Representation and computation in 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

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 driven significant advances in 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, as 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 visual WM formulated at different levels of implementation have converged on common principles relating capacity to variability and uncertainty. Here we review recent research in human visual WM from a computational perspective and the latest developments in identifying neural mechanisms that support it.

Introduction

Since the dawn of perception research, theoretical frameworks have been built around the notions of representation and computation (Wade and Swanston, 2013). 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 (Knill and Pouget, 2004). To make meaningful progress in understanding representation and computation, experiments must be combined with mathematical models.

While this agenda has been pursued with great success in perception research, the field of visual WM research has been different. This field initially held 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. Recall can be understood as inference based on the memory state about what features were present or how they relate to features of other objects. 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

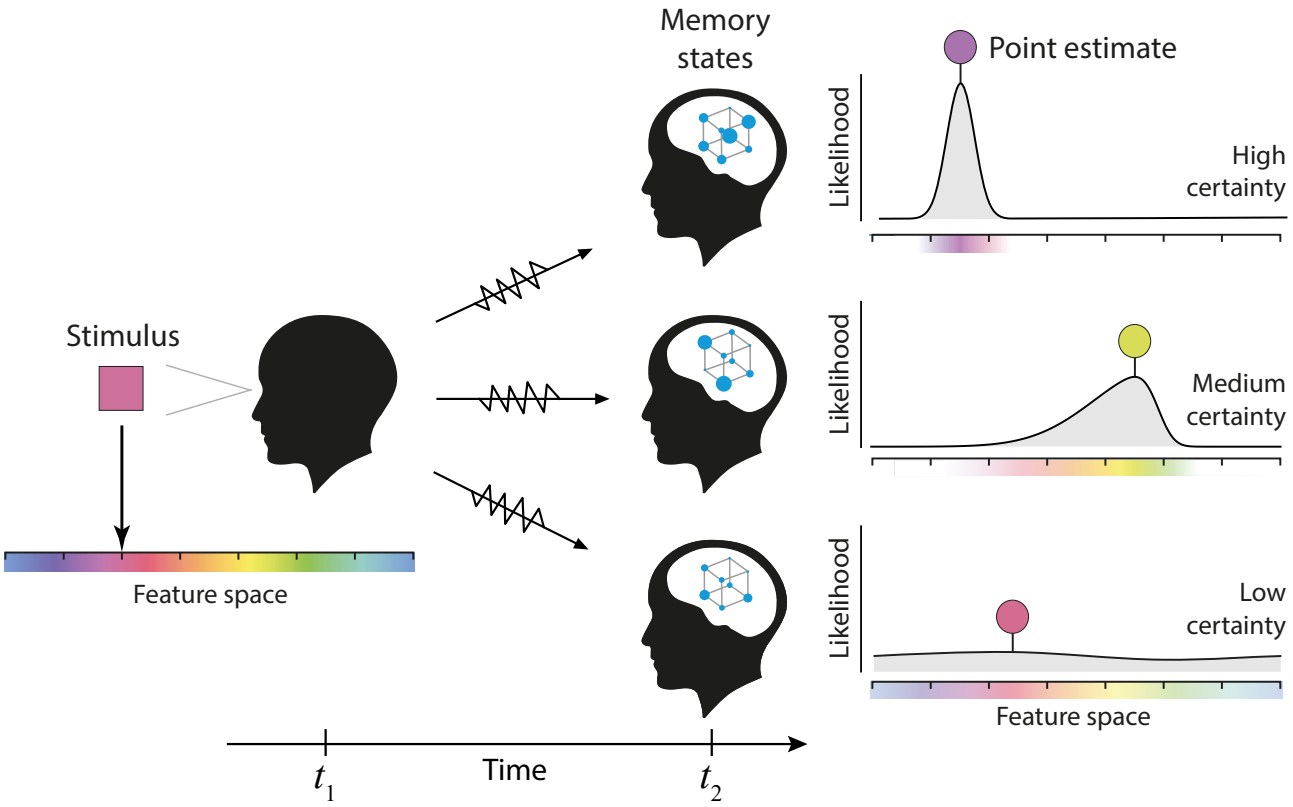

Figure 1: Recall as inference about the past. In this minimal illustration, viewing a single colour patch drawn from a continuous space of hues (left) at time t_1 induces stochastic changes in the neural system that propagate in time, resulting in one of many possible "memory states" (middle) at time 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).

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 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", illustrated by the coloured pins), 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 because the likelihood could also be asymmetric (centre right) or even multimodal.

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 (Trommershauser et al., 2011) 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 set size increases, error becomes more broadly distributed and reported confidence on average 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; Rademaker et al., 2012), 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 (Yoo et al., 2018; Honig et al., 2020; Jabar et al., 2020) (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; Yoo et al., 2018; Honig et al., 2020; Jabar et al., 2020), consistent with the studies that use confidence ratings. Moreover, in parallel to perceptual studies (Acerbi et al., 2014), observers combine their memory-based likelihood with prior information about a feature, even if that information varies from trial to trial (Honig et al., 2020).

Uncertainty can also be assessed in change detection tasks by asking whether uncertainty is taken into account implicitly in observers' decisions (Keshvari et al., 2012; Yoo et al., 2021), an approach that is particularly useful in non-human animals (Devkar et al., 2017). The basic idea is that a large change in the internal representation between the memory and the probe provides less evidence for a true change if uncertainty is higher than if it is lower (Fig. 2F). Variations in uncertainty not only arise spontaneously, but can also be experimentally induced by varying the reliability of the stimulus information from trial to trial and from item to item, a technique borrowed from the cue combination literature. These change detection studies rely on formal model comparison to conclude that observers take into account memory uncertainty in their decision.

Taken together, 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. At a fundamental level, 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 (Meyniel et al., 2015; Fleming and Daw, 2017; Yeon and Rahnev, 2020). WM researchers have started to study the analogous question (Jabar et al., 2020), with initial evidence suggesting the use of the likelihood function beyond its width.

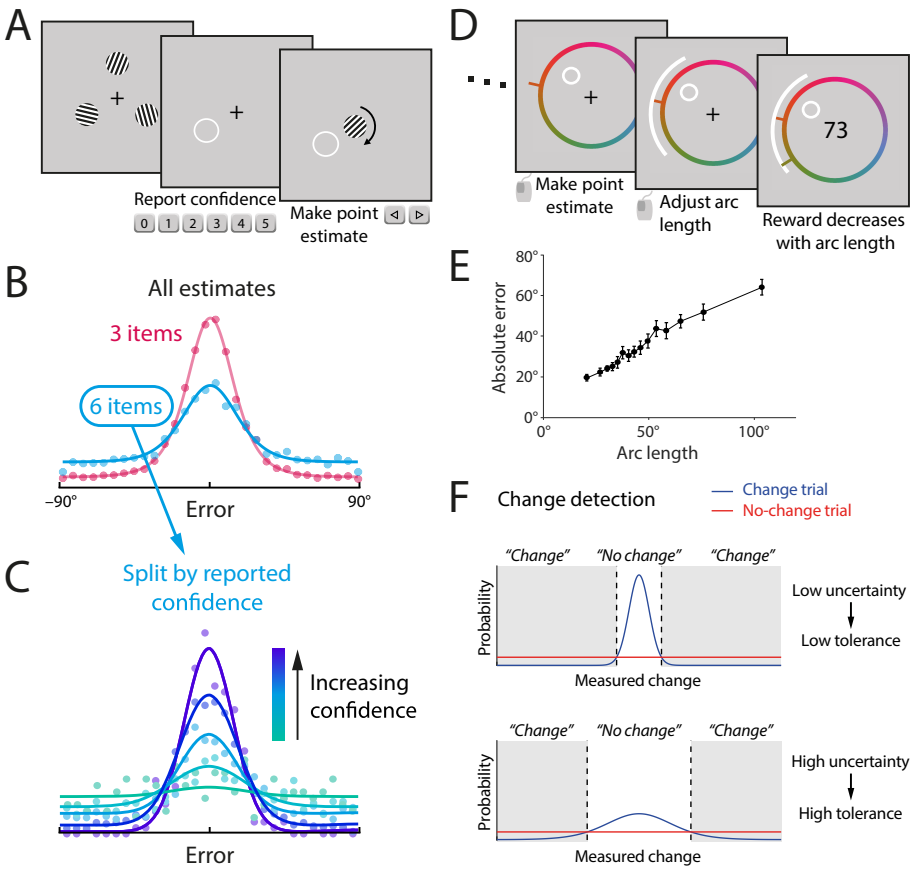

Figure 2: Tools for measuring WM uncertainty. (A) A typical task testing orientation recall with confidence reported on an ordinal scale. (B) Increasing the number of items to be remembered (the set size) reduces the signal strength relative to noise, increasing variability (broadening of error distribution). (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 Rademaker et al. (2012). (D–E) Reporting a confidence interval (D); arc length is correlated with absolute error in the point estimate (E). Adapted from Honig et al. (2020). (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 Yoo et al. (2021).

Models

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 here all imply a richer underlying stimulus representation that carries information about memory uncertainty, as described above.

Population coding accounts (Bays, 2014; Schneegans and Bays, 2017a), inspired by similar models of attention, sensory integration and decision-making (Pouget et al., 2000; Ma et al., 2006; Jazayeri and Movshon, 2006; Ohshiro et al., 2011; Reynolds and Heeger, 2009), 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, explaining why recall fidelity declines with the number of items held simultaneously in memory, and also effects of stimulus salience and behavioural priority on recall.

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; Schneegans et al. 2020). This provides a connection to cognitive models that describe resource allocation as distributing a limited (but potentially very large) number of discrete samples between memory items (Palmer, 1990; Zhang and Luck, 2008; Sewell et al., 2014), a concept that was originally proposed to model selective attention (Shaw, 1980) and that was later successfully applied to multiple-object tracking (Ma and Huang, 2009; Vul et al., 2009). The analogy with stochastic spiking imposes a key additional element needed to provide the best fits to continuous recall data: that the number of samples varies randomly and independently between items. While samples are discrete in this account (Zhang and Luck, 2008), random variability in their number fits poorly with the older concept of "slots", and the allocated resource (the mean number of samples) is a continuous variable.

As an alternative perspective related to population coding, the TCC model (Schurgin et al., 2020) describes the output of WM as a noisy familiarity signal with a mean that decays as a function of distance (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 distance 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 stimuli are from each other (Kriegeskorte and Wei, 2021); a proposed relationship with psychological similarity is a current subject of debate (Schurgin et al., 2020; Tomić and Bays, 2022).

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 (Van den Berg et al., 2012; Fougnie et al., 2012) 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 (Schneegans et al., 2020).

