogel, E. K. Visual working memory represents a fixed number of items regardless of complexity. *Psychol. science* **18**, 622–628 (2007).

276. Pratte, M. S. Set size effects on working memory precision are not due to an averaging of slots. *Attention, Perception, & Psychophys.* **82**, 2937–2949 (2020).

277. Bays, P. M. Failure of self-consistency in the discrete resource model of visual working memory. *Cogn. Psychol.* **105**, 1–8 (2018).

278. Devkar, D. T., Wright, A. A. & Ma, W. J. The same type of visual working memory limitations in humans and monkeys. *J. vision* **15**, 13–13 (2015).

279. Pratte, M. S., Park, Y. E., Rademaker, R. L. & Tong, F. Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *J. experimental psychology. Hum. perception performance* **43**, 6, DOI: 10.1037/xhp0000302 (2017).

280. Pashler, H. Familiarity and visual change detection. *Percept. & psychophysics* **44**, 369–378 (1988).

281. Oostwoud Wijdenes, L., Marshall, L. & Bays, P. M. Evidence for Optimal Integration of Visual Feature Representations across Saccades. *J. Neurosci.* **35**, 10146–10153, DOI: 10.1523/JNEUROSCI.1040-15.2015 (2015).

282. Wolf, C. & Schütz, A. C. Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *J. Vis.* **15**, 1–1 (2015).

283. Ganmor, E., Landy, M. S. & Simoncelli, E. P. Near-optimal integration of orientation information across saccades. *J. Vis.* **15**, 8–8, DOI: 10.1167/15.16.8 (2015).

284. Kong, G., Kroell, L. M., Schneegans, S., Aagten-Murphy, D. & Bays, P. M. Transsaccadic integration relies on a limited memory resource. *J. Vis.* **21**, 24–24, DOI: 10.1167/jov.21.5.24 (2021). Publisher: The Association for Research in Vision and Ophthalmology.

285. Stewart, E. E. M. & Schütz, A. C. Optimal trans-saccadic integration relies on visual working memory. *Vis. Res.* **153**, 70–81, DOI: 10.1016/j.visres.2018.10.002 (2018).

286. Stewart, E. E. M. & Schütz, A. C. Transsaccadic integration benefits are not limited to the saccade target. *J. Neurophysiol.* **122**, 1491–1501, DOI: 10.1152/jn.00420.2019 (2019). Publisher: American Physiological Society.

287. Ohl, S. & Rolfs, M. Saccadic Eye Movements Impose a Natural Bottleneck on Visual Short-Term Memory. *J. Exp. Psychol. Learn. Mem. Cogn.* DOI: 10.1037/xlm0000338 (2016).

288. Udale, R., Tran, M. T., Manohar, S. & Husain, M. Dynamic in-flight shifts of working memory resources across saccades. *J. Exp. Psychol. Hum. Percept. Perform.* **48**, 21, DOI: 10.1037/xhp0000960 (2022). Publisher: US: American Psychological Association.

289. Shao, N. *et al.* Saccades elicit obligatory allocation of visual working memory. *Mem. & Cogn.* **38**, 629–640 (2010). Publisher: Springer.

290. Hanning, N. M., Jonikaitis, D., Deubel, H. & Szinte, M. Oculomotor selection underlies feature retention in visual working memory. *J. Neurophysiol.* **115**, 1071–1076, DOI: 10.1152/jn.00927.2015 (2016). Publisher: American Physiological Society.

291. Heuer, A., Ohl, S. & Rolfs, M. Memory for action: a functional view of selection in visual working memory. *Vis. Cogn.* **28**, 388–400, DOI: 10.1080/13506285.2020.1764156 (2020). Publisher: Routledge _eprint: https://doi.org/10.1080/13506285.2020.1764156.

292. Chen, Y. & Crawford, J. D. Allocentric representations for target memory and reaching in human cortex. *Annals New York Acad. Sci.* **1464**, 142–155, DOI: 10.1111/nyas.14261 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.14261.

293. Aagten-Murphy, D. & Bays, P. M. Functions of Memory Across Saccadic Eye Movements. *Curr. Top. Behav. Neurosci.* DOI: 10.1007/7854_2018_66 (2018).

294. Hanning, N. M. & Deubel, H. Independent Effects of Eye and Hand Movements on Visual Working Memory. *Front. Syst. Neurosci.* **12** (2018).