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 (Bays, 2016b; Schneegans et al., 2020) and variable-precision models (Van den Berg et al., 2017)

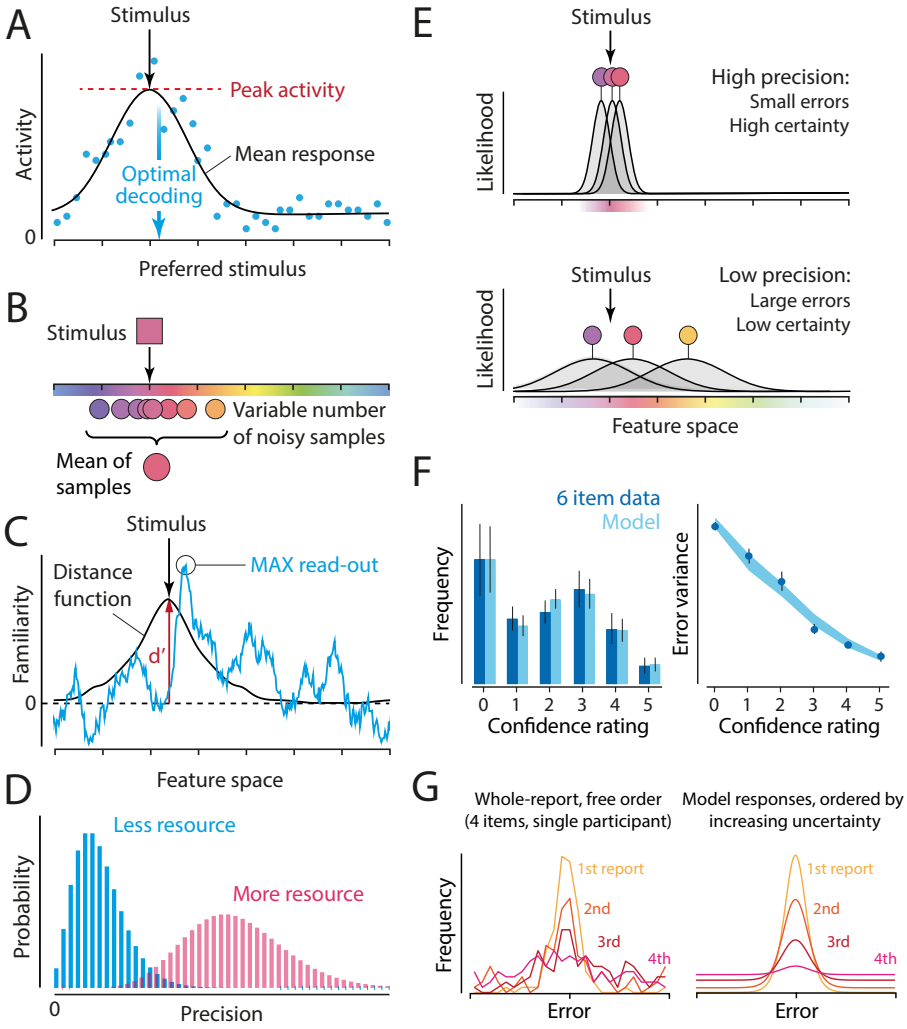

Figure 3: (A–D) Four models of visual WM that share common principles. (A) Encoding-decoding model based on representation in a population code. (B) Sample-based model with stochastic variation in the number of samples. (C) Signal detection model with correlated random noise. (D) Model based on probabilistic variability in mnemonic precision. (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 Van den Berg et al. (2017). (C) 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 Schneegans et al. (2020).

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 (Fig. 3F) 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 (Fig. 3F). Similarly, models based in signal detection theory accurately predict how performance changes with confidence in change detection tasks (Williams et al., 2022b). Moving beyond explicit reports of confidence, subjects can be asked to report the item they recall best (Fougnie et al., 2012) or to recall all items in any order they like (Adam et al., 2017) (Fig. 3G). In the latter case, the error distribution grows progressively wider for later reports, and the results can be quantitatively reproduced on the basis that participants report items in order of decreasing confidence (Schneegans et al., 2020).

A lesson emerging from 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 (Honig et al., 2020), or transformed into a decision about a categorical global variable such as presence of a target (Mazyar et al., 2012) or of a change (Wilken and Ma, 2004; Keshvari et al., 2012, 2013; Devkar et al., 2017; Yoo et al., 2021). 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 (Mazyar et al., 2012; Keshvari et al., 2012; Devkar et al., 2017; Honig et al., 2020; Yoo et al., 2021).

BOX 1: Slots versus resource models

Influential initial models of visual WM (Cowan, 2001; Luck and Vogel, 1997) 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., Palmer, 1990; Wilken and Ma, 2004; Bays and Husain, 2008), the simple picture painted by slot models was no longer sufficient to capture the data.

Faced with the argument that representational noise governed by a continuously distributed resource made the concept of a fixed item limit 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., Alvarez and Cavanagh, 2004), led to the proposal that the limit of four slots coexisted with noisy storage within each slot (e.g., Awh et al., 2007). 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 (Zhang and Luck, 2008). 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; Schneegans et al., 2020), for failures in self-consistency (e.g., Pratte, 2020; Bays, 2018a) and for failing to fit performance across set sizes as accurately as the best resource models without the slot constraint (Van den Berg et al., 2014; Devkar et al., 2015; Keshvari et al., 2013; Bays, 2014).

This has led to the abandonment of the slots-plus-averaging model and a return to slot models that allow for memory precision to be resource-based and vary continuously, but claim there might additionally be an upper bound on how many representations can exist (e.g., Van den Berg et al., 2014; Adam et al., 2017; Ngiam et al., 2019). Arguments for this kind of model are usually 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) despite the fact that all current resource models predict such zero-precision estimates (or estimates indistinguishably close to zero) as arising from probabilistic variation in precision (Fig 3). 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; Schneegans et al., 2020). Thus, pure resource accounts are criticized on the basis of patterns of data that they accurately predict, 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 fit K values to change detection data (counts of how many items are present versus absent), which depend on an all-or-none view of memory that has long been abandoned, leading to many studies mistaking response biases for memory limits (e.g.,Williams et al., 2022b). Similarly, 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 (Taylor and Bays, 2020; Schurgin et al., 2020). For example, in change detection, a simple variable-precision model best accounts for *apparent guesses*, even though it does not contain a guessing component (Keshvari et al., 2013). Overall, then, the field should carefully specify what is means 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 adding in an additional assumption of complete memory failure that is superfluous to an accurate account of empirical performance.

Resource allocation, rationality and incentives

While the nature of memory resources has been a point of intense debate in the WM community, the underlying cause of resource limitations has received less attention. Most modern models of visual WM are based on fixed pools of resources (Bays and Husain, 2008; Zhang and Luck, 2008), but allow flexibility in how those resources are allocated. This flexibility is necessary to account for a range of findings in which observers prioritize the precision of certain memoranda over others, as a result of differences in their attentional salience or relevance to behavioural goals (Emrich et al., 2017; Gorgoraptis et al., 2011; Rajsic et al., 2016). 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 (Bays, 2014; Yoo et al., 2018) 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.

More direct manipulations of incentives have also been successful in modulating performance. 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 (Klyszejko et al., 2014). While in that study, attentional priority and reward coincided, in another study reward improved performance even when these cues were dissociated (Brissenden et al., 2021). Finally, reward-associated items are remembered better even when task-irrelevant (Gong and Li, 2014).

These results are compatible with a structural constraint, perhaps neurophysiological in origin, on the representational capacity of the WM system. An alternative perspective is based on the theory of resource rationality (Lieder and Griffiths, 2020), which proposes that the brain attempts to maximize performance in a given task while at the same time minimizing a biologically relevant cost. In the case of WM, this could for example be the cost of neural spiking (Attwell and Laughlin, 2001; Lennie, 2003), which would naturally map to costs on samples or precision in the memory representations. The balance between performance and cost would be controlled by a relative scaling factor, and the behaviour of a resource-rational agent would generally deviate from optimality, if optimality is only defined in terms of maximizing performance.

In delayed-estimation tasks, effects of set size and probe probability have been successfully captured on the basis that the values of mean precision for each of the items are resource-rational under a cost linear in precision (Van den Berg and Ma, 2018). 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, e.g. by changing the attainable rewards, from one trial to another. An increased reward should shift the balance towards higher performance by compensating for the higher associated cost. In orientation delayed estimation experiments, WM performance did not improve when a monetary reward was higher, regardless of whether it was manipulated between or within subjects (Van den Berg et al., 2020), nor when the total attainable reward was raised by increasing cue validity (Brissenden et al., 2021). It is possible that the differences in reward were too small to elicit an effect, but the results suggest that the total resource available is not sensitive to reward manipulations. Interestingly, in one 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 (Bengson and Luck, 2016). However, in another study, "gamification" of a working memory task increased motivation but did not improve recall performance (Mystakidou and van den Berg, 2020).

Taken together, it seems that resource allocation in WM is highly responsive to reward differences between items or locations, while evidence for effects at the trial 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.

WM limitations have also been recognized as being an important factor in reward-based instrumental learning (Yoo and Collins, 2022). 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 (Collins and Frank, 2012; Collins, 2018). Further work should test alternative, resource-based models of WM within this task.

Within a resource-rational framework, an interesting question is why WM even exists. After all, storing a memory through spiking activity is on the surface more costly than through changes in synaptic strengths, as is done in long-term memory. Presumably, the latter mechanism has its own costs, for example associated with interference between the large number of items stored (Engle, 2002).

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 (Brewer and Treyens, 1981); 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 was entirely static, we wouldn't have to encode any information in WM. If it was 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 (Bates and Jacobs, 2020; Brady et al., 2009) and rate-distortion theory (Orhan et al., 2014; Sims et al., 2012), 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 to see (Lew and Vul, 2015; Orhan and Jacobs, 2013; Brady and Tenenbaum, 2013). 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 (Brady and Tenenbaum, 2013).

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 (Girshick et al., 2011; Huttenlocher et al., 2004). 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 (Ganguli and Simoncelli, 2014; Wei and Stocker, 2015; Morais and Pillow, 2018) . These principles can be naturally incorporated into population coding models of WM (Fig. 3A) via an optimal redistribution of tuning functions (Taylor and Bays, 2020), 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 (Luck and Vogel, 1997; Cowan, 2001). Such frameworks have generally formalized the usage of prior knowledge via the concept of chunking (Miller, 1956; Cowan, 2001). The most common conception of chunking in WM is that we learn co-occurrences and use these to create chunks in long-term memory. The content of WM is then often thought to point to 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 (Cowan, 2001; Miller, 1956; Simon, 1974). 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 (Bae et al., 2015; Hardman et al., 2017). In an information theoretic framework, chunking can be recast as an approximation to more general compression schemes: that is, chunking can be seen not as an alternative to compression but as a means of implementing such compression in models where items are treated like discrete units (Brady et al., 2009; Mathy and Feldman, 2012; Norris et al., 2020).

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 (Sims et al., 2012) or newly learned co-occurrence statistics (Brady et al., 2009; Ngiam et al., 2019). 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 (Alvarez and Cavanagh, 2004; Asp et al., 2021; Starr et al., 2020), and better with realistic objects and configurations of objects compared to simple meaningless stimuli or random configurations of objects (Brady and Störmer, 2021; Kaiser et al., 2015; Hu and Jacobs, 2021; O'Donnell et al., 2018).

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 et al. 2009 show 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 (Park and Pillow, 2020).

There are also limits to encoding flexibility (Weber et al., 2019; Benucci et al., 2013), 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. Furthermore, the adaptability may differ across the cortical hierarchy and levels of processing, i.e., the early visual system may adapt more slowly than WM. Indeed, a classic work addressing flexibility of encoding (Miller, 1956) showed participants could not automatically adapt to new situations. Miller found similar performance in remembering binary digits and decimal digits despite the severely reduced information load in having only 2 options rather than 10. Subsequent research showed people appear to encode both kind of digits phonologically (Jacquemot and Scott, 2006), rather than making use of the most efficient encoding strategy for remembering them only with respect to the possible options. Thus, although there do appear to be situations where people adapt their encoding to environmental statistics, the limits of how adaptive people can be in their encoding strategy, and the relative role of encoding vs. decoding in benefiting from environmental regularities, remain important open questions.

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 (Shin and Ma, 2017), broadly consistent with previous results from delayed estimation (Fougnie et al., 2010; Bays et al., 2011). In a delayed comparison task for a single object, performance slightly suffered when the number of relevant features dimensions was increased, but the decrease in performance was much smaller than would be expected if resources were fully shared across features (Palmer et al., 2015). Some change detection studies have also reported a modest decline of accuracy when features are added (Oberauer and Eichenberger, 2013; Hardman and Cowan, 2015). 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 (Shin and Ma, 2017). A separate indicator that resource pools for different features are not completely independent comes from experiments in which a retrospective cue indicates the feature dimension to be tested in a continuous report. Several studies have found a performance benefit from valid cues, and a cost of invalid ones, suggesting that resources can to some degree be shifted across feature dimensions (Ye et al., 2016; Park et al., 2017; Hajonides et al., 2020).

A second meaning is whether or not an irrelevant feature of a relevant object is automatically represented in WM. Several studies employing surprise tests with discrete report of previously irrelevant sample features observed near-chance performance (Chen and Wyble, 2016; Wyble et al., 2019), and decoding from fMRI or EEG data has shown little evidence for maintenance of task-irrelevant features (Yu and Shim, 2017; Bocincova and Johnson, 2019). However, surprise tests with a continuous report for a colour

or orientation showed evidence that irrelevant features were maintained, albeit only weakly (Shin and Ma, 2016; Swan et al., 2016). Stimulus location appears to take a special role in that it is robustly recalled even when task-irrelevant (Chen and Wyble, 2015; Kondo and Saiki, 2012; Foster et al., 2017; Cai et al., 2019), although with reduced precision (Tam and Wyble, 2022). The temporal order of sequentially presented stimuli may likewise be maintained automatically (Heuer and Rolfs, 2021). Irrelevant features have also been observed to produce inter-trial priming (Jiang et al., 2016) and to affect visual attention in a secondary visual search task (Hollingworth and Bahle, 2020; Harrison et al., 2021).

While task-irrelevant features appear at best to be relatively poorly represented at recall, 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 (Marshall and Bays, 2013). One possible explanation is 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. This is consistent with change localization performance for other features in the study of Shin and Ma (2017).

A third meaning is whether for a given feature dimension, resource "leaks away" to objects that are neutral in that feature (e.g., a circle is neutral for orientation) but that are task-relevant because of other features. In WM tasks in which $2N$ features were divided over either N or $2N$ objects, this was found to be the case both for orientation and colour (Shin and Ma, 2017; Fougnie et al., 2010). Two further studies indicate that to prevent this "leaking away" of resources, it is sufficient for different features to share the same location, even if they are not fully integrated into a smaller number of objects (Wang et al., 2016; Markov et al., 2019).

Theoretical proposals attempting to unify the different aspects of the feature/object question have included that of a hierarchically structured feature bundle (Brady et al., 2011) and of partially packaged resource (Shin and Ma, 2017). 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 (Bays et al., 2009; Huang, 2020a; Pratte, 2019).

Our understanding of this type of error has substantially improved in recent years. Swap frequency depends on the feature (or features) used as a cue (Rajsic and Wilson, 2014; Rajsic et al., 2017), and they occur most often between a target and a non-target item that are similar in their cue feature (Bays, 2016a; Emrich and Ferber, 2012; Rerko et al., 2014; Souza et al., 2014; Sahan et al., 2019). 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 (Wheeler and Treisman, 2002). 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

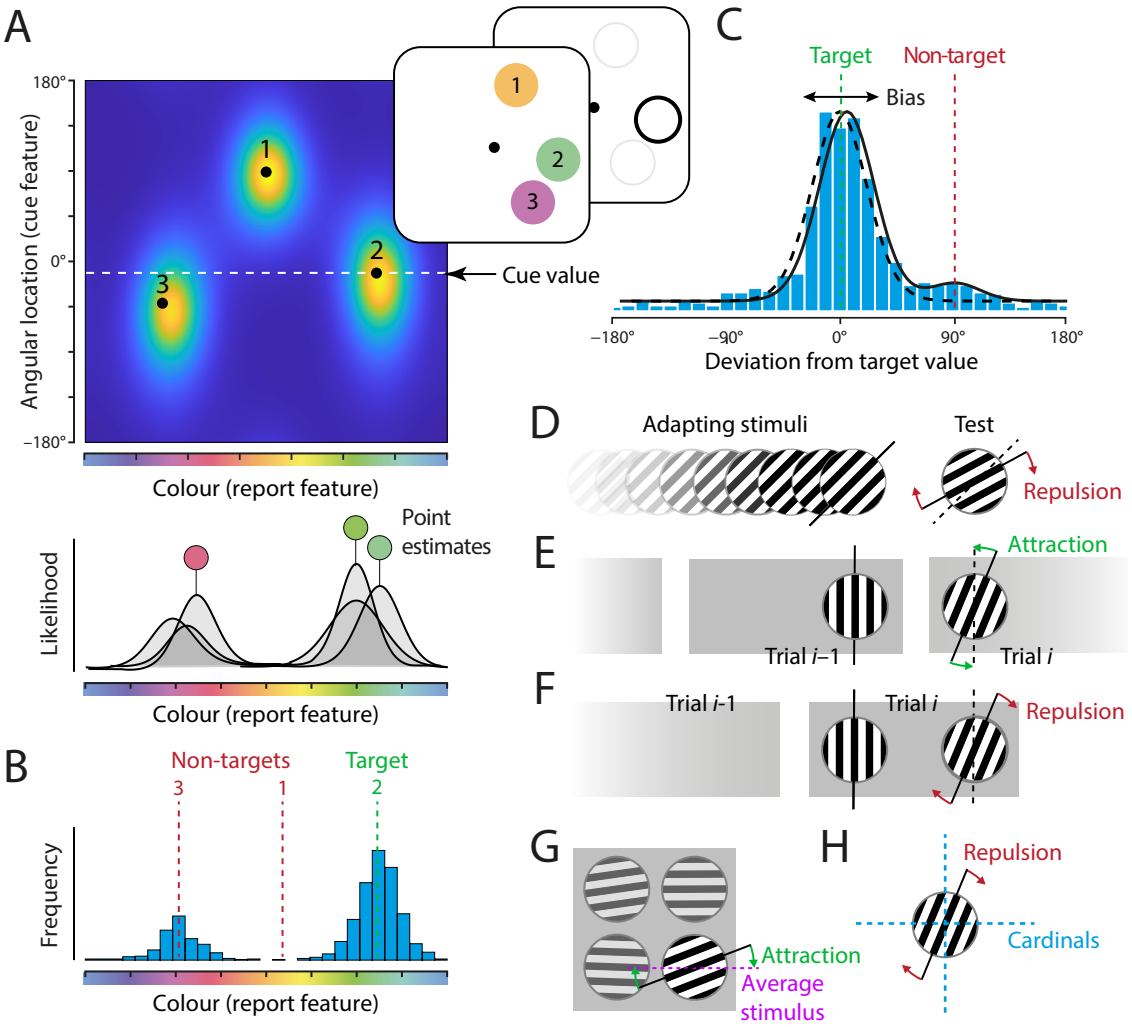

Figure 4: (A & B) Swap errors arising from cue feature similarity in a conjunctive coding model. (A) Example of a likelihood function over all possible combinations of cue and report feature value based on a fully conjunctive memory representation of a memory array (shown in inset, numbers for reference), 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 Golomb et al. (2014). (D–H) A diverse range of factors contributing to VWM biases.

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 (McMaster et al., 2022).

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 (Schneegans and Bays, 2017a; Swan and Wyble, 2014; Schneegans et al., 2015), or through rapidly formed synaptic connections between neurons sensitive for a single feature (Manohar et al., 2019; Oberauer and Lin, 2017). Some models additionally incorporate separate single-feature representations to allow a more efficient coding of memoranda (Matthey et al., 2015; Oberauer and Lin, 2017; Schneegans et al., 2015). Both approaches have been shown to fit the effects of cue similarity on swap errors in quantitative detail (Schneegans and Bays, 2017a; Oberauer and Lin, 2017), and both have also been used to explain error patterns in change detection tasks (Swan and Wyble, 2014; Lin and Oberauer, 2022). An interesting recent extension of a conjunctive coding model additionally describes feature binding across multiple levels of visual processing (Hedayati et al., 2021).

Among visual features, location has long been considered to have a special role in both perception and WM (Treisman and Zhang, 2006; Huang, 2020b). In addition to being encoded automatically even when task-irrelevant, location is a particularly effective retrieval cue (Rajsic et al., 2017), and spatial congruency between stimuli affects recall performance (Golomb et al., 2014; Teng and Postle, 2021). It has been 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 (Schneegans and Bays, 2017a). 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 (Bays et al., 2011; Fougnie and Alvarez, 2011; Kovacs and Harris, 2019; Markov et al., 2021; but see Sone et al., 2021 for an alternative account). More recent work further indicates that for sequentially presented stimuli, presentation time may take a similar role as location in binding visual features (Schneegans et al., 2021, 2022; Heuer and Rolfs, 2021).