295. Heuer, A., Crawford, J. D. & Schubö, A. Action relevance induces an attentional weighting of representations in visual working memory. *Mem. & Cogn.* **45**, 413–427, DOI: 10.3758/s13421-016-0670-3 (2017).

296. Heuer, A. & Schubö, A. Separate and combined effects of action relevance and motivational value on visual working memory. *J. Vis.* **18**, 14, DOI: 10.1167/18.5.14 (2018).

297. Byrne, P. A. & Crawford, J. D. Cue Reliability and a Landmark Stability Heuristic Determine Relative Weighting Between Egocentric and Allocentric Visual Information in Memory-Guided Reach. *J. Neurophysiol.* **103**, 3054–3069, DOI: 10.1152/jn.01008.2009 (2010).

298. Fiehler, K., Wolf, C., Klinghammer, M. & Blohm, G. Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Front. Hum. Neurosci.* **8**, DOI: 10.3389/fnhum.2014.00636 (2014).

299. Aagten-Murphy, D. & Bays, P. M. Independent working memory resources for egocentric and allocentric spatial information. *PLoS computational biology* **15**, e1006563 (2019). Publisher: Public Library of Science.

Acknowledgements
PMB was supported by a personal fellowship from the Wellcome Trust (Grant number 106926). TFB was supported by NSF BCS-2141189 and NSF BCS-2146988.

Author contributions
All authors contributed equally to this work.

Competing interests
The authors declare no competing interests.

Boxes:

BOX 1: Slots versus resource models

Influential initial models of visual WM[3,77] were often based on the idea that, to be remembered, an object must be stored in one of a fixed number of memory *slots*, such that up to around four items could be remembered without error and beyond that limit no further items could be remembered at all. Such models were simple and made strong predictions that initially appeared to be borne out in tasks such as change detection, leading them to be highly influential. However, as evidence grew that items in memory were subject to significant variability, and that this noise increased with memory load even from one to two items (e.g.,[30,45,274]), the simple picture painted by slot models was no longer sufficient to capture the data. Alternative *resource* models were developed in which a fixed quantity of representational signal is distributed between memory items, with no fixed upper limit on the number of items represented.

Faced with the argument that noise-based accounts made the notion of slots redundant, attempts to adapt slot models have taken two main routes. First, early evidence that certain changes to complex object can be detected when it is the only item in memory but not when multiple items must be remembered (e.g.,[85]), led to the proposal that the limit of four slots coexisted with noisy storage within each slot (e.g.,[275]). Second, the influential *slots-plus-averaging model* proposed to adapt the slot model by allowing a single item to be represented in multiple slots, with averaging of the independent representations[31]. However, this model has been criticized on multiple fronts: for being functionally identical to a discrete resource model (specifically, the sample-size model, with samples re-branded as slots;[29]), for failures in self-consistency (e.g.,[276,277]) and for failing to fit performance across set sizes as accurately as the best resource models without the slot constraint[19,41,42,278]. While one study[279] argued that the better fit of resource models disappeared when stimulus-dependent sources of variability were accounted for, subsequent work[76] showed that a resource-based model incorporating efficient coding fit the same data better while also predicting the patterns of stimulus-dependent variability from first principles (see *WM in a structured environment*).

This has resulted in the slots-plus-averaging model losing favour and a return to slot models that allow for memory precision to be resource-based and vary continuously, but with an additional upper bound on how many representations can exist and hence on overall performance (e.g.,[44]). Arguments for this kind of model are typically based on observations interpreted as "true guesses" (i.e. responses that do not appear to be based on any knowledge of the previously-presented stimulus). However, all current resource models predict such zero-precision estimates (or estimates indistinguishably close to zero) as arising from probabilistic variation in precision (Fig 3), and when models have been formally fit to such data, resource models have been found to reproduce the patterns interpreted as guesses without needing an additional mechanism (e.g. in whole-report delayed estimation;[29]). Thus, pure resource accounts are criticized on the basis of patterns of data that they accurately account for, with those patterns claimed as evidence for an alternative model that has not been fully formulated in quantitative terms and has not been shown to reproduce the data.