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 (Read et al., 2016; Rhodes et al., 2017; Pertzov et al., 2015), nor with most other clinical conditions (Della Sala et al., 2012; Lugtmeijer et al., 2021). However, a specific binding impairment has been observed in association with Alzheimer's disease (Liang et al., 2016; Della Sala et al., 2012) and has been proposed as a diagnostic tool to differentiate Alzheimer's from other forms of dementia (Martínez et al., 2019).

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 et al. (2014) 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 (Fornaciai and Park, 2020; Czoschke et al., 2020).

Classical adaptation effects (Webster, 2015), 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" (Cicchini et al., 2021; Kiyonaga et al., 2017) 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 (Bliss et al., 2017; Barbosa and Compte, 2020; Fritsche et al., 2017). 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 (Bergen and Jehee, 2019; Fritsche et al., 2020; Cicchini et al., 2018); 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; Bae and Luck 2017; Czoschke et al. 2019; Kang and Choi 2015). 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 (Bae and Luck, 2017; Lively et al., 2021; Chunharas et al., 2022). 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 interitem confusion (Scotti et al., 2021). By contrast, when many items are held in mind, or when memories are weak for another reason (Dubé et al., 2014), items tend to be reported as more similar to each other than they really were (Fig 4G; Orhan and Jacobs, 2013; Brady and Alvarez, 2011; Lively et al., 2021; Chunharas et al., 2022; Papenmeier and Timm, 2021). 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; Sheth and Shimojo, 2001), and others that may arise at the decision stage (e.g. reference repulsion; Luu and Stocker, 2021). 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 (Rademaker et al., 2018; Schneegans and Bays, 2018). However, this effect is relatively subtle and variable (Shin et al., 2017) 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 (Compte et al., 2000; Johnson et al., 2009; Wei et al., 2012). Several memory decoding studies have observed gradual changes in encoded feature values over time that correlate with response errors, consistent with this theoretical account (Wimmer et al., 2014; Lim et al., 2019; Wolff et al., 2020). This account is further supported by behavioural results comparing response errors and latencies across different set size and delay conditions (Schneegans and Bays, 2018), and is consistent with findings from signal detection analyses of behavioural data indicating that deterioration of memory is driven by accumulation of internal noise (Kuuramo et al., 2022).

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 (Carroll et al., 2014; Kutschireiter et al., 2022). 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 (Orhan and Ma, 2019).

Deterioration of memory over time may also be driven by interference between multiple memory items (Pertzov et al., 2017). One proposed model explains this effect by a combination of sharing representational resources in an attractor model with efficient encoding (Koyluoglu et al., 2017). 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 (Bouchacourt and Buschman, 2019).

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 (Scotti et al., 2021; Chunharas et al., 2022). Such interactions also occur in continuous attractor models as a result of mutual excitation and inhibition between active sub-populations (Almeida et al., 2015; Johnson et al., 2009; Wei et al., 2012), 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 (Fuster and Alexander, 1971; Funahashi et al., 1989; Wimmer et al., 2014; Hart and Huk, 2020). Persistent activity has also been observed in rare electrophysiology studies in humans (Kamiński et al., 2017; Kornblith et al., 2017).

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 (Brouwer and Heeger, 2009; Ester et al., 2013). In this type of study, it has often been

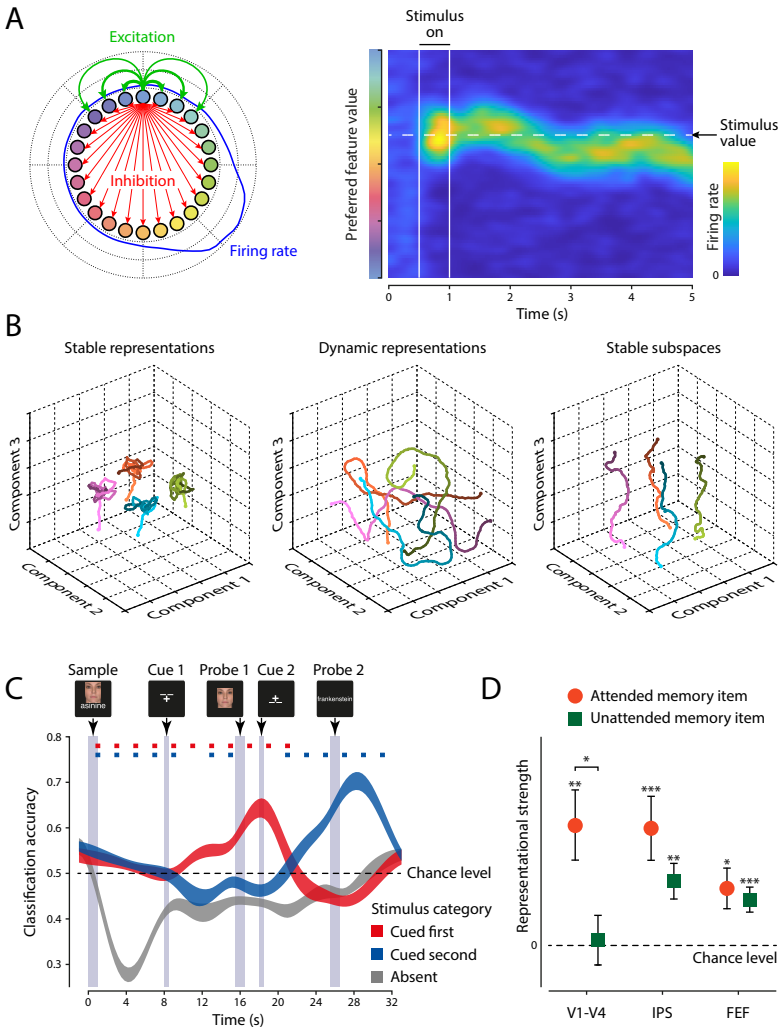

Figure 5: Neural mechanisms of visual WM. (A) Left: Architecture of a continuous attractor model of WM, with neurons shown as circles coloured with their preferred feature value, arranged on a ring reflecting the topology of the feature space. The pattern of synaptic connectivity is shown for one example 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 Rose et al., 2016). Decoding strength for the category of a currently uncued item transiently drops to chance level, suggestive of representation in an activity-silent state. (B) 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 Christophel et al., 2018).

found that there is little generalization in decoder efficacy between sample and delay period (Stokes et al., 2013; Wolff et al., 2017), or between different phases of the delay period (Sreenivasan et al., 2014; Meyers et al., 2008; Cavanagh et al., 2018). While changes in neural representations immediately after stimulus presentation may reflect transitions from perceptual and iconic memory (Coltheart, 1980) to WM, qualitative changes in representational format during maintenance are inconsistent with traditional conceptualizations of WM as implemented in atttractor models. This has lead to postulates that WM activity is substantially more dynamic than previously recognized (Stokes, 2015; Postle, 2015) (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 (Baeg et al., 2003; MacDonald et al., 2011; Scott et al., 2017; Stokes et al., 2013). 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 (Orhan and Ma, 2019).

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 (Murray et al., 2017; Parthasarathy et al., 2019; Spaak et al., 2017; Cueva et al., 2020). 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 (Wolff et al., 2020).

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 (Oostwoud Wijdenes et al., 2015; Wolf and Schütz, 2015; Ganmor et al., 2015) and that this process relies on the allocation of limited VWM resources to behaviourally relevant objects in advance of the eye movement (Kong et al., 2021; Stewart and Schütz, 2018). Multiple object representations can be integrated across a saccade, including objects that are never brought into foveal vision (Oostwoud Wijdenes et al., 2015; Stewart and Schütz, 2019); 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 (Ohl and Rolfs, 2016; Udale et al., 2022; Bays and Husain, 2008; Shao et al., 2010; Hanning et al., 2016).

WM has a broad role in supporting goal-directed movement (see Heuer et al., 2020; Chen and Crawford,

2020; Aagten-Murphy and Bays, 2018 for detailed reviews). Recent studies have demonstrated enhanced recall for visual items at locations relevant to reaching movements (Hanning and Deubel, 2018; Heuer et al., 2017) and also for feature dimensions relevant to a movement, e.g. object size for grasp (Heuer and Schubö, 2018). 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 (Chen and Crawford, 2020), 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 (Byrne and Crawford, 2010; Fiehler et al., 2014), 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 (Aagten-Murphy and Bays, 2019). 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.

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 very limited portion of the total memory content, typically just a single item, is represented through neural activity. This active memory is sometimes equated with the "focus of attention" in previous models (Oberauer, 2002; Lewis-Peacock et al., 2012). All other items are proposed to be held in an activity-silent state (Stokes, 2015) realized through mechanisms classically associated with long term memory, such as rapid synaptic plasticity or short-term changes in neural excitability (Mongillo et al., 2008; Barak and Tsodyks, 2014).

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 (LaRocque et al., 2013, 2017) 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 (Sprague et al., 2016), and transiently following a transcranial magnetic stimulation pulse (Rose et al., 2016) or a salient, but task-irrelevant visual stimulus Wolff et al. (2017). 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.

One important limitation in all of these studies is that their conclusions are based on null results, namely failures to decode certain stimuli from neural activity. Subsequent studies called the claims about activity-silent WM into question by successfully decoding the identity of unattended items (Christophel et al., 2018; Iamshchinina et al., 2021) (Figure 5D), even in data that had previously been used as support for

activity-silent states (Barbosa et al., 2021). Schneegans and Bays (2017b) 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 (Vogel and Machizawa, 2004; Luria et al., 2016), saturating at higher set sizes (Bays, 2018b), and memory load can also be estimated through classification methods applied to multivariate EEG (Adam et al., 2020) or fMRI data (Emrich et al., 2013). 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 et al. (2019) 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, a different interpretation of the findings supporting activity-silent memory proposes that they do not demonstrate a specific neural WM mechanism, but rather reflect contributions of classical long-term memory in WM tasks (Beukers et al., 2021; Foster et al., 2019).

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 (Stokes et al., 2020). Models that assume that only a single item can be in the focus of attention (Oberauer, 2002; Oberauer and Lin, 2017), giving it a privileged role in influencing visual attention, contrast with alternative conceptualizations in which the focus of attention can encompass multiple items (Cowan, 2011). This debate takes a more concrete form in the question of whether only one (Olivers et al., 2011; Ort et al., 2018) or multiple WM representations (Beck et al., 2012; Bahle et al., 2020) 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 (Ort et al., 2019). 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 (Williams et al., 2022a).