Importantly, while slot models have changed over time from simple models that made strong predictions to resources-plus-guessing models that retain little of the original slot concept, the wider field has not always kept track of this evolution. For example, many researchers continue to report K values based on change detection data (counts of how many items are "in memory"), even though the all-or-nothing assumption underlying the calculation of K[3,280] is incompatible even with modern slot models, which assume that items are not simply present or absent from memory but at minimum also have an associated precision[31,44]. This may lead to researchers misinterpreting response biases as memory limits[43]. Relatedly, many studies fit mixture models that assume a some-or-none mixture of imprecise memories and guesses to continuous reproduction data to account for the long tail of errors, even though such models have been shown not to isolate independent precision and guess rate parameters[36,266]. In a change detection study, a variable-precision model accounted best for *apparent guesses*, even though it did not contain a guessing component[41]. Overall, then, the field should carefully specify what is meant when appealing to slot models, since such models are not generally slot-like in their character anymore, allowing for many kinds of continuous variation but specifying an additional item limit that is superfluous in accounting for empirical performance.

BOX 2: Sensorimotor functions of visual WM

Visual WM has been conceptualized as a workspace in which visual object representations are not only maintained but also manipulated (as in mental rotation), compared (as in visual search) or integrated with new input. WM has long been assumed to play a critical role in bridging interruptions of sensory input, so that processing does not have to start anew when the input is restored. In vision, common forms of interruption affecting the processing of objects in our environment include dynamic occlusions by other objects (e.g. as a result of motion parallax), movements of the head or body that briefly take the object out of the field of view, and whole-field interruptions in the form of blinks and saccadic shifts of gaze.

Saccades are the most frequent form of interruption to visual input, dislocating and briefly smearing the retinal image several times per second during natural vision. Recent studies have shown that information about an object obtained in sequential gaze fixations is integrated in a statistically near-optimal manner[281–283] and that this process relies on the allocation of limited VWM resources to

behaviourally relevant objects in advance of the eye movement[284,285]. Multiple object representations can be integrated across a saccade, including objects that are never brought into foveal vision[281,286]; however, dynamic allocation of WM resources to upcoming saccade targets seems to be obligatory and to require the withdrawal of resources from previously fixated objects[274,287–290].

WM has a broad role in supporting goal-directed movement (see[291–293] for detailed reviews). Recent studies have demonstrated enhanced recall for visual items at locations relevant to reaching movements[294,295] and also for feature dimensions relevant to a movement, e.g. object size for grasp[296]. These benefits have been observed even for movements specified shortly after disappearance of the memory array, perhaps reflecting reallocation of WM resources supported by shifts of attentional focus within sensory memory.

Action planning is thought to rely on representations of spatial location in multiple reference frames[292], that is, the encoding of an object's location relative to a stable visual landmark (allocentric coding) may be at least as relevant to action as its location in the visual field (a form of egocentric coding). The presence of a landmark at both encoding and retrieval enhances recall of object locations[297,298], increasing precision for items near to the landmark in a manner consistent with integration of allocentric and egocentric representations of an object's location maintained in independent WM stores[299]. The ability to supplement memory of an object's individual spatial location with memory for its location in relation to another object, seemingly without cost, is conceptually similar to some descriptions of inter-item interaction and ensemble representation in visual WM (see main text); future work could aim to synthesise these accounts.

Figures:

Figure 1: Recall as inference about the past. In this minimal illustration, viewing a single colour patch drawn from a continuous space of hues (left) at time t_1 induces stochastic changes in the neural system that propagate in time, resulting in one of many possible "memory states" (middle) at time t_2 when the memory is probed. The information a memory state contains about the stimulus hue is described by a likelihood function (right), the probability of obtaining that particular memory state given each stimulus hue that could have been presented at time t_1. If, as in a typical delayed estimation task, the observer is asked to select a single hue that best matches the memory (a "point estimate"), a good choice might be the maximum-likelihood estimate (coloured pins). However, the full likelihood function contains richer information about the plausibility of different hues that, to the extent the observer has access to it, may be revealed using other experimental methods (see Fig. 2).