Another proposal is that WM is maintained by intermittent bursts of activity (Lundqvist et al., 2016, 2018b), bridged by mechanisms such as synaptic plasticity (Mongillo et al., 2008; Barak and Tsodyks, 2014). 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 (Shafi et al., 2007). The debate on the degree of persistence in neural firing during WM maintenance is still ongoing (Lundqvist et al., 2018a; Constantinidis et al., 2018). 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 (Pomper and Ansorge, 2021).

Anatomical localization of visual WM representations

The debate on the activity state of WM representations is closely linked to the question of their anatomical localization, although the latter has generally been studied without the possibility of activity-silent memory in mind. The first decoding studies had found that visual WM content could be decoded from early visual cortices (like V1 to V4), but not areas in parietal or prefrontal cortex (Serences et al., 2009; Harrison and Tong, 2009). These findings supported the idea that the same neural populations that are involved in sensory processing of visual features are also recruited for maintaining them in WM (Pasternak and

Greenlee, 2005). However, in subsequent studies memory content has been successfully decoded from a wider range of cortical areas, including parietal and prefrontal cortex (Ester et al., 2015; Christophel et al., 2018; Yu and Shim, 2017; Li et al., 2021), and the earlier results may primarily reflect the technical difficulties in decoding from cortical areas in which neurons with different selectivities are finely dispersed (Riley and Constantinidis, 2016).

It has been proposed that representations in extrasensory cortex serve to provide stability against interference from new sensory inputs (Bettencourt and Xu, 2016; Lorenc et al., 2018). In dual retro-cue paradigms, it has also been observed that the currently unattended item can be decoded predominantly from higher cortical areas (Christophel et al., 2018; Iamshchinina et al., 2021). Xu (2020) interpreted these results as evidence that extrasensory cortex (in particular, the intraparietal sulcus) is the true locus of WM representations, and attributes decoding from early visual cortex to feedback projections from these areas. Others argue in favor of distributed representations (Christophel et al., 2017; Lorenc and Sreenivasan, 2021), supported by findings that WM representations and new visual inputs can coexist in early visual cortex (Rademaker et al., 2019a), and that decoding precision in these areas correlates more strongly with behavioural performance than it does for higher cortical areas (Iamshchinina et al., 2021).

The role of prefrontal cortex is likewise debated, with some authors assigning it a central role in storing memory content (Riley and Constantinidis, 2016), while others argue that its primary function in WM tasks is behavioural control (D'Esposito and Postle, 2015). Results from decoding studies in humans indicate that the prefrontal cortex contains more categorical information about stimuli compared to earlier visual areas (Sreenivasan et al., 2014), and also represents meta-information about the role of different stimuli within a task (Olmos-Solis et al., 2021). A recent study using monkey electrophysiology found that the prefrontal cortex mediates selection of items in WM through mechanisms that are shared with visual attention (Panichello and Buschman, 2021), consistent with a role in behavioural control. However, the same study also found that stimulus features could be decoded from prefrontal cortex activity, with their representational format changing during the selection process. The findings are generally in line with the idea that neurons in prefrontal cortex show mixed selectivity for a wide range of stimulus and task features (Fusi et al., 2016). It should also be noted that the role of prefrontal cortex for WM may differ between humans and non-human primates, given that the latter typically undergo extensive training for the specific tasks they perform (Qi et al., 2011).

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 (Cohen et al., 2022; Bays, 2016b; Taylor and Bays, 2020).

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., Zhou et al., 2022)? What mechanisms allow sensory input to be segregated from existing WM representations, or integrated with it, according to behavioural requirements (e.g., Rademaker et al., 2019b)? Are errors in long-term memory representations fundamentally different from those in WM and perception (Miner et al., 2020), 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 (Draschkow et al., 2021; Kristjánsson and Draschkow, 2021; Issen and Knill, 2012), 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., Hedayati et al., 2022).

References

Aagten-Murphy, D. and Bays, P. M. (2018). Functions of Memory Across Saccadic Eye Movements. *Current Topics in Behavioral Neurosciences*.

Aagten-Murphy, D. and Bays, P. M. (2019). Independent working memory resources for egocentric and allocentric spatial information. *PLoS computational biology*, 15(2):e1006563. Publisher: Public Library of Science.

Acerbi, L., Vijayakumar, S., and Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS computational biology*, 10(6):e1003661.

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

Adam, K. C. S., Vogel, E. K., and Awh, E. (2020). Multivariate analysis reveals a generalizable human electrophysiological signature of working memory load. *Psychophysiology*, 57(12).

Almeida, R., Barbosa, J., and Compte, A. (2015). Neural circuit basis of visuo-spatial working memory precision: A computational and behavioral study. *Journal of Neurophysiology*, 114(3):1806–1818.

Alvarez, G. A. and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science*, 15(2):106–111.

Asp, I. E., Störmer, V. S., and Brady, T. F. (2021). Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful. *Journal of cognitive neuroscience*, 33(5):902–918.

Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145.

Awh, E., Barton, B., and Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science*, 18(7):622–628.

Bae, G.-Y. and Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, & Psychophysics*, 79(8):2376–2395.

Bae, G.-Y., Olkkonen, M., Allred, S. R., and Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144(4):744.

Baeg, E., Kim, Y., Huh, K., Mook-Jung, I., Kim, H., and Jung, M. (2003). Dynamics of Population Code for Working Memory in the Prefrontal Cortex. *Neuron*, 40(1):177–188.

Bahle, B., Thayer, D. D., Mordkoff, J. T., and Hollingworth, A. (2020). The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *Journal of Experimental Psychology: General*, 149(5):967–983.

Barak, O. and Tsodyks, M. (2014). Working models of working memory. *Current Opinion in Neurobiology*, 25:20–24.

Barbosa, J. and Compte, A. (2020). Build-up of serial dependence in color working memory. *Scientific Reports*, 10(1):10959. Number: 1 Publisher: Nature Publishing Group.

Barbosa, J., Lozano-Soldevilla, D., and Compte, A. (2021). Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. *PLOS Biology*, 19(10):e3001436.

Bates, C. J. and Jacobs, R. A. (2020). Efficient data compression in perception and perceptual memory. *Psychological review*, 127(5):891.

Bays, P. M. (2014). Noise in Neural Populations Accounts for Errors in Working Memory. *Journal of Neuroscience*, 34(10):3632–3645.

Bays, P. M. (2016a). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, 6(1):19203.

Bays, P. M. (2016b). A signature of neural coding at human perceptual limits. *Journal of Vision*, 16(11):4.

Bays, P. M. (2018a). Failure of self-consistency in the discrete resource model of visual working memory. *Cognitive Psychology*, 105:1–8.

Bays, P. M. (2018b). Reassessing the Evidence for Capacity Limits in Neural Signals Related to Working Memory. *Cerebral Cortex*, 28(4):1432–1438.

Bays, P. M., Catalao, R. F. G., and Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7–7.

Bays, P. M. and Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, 321(5890):851–854.

Bays, P. M., Wu, E. Y., and Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49(6):1622–1631.

Beck, V. M., Hollingworth, A., and Luck, S. J. (2012). Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychological Science*, 23(8):887–898.

Bengson, J. J. and Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin & Review*, 23(1):265–270.

Benucci, A., Saleem, A. B., and Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature neuroscience*, 16(6):724–729.

Bergen, R. S. v. and Jehee, J. F. M. (2019). Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *Journal of Neuroscience*, 39(41):8164–8176. Publisher: Society for Neuroscience Section: Research Articles.

Bettencourt, K. C. and Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, 19(1):150–157.

Beukers, A. O., Buschman, T. J., Cohen, J. D., and Norman, K. A. (2021). Is Activity Silent Working Memory Simply Episodic Memory? *Trends in Cognitive Sciences*, 25(4):284–293.

Bliss, D. P., Sun, J. J., and D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, 7(1):14739. tex.ids= BlissEtAl2017a number: 1 publisher: Nature Publishing Group.

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

Bouchacourt, F. and Buschman, T. J. (2019). A Flexible Model of Working Memory. *Neuron*, 103(1):147–160.e8.

Brady, T. F. and Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, 22(3):384–392.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4):487.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, 11(5):4–4.

Brady, T. F. and Störmer, V. S. (2021). The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Brady, T. F. and Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, 120(1):85–109.

Brewer, W. F. and Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive psychology*, 13(2):207–230.

Brissenden, J. A., Adkins, T. J., Hsu, Y. T., and Lee, T. G. (2021). Reward influences the allocation but not the availability of resources in visual working memory. *bioRxiv*.

Brouwer, G. J. and Heeger, D. J. (2009). Decoding and Reconstructing Color from Responses in Human Visual Cortex. *Journal of Neuroscience*, 29(44):13992–14003.

Byrne, P. A. and Crawford, J. D. (2010). Cue Reliability and a Landmark Stability Heuristic Determine Relative Weighting Between Egocentric and Allocentric Visual Information in Memory-Guided Reach. *Journal of Neurophysiology*, 103(6):3054–3069.

Cai, Y., Sheldon, A. D., Yu, Q., and Postle, B. R. (2019). Overlapping and distinct contributions of stimulus location and of spatial context to nonspatial visual short-term memory. *Journal of Neurophysiology*, 121(4):1222–1231.

Carroll, S., Josić, K., and Kilpatrick, Z. P. (2014). Encoding certainty in bump attractors. *Journal of computational neuroscience*, 37(1):29–48.

Cavanagh, S. E., Towers, J. P., Wallis, J. D., Hunt, L. T., and Kennerley, S. W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nature Communications*, 9(1):3498.

Chen, H. and Wyble, B. (2015). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, 107:76–85.

Chen, H. and Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2):225–234.

Chen, Y. and Crawford, J. D. (2020). Allocentric representations for target memory and reaching in human cortex. *Annals of the New York Academy of Sciences*, 1464(1):142–155. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.14261.

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., and Haynes, J.-D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, 21(4):494–496.

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., and Haynes, J.-D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, 21(2):111–124.

Chunharas, C., Rademaker, R. L., Brady, T. F., and Serences, J. T. (2022). An adaptive perspective on visual working memory distortions. *Journal of Experimental Psychology: General*.

Cicchini, G. M., Benedetto, A., and Burr, D. C. (2021). Perceptual history propagates down to early levels

of sensory analysis. *Current Biology*, 31(6):1245–1250.e2.
Cicchini, G. M., Mikellidou, K., and Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B*, 285(1890):20181722. Publisher: The Royal Society.
Cohen, M., Keefe, J. M., and Brady, T. (2022). Perceptual awareness occurs along a graded continuum: Evidence from psychophysical scaling.
Collins, A. G. (2018). The tortoise and the hare: Interactions between reinforcement learning and working memory. *Journal of cognitive neuroscience*, 30(10):1422–1432.
Collins, A. G. and Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7):1024–1035.
Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, 27(3):183–228.
Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X.-J. (2000). Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model. *Cerebral Cortex*, 10(9):910–923.
Constantinidis, C., Funahashi, S., Lee, D., Murray, J. D., Qi, X.-L., Wang, M., and Arnsten, A. F. (2018). Persistent Spiking Activity Underlies Working Memory. *The Journal of Neuroscience*, 38(32):7020–7028.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1):87–114.
Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia*, 49(6):1401–1406.
Cueva, C. J., Saez, A., Marcos, E., Genovesio, A., Jazayeri, M., Romo, R., Salzman, C. D., Shadlen, M. N., and Fusi, S. (2020). Low-dimensional dynamics for working memory and time encoding. *Proceedings of the National Academy of Sciences*, 117(37):23021–23032.
Czoschke, S., Fischer, C., Beitner, J., Kaiser, J., and Bledowski, C. (2019). Two types of serial dependence in visual working memory. *British Journal of Psychology*, 110(2):256–267. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12349.
Czoschke, S., Peters, B., Rahm, B., Kaiser, J., and Bledowski, C. (2020). Visual objects interact differently during encoding and memory maintenance. *Attention, Perception, & Psychophysics*, 82(3):1241–1257.
Della Sala, S., Parra, M. A., Fabi, K., Luzzi, S., and Abrahams, S. (2012). Short-term memory binding is impaired in AD but not in non-AD dementias. *Neuropsychologia*, 50(5):833–840.
D'Esposito, M. and Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, 66(1):115–142.
Devkar, D., Wright, A. A., and Ma, W. J. (2017). Monkeys and humans take local uncertainty into account when localizing a change. *Journal of Vision*, 17(11):4–4.
Devkar, D. T., Wright, A. A., and Ma, W. J. (2015). The same type of visual working memory limitations in humans and monkeys. *Journal of vision*, 15(16):13–13.
Draschkow, D., Kallmayer, M., and Nobre, A. C. (2021). When Natural Behavior Engages Working Memory. *Current Biology*, 31(4):869–874.e5.
Dubé, C., Zhou, F., Kahana, M. J., and Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, 96:8–16.
Emrich, S. M. and Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision*, 12(4):12–12.
Emrich, S. M., Lockhart, H. A., and Al-Aidroos, N. (2017). Attention Mediates the Flexible Allocation of Visual Working Memory Resources. *Journal of Experimental Psychology. Human Perception and Performance*.

Emrich, S. M., Riggall, A. C., LaRocque, J. J., and Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *Journal of Neuroscience*, 33(15):6516–6523.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science*, 11(1):19–23.

Ester, E. F., Anderson, D. E., Serences, J. T., and Awh, E. (2013). A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience*, 25(5):754–761.

Ester, E. F., Sprague, T. C., and Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, 87(4):893–905.

Fiehler, K., Wolf, C., Klinghammer, M., and Blohm, G. (2014). Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Frontiers in Human Neuroscience*, 8.

Fleming, S. M. and Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological review*, 124(1):91.

Fornaciai, M. and Park, J. (2020). Attractive serial dependence between memorized stimuli. *Cognition*, 200:104250. Publisher: Elsevier.

Foster, J. J., Bsales, E. M., Jaffe, R. J., and Awh, E. (2017). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, 27(20):3216–3223.e6.

Foster, J. J., Vogel, E. K., and Awh, E. (2019). Working memory as persistent neural activity. Preprint, PsyArXiv.

Fougnie, D. and Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, 11(12):3–3.

Fougnie, D., Asplund, C. L., and Marois, R. (2010). What are the units of storage in visual working memory? *Journal of vision*, 10(12):27–27.

Fougnie, D., Suchow, J. W., and Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature communications*, 3(1):1–8.

Fritsche, M., Mostert, P., and de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, 27(4):590–595.

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

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349.

Fusi, S., Miller, E. K., and Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37:66–74.

Fuster, J. M. and Alexander, G. E. (1971). Neuron Activity Related to Short-Term Memory. *Science*, 173(3997):652–654.

Ganguli, D. and Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural computation*.

Ganmor, E., Landy, M. S., and Simoncelli, E. P. (2015). Near-optimal integration of orientation information across saccades. *Journal of Vision*, 15(16):8–8.

Girshick, A. R., Landy, M. S., and Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7):926–932.

Golomb, J. D., Kupitz, C. N., and Thiemann, C. T. (2014). The influence of object location on identity: A "spatial congruency bias". *Journal of Experimental Psychology: General*, 143(6):2262–2278.

Gong, M. and Li, S. (2014). Learned reward association improves visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2):841.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., and Husain, M. (2011). Dynamic Updating of Working Memory Resources for Visual Objects. *Journal of Neuroscience*, 31(23):8502–8511.

Hajonides, J. E., van Ede, F., Stokes, M. G., and Nobre, A. C. (2020). Comparing the prioritization of items and feature-dimensions in visual working memory. *Journal of Vision*, 20(8):25.

Hanning, N. M. and Deubel, H. (2018). Independent Effects of Eye and Hand Movements on Visual Working Memory. *Frontiers in Systems Neuroscience*, 12.

Hanning, N. M., Jonikaitis, D., Deubel, H., and Szinte, M. (2016). Oculomotor selection underlies feature retention in visual working memory. *Journal of Neurophysiology*, 115(2):1071–1076. Publisher: American Physiological Society.

Hardman, K. O. and Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2):325.

Hardman, K. O., Vergauwe, E., and Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1):30.

Harrison, G. W., Kang, M., and Wilson, D. E. (2021). Remembering more than you can say: Re-examining "amnesia" of attended attributes. *Acta Psychologica*, 214:103265.

Harrison, S. A. and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238):632–635.

Hart, E. and Huk, A. C. (2020). Recurrent circuit dynamics underlie persistent activity in the macaque frontoparietal network. *eLife*, 9:e52460.

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

Hedayati, S., O'Donnell, R. E., and Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour*, 6(5):709–719.

Heuer, A., Crawford, J. D., and Schubö, A. (2017). Action relevance induces an attentional weighting of representations in visual working memory. *Memory & Cognition*, 45(3):413–427.

Heuer, A., Ohl, S., and Rolfs, M. (2020). Memory for action: a functional view of selection in visual working memory. *Visual Cognition*, 28(5-8):388–400. Publisher: Routledge _eprint: https://doi.org/10.1080/13506285.2020.1764156.

Heuer, A. and Rolfs, M. (2021). Incidental encoding of visual information in temporal reference frames in working memory. *Cognition*, 207:104526.

Heuer, A. and Schubö, A. (2018). Separate and combined effects of action relevance and motivational value on visual working memory. *Journal of Vision*, 18(5):14.

Hollingworth, A. and Bahle, B. (2020). Feature-based guidance of attention by visual working memory is applied independently of remembered object location. *Attention, Perception, & Psychophysics*, 82(1):98–108.

Honig, M., Ma, W. J., and Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences*, 117(15):8391–8397.

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

Huang, L. (2020a). Distinguishing target biases and strategic guesses in visual working memory. *Attention, Perception, & Psychophysics*, 82(3):1258–1270.

Huang, L. (2020b). Unit of visual working memory: A Boolean map provides a better account than an

object does. *Journal of Experimental Psychology: General*, 149(1):1–30.
Huttenlocher, J., Hedges, L. V., Corrigan, B., and Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition*, 93(2):75–97.
Iamshchinina, P., Christophel, T. B., Gayet, S., and Rademaker, R. L. (2021). Essential considerations for exploring visual working memory storage in the human brain. *Visual Cognition*, 29(7):425–436.
Issen, L. A. and Knill, D. C. (2012). Decoupling eye and hand movement control: Visual short-term memory influences reach planning more than saccade planning. *Journal of Vision*, 12(1).
Jabar, S. B., Sreenivasan, K. K., Lentzou, S., Kanabar, A., Brady, T. F., and Fougnie, D. (2020). Using a betting game to reveal the rich nature of visual working memories. *bioRxiv*.
Jacquemot, C. and Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in cognitive sciences*, 10(11):480–486.
Jazayeri, M. and Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature neuroscience*, 9(5):690–696.
Jiang, Y. V., Shupe, J. M., Swallow, K. M., and Tan, D. H. (2016). Memory for recently accessed visual attributes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(8):1331–1337.
Johnson, J. S., Spencer, J. P., Luck, S. J., and Schöner, G. (2009). A Dynamic Neural Field Model of Visual Working Memory and Change Detection. *Psychological Science*, 20(5):568–577.
Kaiser, D., Stein, T., and Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*, 22(6):1784–1790.
Kamiński, J., Sullivan, S., Chung, J. M., Ross, I. B., Mamelak, A. N., and Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. *Nature Neuroscience*, 20(4):590–601.
Kang, M.-S. and Choi, J. (2015). Retrieval-induced inhibition in short-term memory. *Psychological Science*, 26(7):1014–1025. Publisher: Sage Publications Sage CA: Los Angeles, CA.
Keshvari, S., Van den Berg, R., and Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS One*, 7(6):e40216.
Keshvari, S., Van den Berg, R., and Ma, W. J. (2013). No evidence for an item limit in change detection. *PLoS computational biology*, 9(2):e1002927.
Kiyonaga, A., Scimeca, J. M., Bliss, D. P., and Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. *Trends in Cognitive Sciences*, 21(7):493–497.
Klyszejko, Z., Rahmati, M., and Curtis, C. E. (2014). Attentional priority determines working memory precision. *Vision research*, 105:70–76.
Knill, D. C. and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719.
Kondo, A. and Saiki, J. (2012). Feature-Specific Encoding Flexibility in Visual Working Memory. *PLoS ONE*, 7(12):e50962.
Kong, G., Kroell, L. M., Schneegans, S., Aagten-Murphy, D., and Bays, P. M. (2021). Transsaccadic integration relies on a limited memory resource. *Journal of Vision*, 21(5):24–24. Publisher: The Association for Research in Vision and Ophthalmology.
Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I., and Mormann, F. (2017). Persistent Single-Neuron Activity during Working Memory in the Human Medial Temporal Lobe. *Current Biology*, 27(7):1026–1032.
Kovacs, O. and Harris, I. M. (2019). The role of location in visual feature binding. *Attention, Perception, & Psychophysics*, 81(5):1551–1563.
Koyluoglu, O. O., Pertzov, Y., Manohar, S., Husain, M., and Fiete, I. R. (2017). Fundamental bound on the persistence and capacity of short-term memory stored as graded persistent activity. *eLife*, 6:e22225.

Kriegeskorte, N. and Wei, X.-X. (2021). Neural tuning and representational geometry. *Nature Reviews Neuroscience*, pages 1–16. Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Neural decoding;Neural encoding Subject_term_id: neural-decoding;neural-encoding.