Figure 2: Tools for measuring WM uncertainty. (A) A typical task testing orientation recall with confidence reported on an ordinal scale. (B) Increasing the number of items to be remembered (the set size) reduces the signal strength relative to noise, increasing variability (broadening of error distribution). (C) Even within a given set size (here, six items) error distributions can be decomposed on the basis of subjective confidence ratings into components that differ in precision. Panels A–C adapted from[5]. (D–E) Reporting a confidence interval (D); arc length is correlated with absolute error in the point estimate (E). Adapted from[7]. (F) In change detection, the optimal decision criterion depends on uncertainty. The x-axis represents the measured change based on noisy WM representations in a single-item change detection task. The lines represent the probability distribution of the measured change on change (blue) and no-change (red) trials. The grey areas indicate where the optimal observer would report a change. When uncertainty is high, the optimal observer tolerates a larger measured change before reporting "change". Adapted from[11].

Figure 3: (A–D) Four models of visual WM that share common principles and predict similar patterns of error in recall (see main text for details). (A) Encoding-decoding model based on representation in a population code. Stimulus features are encoded in the activity of idealized neurons individually tuned to different parts of the feature space (inset). Recall errors arise at decoding due to limited activity amplitude and Poisson variability in spike counts. (B) Sample-based model with stochastic variation in number of samples. Recall errors arise from averaging over a limited and variable number of individually noisy samples. (C) Signal detection model with correlated random noise. Recall error arises from the addition of noise to an underlying familiarity function that peaks at the stimulus feature. (D) Model based on probabilistic variability in mnemonic precision. Recall errors comprise scale mixtures of normal distributions with differing precisions. (E) Relationship between variability and uncertainty common to these models: memories that are compatible with a narrow range of stimuli (high certainty as measured by likelihood width; top) correspond to point estimates with low variability (coloured pins; top); low certainty memories correspond to high variability estimates (bottom). (F) Confidence ratings (from task shown in Fig. 2A) can be explained as a logarithmic transformation of precision and fit jointly with error. Adapted from[24]. (G) Whole-report delayed estimation with the reporting order chosen by the participant. The estimate distribution gets wider for later responses (left), consistent with selecting items in order of increasing uncertainty (right). Adapted from[29].

Figure 4: (A & B) Swap errors arising from cue feature similarity in a conjunctive coding model. (A) Example of a likelihood function over all possible combinations of cue and report feature value based on a fully conjunctive memory representation of a memory array (shown in inset, numbers for reference), with random noise. Numbered points indicate the true feature combinations of target (item 2) and non-target items. Likelihood of the report feature value associated with the cue (matching the cue value of the target item, dashed white line) is shown in the lower part of the panel, with corresponding decoded estimates, for three repetitions with the same stimuli but independent noise. (B) Distribution of decoded report feature values over many repetitions. While the majority of decoded values are concentrated around the report feature of the target item (green dashed line), a substantial proportion are close to the report feature values of non-target items (red dashed lines), in particular item 3 which has a similar cue feature value (angular location) as the target. (C) Recall error distributions display dissociable contributions from swap errors (secondary peak at non-target value) and biases (shift or skew of central peak away from target value). Data from[137]. (D–H) A diverse range of factors contributing to VWM biases.

Figure 5: Dynamics of WM representations. (A) Left: Architecture of a continuous attractor model of WM, with neurons shown as circles coloured with their preferred feature value, arranged on a ring reflecting the topology of the feature space. The pattern of synaptic connectivity is shown for one example neuron, with local excitatory connections (green) and global inhibitory connections (red). The blue circular plot shows neural firing rate briefly after stimulus presentation. Right: Neural activity during a single WM trial, showing persistent firing after stimulus offset due to recurrent excitation, and random drift in the represented feature value over time due to noise in neural activity. (B) State-space plots of WM activity for different coding schemes. The plots show a projection of the high-dimensional space of activities in a neural population onto a low-dimensional state space. Each coloured line shows the time course of the activity state in a single trial (from light to dark), with different colours corresponding to different memorized feature values. Left: Stable neural representations, in which activity states remain largely fixed for the duration of the trial, except for effects of noise and possibly an initial transient phase. Middle: Dynamic representation, with activity states changing along different trajectories for different features. Right: Representation with stable sub-spaces (here in components 1 and 2), but dynamic in orthogonal spaces (here component 3 reflects time). (C) Time course of decoding strengths from fMRI data for different stimulus categories in a dual retro-cue task (adapted from[227]). Decoding strength for the category of a currently uncued item transiently drops to chance level, suggestive of representation in an activity-silent state. (D) Decoding strength for features of different sample stimuli in another dual retro-cue task. Here, decoding strength in higher cortical areas is significantly above chance for a currently uncued memory item (adapted from[234]).