Kristjánsson, Á. and Draschkow, D. (2021). Keeping it real: Looking beyond capacity limits in visual cognition. *Attention, Perception, & Psychophysics*, 83(4):1375–1390.

Kutschireiter, A., Basnak, M. A., and Drugowitsch, J. (2022). Bayesian inference in ring attractor networks.

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

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. (2013). Decoding Attended Information in Short-term Memory: An EEG Study. *Journal of Cognitive Neuroscience*, 25(1):127–142.

LaRocque, J. J., Riggall, A. C., Emrich, S. M., and Postle, B. R. (2017). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex*, page cercor;bhw283v1.

Lennie, P. (2003). The cost of cortical computation. *Current biology*, 13(6):493–497.

Lew, T. F. and Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, 15(4):10.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., and Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, 24(1):61–79.

Li, J., Huang, Q., Han, Q., Mi, Y., and Luo, H. (2021). Temporally coherent perturbation of neural dynamics during retention alters human multi-item working memory. *Progress in Neurobiology*, 201:102023.

Liang, Y., Pertzov, Y., Nicholas, J. M., Henley, S. M., Crutch, S., Woodward, F., Leung, K., Fox, N. C., and Husain, M. (2016). Visual short-term memory binding deficit in familial Alzheimer's disease. *Cortex*, 78:150–164.

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

Lim, P. C., Ward, E. J., Vickery, T. J., and Johnson, M. R. (2019). Not-so-working Memory: Drift in Functional Magnetic Resonance Imaging Pattern Representations during Maintenance Predicts Errors in a Visual Working Memory Task. *Journal of Cognitive Neuroscience*, 31(10):1520–1534.

Lin, H.-Y. and Oberauer, K. (2022). An interference model for visual working memory: Applications to the change detection task. *Cognitive Psychology*, 133:101463.

Lively, Z., Robinson, M. M., and Benjamin, A. S. (2021). Memory Fidelity Reveals Qualitative Changes in Interactions Between Items in Visual Working Memory. *Psychological Science*, 32(9):1426–1441.

Lorenc, E. S. and Sreenivasan, K. K. (2021). Reframing the debate: The distributed systems view of working memory. *Visual Cognition*, 29(7):416–424.

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R., and D'Esposito, M. (2018). Flexible Coding of Visual Working Memory Representations during Distraction. *The Journal of Neuroscience*, 38(23):5267–5276.

Luck, S. J. and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281.

Lugtmeijer, S., Schneegans, S., Lammers, N. A., Geerligs, L., de Leeuw, F. E., de Haan, E. H., Bays, P. M., and Kessels, R. P. (2021). Consequence of stroke for feature recall and binding in visual working

memory. *Neurobiology of Learning and Memory*, 179:107387.
Lundqvist, M., Herman, P., and Miller, E. K. (2018a). Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *The Journal of Neuroscience*, 38(32):7013–7019.
Lundqvist, M., Herman, P., Warden, M. R., Brincat, S. L., and Miller, E. K. (2018b). Gamma and beta bursts during working memory readout suggest roles in its volitional control. *Nature Communications*, 9(1):394.
Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., and Miller, E. K. (2016). Gamma and Beta Bursts Underlie Working Memory. *Neuron*, 90(1):152–164.
Luria, R., Balaban, H., Awh, E., and Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, 62:100–108.
Luu, L. and Stocker, A. A. (2021). Categorical judgments do not modify sensory representations in working memory. *PLOS Computational Biology*, 17(6):e1008968.
Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–1438.
Ma, W. J. and Huang, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision*, 9(11):3–3.
MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. *Neuron*, 71(4):737–749.
Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., and Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience & Biobehavioral Reviews*, 101:1–12.
Markov, Y. A., Tiurina, N. A., and Utochkin, I. S. (2019). Different features are stored independently in visual working memory but mediated by object-based representations. *Acta Psychologica*, 197:52–63.
Markov, Y. A., Utochkin, I. S., and Brady, T. F. (2021). Real-world objects are not stored in holistic representations in visual working memory. *Journal of Vision*, 21(3):18.
Marshall, L. and Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, 13(2):21–21.
Martínez, J. F., Trujillo, C., Arévalo, A., Ibáñez, A., and Cardona, J. F. (2019). Assessment of Conjunctive Binding in Aging: A Promising Approach for Alzheimer's Disease Detection. *Journal of Alzheimer's Disease*, 69(1):71–81.
Mathy, F. and Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, 122(3):346–362.
Matthey, L., Bays, P. M., and Dayan, P. (2015). A Probabilistic Palimpsest Model of Visual Short-term Memory. *PLOS Computational Biology*, 11(1):e1004003.
Mazyar, H., Van den Berg, R., and Ma, W. J. (2012). Does precision decrease with set size? *Journal of vision*, 12(6):10–10.
McMaster, J. M. V., Tomić, I., Schneegans, S., and Bays, P. M. (2022). Swap errors in visual working memory are fully explained by cue-feature variability. *Cognitive Psychology*.
Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., and Poggio, T. (2008). Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *Journal of Neurophysiology*, 100(3):1407–1419.
Meyniel, F., Sigman, M., and Mainen, Z. F. (2015). Confidence as bayesian probability: From neural origins to behavior. *Neuron*, 88(1):78–92.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.
Miner, A. E., Schurgin, M. W., and Brady, T. F. (2020). Is working memory inherently more "precise" than long-term memory? extremely high fidelity visual long-term memories for frequently encountered

objects. *Journal of Experimental Psychology: Human Perception and Performance*, 46(8):813.
Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic Theory of Working Memory. *Science*, 319(5869):1543–1546.
Morais, M. and Pillow, J. W. (2018). Power-law efficient neural codes provide general link between perceptual bias and discriminability. *Advances in Neural Information Processing Systems*, 31.
Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., and Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 114(2):394–399.
Mystakidou, M. and van den Berg, R. (2020). More motivated but equally good: No effect of gamification on visual working memory performance.
Ngiam, W. X., Brissenden, J. A., and Awh, E. (2019). "memory compression" effects in visual working memory are contingent on explicit long-term memory. *Journal of Experimental Psychology: General*, 148(8):1373.
Norris, D., Kalm, K., and Hall, J. (2020). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(5):872.
Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):411–421.
Oberauer, K. and Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & cognition*, 41(8):1212–1227.
Oberauer, K. and Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1):21–59.
O'Donnell, R. E., Clement, A., and Brockmole, J. R. (2018). Semantic and functional relationships among objects increase the capacity of visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(7):1151.
Ohl, S. and Rolfs, M. (2016). Saccadic Eye Movements Impose a Natural Bottleneck on Visual Short-Term Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*.
Ohshiro, T., Angelaki, D. E., and DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, 14(6):775–782.
Olivers, C. N., Peters, J., Houtkamp, R., and Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, page S1364661311000854.
Olmos-Solis, K., van Loon, A. M., and Olivers, C. N. (2021). Content or status: Frontal and posterior cortical representations of object category and upcoming task goals in working memory. *Cortex*, 135:61–77.
Oostwoud Wijdenes, L., Marshall, L., and Bays, P. M. (2015). Evidence for Optimal Integration of Visual Feature Representations across Saccades. *Journal of Neuroscience*, 35(28):10146–10153.
Orhan, A. E. and Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological review*, 120(2):297.
Orhan, A. E. and Ma, W. J. (2019). A diverse range of factors affect the nature of neural representations underlying short-term memory. *Nature Neuroscience*, 22(2):275–283.
Orhan, A. E., Sims, C. R., Jacobs, R. A., and Knill, D. C. (2014). The adaptive nature of visual working memory. *Current directions in psychological science*, 23(3):164–170.
Ort, E., Fahrenfort, J. J., and Olivers, C. N. L. (2018). Lack of free choice reveals the cost of multiple-target search within and across feature dimensions. *Attention, Perception, & Psychophysics*, 80(8):1904–1917.
Ort, E., Fahrenfort, J. J., ten Cate, T., Eimer, M., and Olivers, C. N. (2019). Humans can efficiently look for but not select multiple visual objects. *eLife*, 8:e49130.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16(2):332–350.

Palmer, J., Boston, B., and Moore, C. M. (2015). Limited capacity for memory tasks with multiple features within a single object. *Attention, Perception, & Psychophysics*, 77(5):1488–1499.

Panichello, M. F. and Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, 592(7855):601–605.

Papenmeier, F. and Timm, J. D. (2021). Do group ensemble statistics bias visual working memory for individual items? A registered replication of Brady and Alvarez (2011). *Attention, Perception, & Psychophysics*, 83(3):1329–1336.

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

Park, Y. E., Sy, J. L., Hong, S. W., and Tong, F. (2017). Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychological Science*, 28(12):1773–1785.

Parthasarathy, A., Tang, C., Herikstad, R., Cheong, L. F., Yen, S.-C., and Libedinsky, C. (2019). Time-invariant working memory representations in the presence of code-morphing in the lateral prefrontal cortex. *Nature Communications*, 10(1):4995.

Pasternak, T. and Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2):97–107.

Pertzov, Y., Heider, M., Liang, Y., and Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging*, 30(1):26–35.

Pertzov, Y., Manohar, S., and Husain, M. (2017). Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4):528–536.

Pomper, U. and Ansorge, U. (2021). Theta-Rhythmic Oscillation of Working Memory Performance. *Psychological Science*, 32(11):1801–1810.

Postle, B. R. (2015). Neural Bases of the Short-term Retention of Visual Information. In *Mechanisms of Sensory Working Memory*, pages 43–58. Elsevier.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–32.

Pratte, M. S. (2019). Swap errors in spatial working memory are guesses. *Psychonomic Bulletin & Review*, 26(3):958–966.

Pratte, M. S. (2020). Set size effects on working memory precision are not due to an averaging of slots. *Attention, Perception, & Psychophysics*, 82(6):2937–2949.

Qi, X.-L., Meyer, T., Stanford, T. R., and Constantinidis, C. (2011). Changes in Prefrontal Neuronal Activity after Learning to Perform a Spatial Working Memory Task. *Cerebral Cortex*, 21(12):2722–2732.

Rademaker, R. L., Chunharas, C., and Serences, J. T. (2019a). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, 22(8):1336–1344.

Rademaker, R. L., Chunharas, C., and Serences, J. T. (2019b). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature neuroscience*, 22(8):1336–1344.

Rademaker, R. L., Park, Y. E., Sack, A. T., and Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 44(6):925–940.

Rademaker, R. L., Tredway, C. H., and Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of vision*, 12(13):21–21.

Rajsic, J., Sun, S. Z., Huxtable, L., Pratt, J., and Ferber, S. (2016). Pop-out and pop-in: Visual working memory advantages for unique items. *Psychonomic Bulletin & Review*, 23(6):1787–1793.

Rajsic, J., Swan, G., Wilson, D. E., and Pratt, J. (2017). Accessibility limits recall from visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(9):1415–1431.

Rajsic, J. and Wilson, D. E. (2014). Asymmetrical access to color and location in visual working memory. *Attention, Perception, & Psychophysics*, 76(7):1902–1913.

Read, C. A., Rogers, J. M., and Wilson, P. H. (2016). Working memory binding of visual object features in older adults. *Aging, Neuropsychology, and Cognition*, 23(3):263–281.

Rerko, L., Oberauer, K., and Lin, H.-Y. (2014). Spatial Transposition Gradients in Visual Working Memory. *Quarterly Journal of Experimental Psychology*, 67(1):3–15.

Reynolds, J. H. and Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, 61(2):168–185.

Rhodes, S., Parra, M. A., Cowan, N., and Logie, R. H. (2017). Healthy aging and visual working memory: The effect of mixing feature and conjunction changes. *Psychology and Aging*, 32(4):354–366.

Riley, M. R. and Constantinidis, C. (2016). Role of Prefrontal Persistent Activity in Working Memory. *Frontiers in Systems Neuroscience*, 9.

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., and Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 354(6316):1136–1139.

Sahan, M. I., Dalmaijer, E. S., Verguts, T., Husain, M., and Fias, W. (2019). The Graded Fate of Unattended Stimulus Representations in Visuospatial Working Memory. *Frontiers in Psychology*, 10:374.

Schneegans, S. and Bays, P. M. (2017a). Neural Architecture for Feature Binding in Visual Working Memory. *The Journal of Neuroscience*, 37(14):3913–3925.

Schneegans, S. and Bays, P. M. (2017b). Restoration of fMRI Decodability Does Not Imply Latent Working Memory States. *Journal of Cognitive Neuroscience*, 29(12):1977–1994.

Schneegans, S. and Bays, P. M. (2018). Drift in Neural Population Activity Causes Working Memory to Deteriorate Over Time. *The Journal of Neuroscience*, 38(21):4859–4869.

Schneegans, S., Harrison, W. J., and Bays, P. M. (2021). Location-independent feature binding in visual working memory for sequentially presented objects. *Attention, Perception, & Psychophysics*, 83(6):2377–2393.

Schneegans, S., McMaster, J. M. V., and Bays, P. M. (2022). Role of time in binding features in visual working memory. *Psychological Review*.

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

Schneegans, S., Taylor, R., and Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences*.

Schurgin, M. W., Wixted, J. T., and Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, pages 1–17.

Scott, B. B., Constantinople, C. M., Akrami, A., Hanks, T. D., Brody, C. D., and Tank, D. W. (2017). Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron*, 95(2):385–398.e5.

Scotti, P. S., Hong, Y., Golomb, J. D., and Leber, A. B. (2021). Statistical learning as a reference point for memory distortions: Swap and shift errors. *Attention, Perception, & Psychophysics*, 83(4):1652–1672.

Serences, J. T., Ester, E. F., Vogel, E. K., and Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, 20(2):207–214.

Sewell, D. K., Lilburn, S. D., and Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of experimental psychology: human perception and performance*, 40(6):2214.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. (2007). Variability in neuronal

activity in primate cortex during working memory tasks. *Neuroscience*, 146(3):1082–1108.
Shao, N., Li, J., Shui, R., Zheng, X., Lu, J., and Shen, M. (2010). Saccades elicit obligatory allocation of visual working memory. *Memory & Cognition*, 38(5):629–640. Publisher: Springer.
Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. *Attention and performance VIII*, 8:277–295.
Sheth, B. R. and Shimojo, S. (2001). Compression of space in visual memory. *Vision research*, 41(3):329–341.
Shin, H. and Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5):10.
Shin, H. and Ma, W. J. (2017). Visual short-term memory for oriented, colored objects. *Journal of Vision*, 17(9):12.
Shin, H., Zou, Q., and Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, 17(14):10.
Simon, H. A. (1974). How big is a chunk? by combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, 183(4124):482–488.
Sims, C. R., Jacobs, R. A., and Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological review*, 119(4):807.
Sone, H., Kang, M.-S., Li, A. Y., Tsubomi, H., and Fukuda, K. (2021). Simultaneous estimation procedure reveals the object-based, but not space-based, dependence of visual working memory representations. *Cognition*, 209:104579.
Souza, A. S., Rerko, L., Lin, H.-Y., and Oberauer, K. (2014). Focused attention improves working memory: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics*, 76(7):2080–2102.
Spaak, E., Watanabe, K., Funahashi, S., and Stokes, M. G. (2017). Stable and Dynamic Coding for Working Memory in Primate Prefrontal Cortex. *The Journal of Neuroscience*, 37(27):6503–6516.
Sprague, T. C., Ester, E. F., and Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron*, 91(3):694–707.
Sreenivasan, K. K., Vytlacil, J., and D'Esposito, M. (2014). Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex. *Journal of Cognitive Neuroscience*, 26(5):1141–1153.
Starr, A., Srinivasan, M., and Bunge, S. A. (2020). Semantic knowledge influences visual working memory in adults and children. *PloS one*, 15(11):e0241110.
Stewart, E. E. M. and Schütz, A. C. (2018). Optimal trans-saccadic integration relies on visual working memory. *Vision Research*, 153:70–81.
Stewart, E. E. M. and Schütz, A. C. (2019). Transsaccadic integration benefits are not limited to the saccade target. *Journal of Neurophysiology*, 122(4):1491–1501. Publisher: American Physiological Society.
Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, 19(7):394–405.
Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron*, 78(2):364–375.
Stokes, M. G., Muhle-Karbe, P. S., and Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition*, 28(5-8):420–432.
Sutterer, D. W., Foster, J. J., Adam, K. C. S., Vogel, E. K., and Awh, E. (2019). Item-specific delay activity demonstrates concurrent storage of multiple active neural representations in working memory. *PLOS Biology*, 17(4):e3000239.

Swan, G., Collins, J., and Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision*, 16(3):32.

Swan, G. and Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, 76(7):2136–2157.

Tam, J. and Wyble, B. (2022). Location has a privilege, but it is limited: Evidence from probing task-irrelevant location. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Taylor, R. and Bays, P. M. (2020). Theory of neural coding predicts an upper bound on estimates of memory variability. *Psychological review*, 127(5):700.

Teng, C. and Postle, B. R. (2021). Spatial specificity of feature-based interaction between working memory and visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, 47(4):495–507.

Tomić, I. and Bays, P. M. (2022). Perceptual similarity judgments do not predict the distribution of errors in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition (in press)*.

Treisman, A. and Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, 34(8):1704–1719.

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

Udale, R., Tran, M. T., Manohar, S., and Husain, M. (2022). Dynamic in-flight shifts of working memory resources across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 48(1):21. Publisher: US: American Psychological Association.

Van den Berg, R., Awh, E., and Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological review*, 121(1):124.

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

Van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W. J. (2012). Variability in Encoding Precision Accounts for Visual Short-Term Memory Limitations. *Proceedings of the National Academy of Sciences*, 109(22):8780–8785.

Van den Berg, R., Yoo, A. H., and Ma, W. J. (2017). Fechner's law in metacognition: A quantitative model of visual working memory confidence. *Psychological review*, 124(2):197.

Van den Berg, R., Zou, Q., and Ma, W. J. (2020). No effect of monetary reward in a visual working memory task. *BioRxiv*, page 767343.

Vogel, E. K. and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984):748–751.

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

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

Wang, B., Cao, X., Theeuwes, J., Olivers, C. N. L., and Wang, Z. (2016). Location-based effects underlie feature conjunction benefits in visual working memory. *Journal of Vision*, 16(11):12.

Weber, A. I., Krishnamurthy, K., and Fairhall, A. L. (2019). Coding principles in adaptation. *Annual review of vision science*, 5:427–449.

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, 1(1):547–567. _eprint: https://doi.org/10.1146/annurev-vision-082114-035509.

Wei, X.-X. and Stocker, A. A. (2015). A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nature neuroscience*, 18(10):1509–1517.

Wei, Z., Wang, X.-J., and Wang, D.-H. (2012). From Distributed Resources to Limited Slots in Multiple-

Item Working Memory: A Spiking Network Model with Normalization. *Journal of Neuroscience*, 32(33):11228–11240.

Wheeler, M. E. and Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131(1):48–64.

Wilken, P. and Ma, W. J. (2004). A detection theory account of change detection. *Journal of vision*, 4(12):11–11.

Williams, J. R., Brady, T. F., and Störmer, V. S. (2022a). Guidance of attention by working memory is a matter of representational fidelity. *Journal of Experimental Psychology: Human Perception and Performance*.

Williams, J. R., Robinson, M. M., Schurgin, M., Wixted, J., and Brady, T. (2022b). You can't "count" how many items people remember in working memory: The importance of signal detection-based measures for understanding change detection performance. *Journal of Experimental Psychology: Human Perception and Performance*.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, 17(3):431–439.

Wolf, C. and Schütz, A. C. (2015). Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *Journal of Vision*, 15(16):1–1.

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., and Stokes, M. G. (2020). Drifting codes within a stable coding scheme for working memory. *PLOS Biology*, 18(3):e3000625.

Wolff, M. J., Jochim, J., Akyürek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, 20(6):864–871.

Wyble, B., Hess, M., O'Donnell, R. E., Chen, H., and Eitam, B. (2019). Learning how to exploit sources of information. *Memory & Cognition*, 47(4):696–705.

Xu, Y. (2020). Revisit once more the sensory storage account of visual working memory. *Visual Cognition*, 28(5-8):433–446.

Ye, C., Hu, Z., Ristaniemi, T., Gendron, M., and Liu, Q. (2016). Retro-dimension-cue benefit in visual working memory. *Scientific Reports*, 6(1):35573.

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

Yoo, A. H., Acerbi, L., and Ma, W. J. (2021). Uncertainty is maintained and used in working memory. *Journal of vision*, 21(8):13–13.

Yoo, A. H. and Collins, A. G. (2022). How working memory and reinforcement learning are intertwined: a cognitive, neural, and computational perspective. *Journal of cognitive neuroscience*, 34(4):551–568.

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

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

Zhang, W. and Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192):233–235.

Zhou, Y., Curtis, C. E., Sreenivasan, K., and Fougnie, D. (2022). Common neural mechanisms control attention and working memory. *Journal of Neuroscience*.

Acknowledgements

PMB was supported by a personal fellowship from the Wellcome Trust (Grant number 106926).

Author contributions

All authors contributed equally to this work.

Competing interests
The authors declare no competing interests.