Trends in Cognitive Sciences

CellPress

Review

Sensory reformatting for a working visual memory

Anastasia Kiyonaga [1,2,*] and John T. Serences [2,3]

A core function of visual working memory (WM) is to sustain mental representations of recent visual inputs, thereby bridging moments of experience. This is thought to occur in part by recruiting early 'sensory' cortical regions, via flexible fronto-parietal mechanisms. The nature of visual cortex activity during WM has been elusive, but new evidence suggests that early WM representations can transform from a sensory-like code into a format that is shaped by task context and optimized for behavior. Here, we review evidence for transformations in visual cortical WM coding, the various forms they take, and their functional importance. Visual cortex may be an active workspace during WM, where flexible and 'good enough' WM representations serve to interface with perception and action.

Highlights

A growing body of evidence suggests that working memory (WM) representations in early visual cortex can flexibly transform from a perceptual code into a format that is optimized for behavior.

Feedback may impose limits on visual WM resolution while also facilitating various context-dependent transformations to the WM code.

Visual WM may capitalize on sensorimotor cortical areas as an efficient interface between higher order computations and the sensory environment.

Apparent distortions to precise WM representations may reflect adaptive shifts to reciprocally align perception, action, and top-down goals.

Visual working memory needs flexibility

Working memory (WM) serves as a temporal bridge between perception and action, and it operates by transiently storing sensory information to guide behavior [1,2]. For instance, before crossing a busy intersection, you might take several glimpses in each direction to refresh your mental picture of the moving traffic. We cannot perceive both directions simultaneously, so we rely on visual WM to plan a safe decision by keeping essential features of the traffic in mind (e.g., car distance and speed). WM, therefore requires fast, fluid interplay with the environment to flexibly prioritize behaviorally relevant aspects of the scene.

Despite this flexibility, WM research has often aimed to pinpoint a singular mechanism or anatomical locus for WM. Likewise, much ink has been spilled debating whether WM storage critically relies on early sensory, motor, or fronto-parietal and **association cortices** (see Glossary) [3–10]. Yet WM is at the same time credited with a vast array of functions [11]. Humans may use visual WM to cross the street, keep track of multiple players on a field, or plan a series of chess maneuvers. These operations differ in timescale, complexity of contingencies, opportunity to refresh the sensory representation, and motivational stakes. Thus, attempts to pinpoint WM storage will inevitably find inconsistent evidence and may also overlook the system's true priorities. Namely, while successful WM storage has conventionally been construed as a faithful trace of the encoded sensory content, a more adaptive code may often serve behavior better. Aptly, the arc of WM research has recently bent toward asking how the context and intended use of WM content may modify how it is maintained [12].

WM-related signals are now understood to be spread across many interacting brain regions, carrying information at multiple levels of abstraction, and via patterns that fluctuate with time and context [7,13–15]. In this distributed scheme, WM-related activity in fronto-parietal regions is often considered to represent stimulus abstractions and **domain-general** control, while that in occipital cortex is considered to represent **feature-specific** visual content [16–18]. However, given how brain-wide signals wax and wane, and demands in the sensory environment vary, the nature of occipital WM activations remains contested. For instance, it is unclear whether

[1]Department of Cognitive Science, University of California, San Diego, CA, USA
[2]Neuroscience Graduate Program, University of California, San Diego, CA, USA
[3]Department of Psychology, University of California, San Diego, CA, USA

*Correspondence: akiyonaga@ucsd.edu (A. Kiyonaga).

WM representations in sensory cortex can resist interference from incoming sensory stimuli and whether they are suited to meet behavioral needs. Indeed, while more anterior WM signals are often robust to interference, those in visual cortex can be susceptible to various distortions – from new sensory input, from stimuli encountered in the recent past, and from other information held in mind concurrently [19–23].

Rather than a vulnerability, however, this malleability may be the essence of visual cortical WM function. For instance, while fronto-parietal regions are known to support flexible WM, transmission to prefrontal cortex (PFC) can be slow and the representations coarse [24–27]. A complementary flexible processing stage in early sensory cortex could responsively keep WM aligned with the external world and accessible for use. Here, we will review the growing evidence that WM representations in early sensory cortex may be transformed and abstracted for the task context. We will trace the evolving understanding of visual cortical involvement in WM, examine the scope of possible neural coding schemes for sensory WM, and speculate on how these flexible codes might support adaptive behavior.

Shifting perspectives on WM storage

Electrophysiological recordings from macaque prefrontal, parietal, and temporal cortices have historically suggested that WM content is stored in higher-order heteromodal or association regions [3,9,28]. For instance, during a WM delay, PFC neurons show **spiking** patterns that are selectively tuned to the features of WM stimuli, while such WM-related spiking is rarely detected in early sensory regions. The flexible PFC coding properties, protracted neuronal timescale, and long-range connectivity make the area especially well-equipped to support diverse WM needs [29–32]. However, the role that fronto-parietal activity plays in WM content storage has remained an active area of inquiry [3,26,28,33].

Human neuroimaging instead highlights that WM information is distributed across the brain, including the same unimodal sensory regions that perceptually encode the content [13,18]. For instance, multivariate **fMRI** activity patterns across V1 voxels can be used to decode or reconstruct a remembered stimulus orientation during a WM delay [34,35]. This suggests that information about specific remembered orientations is represented in activity patterns across V1 populations, even when the stimulus is not being perceived. Similar findings across many other types of stimuli and task conditions fuel the theory that high-resolution WM content is stored in sensory cortical representations [16,26].

This general perspective is often referred to as 'sensory recruitment', with the idea being that flexible fronto-parietal regions recruit sensory cortex to sustain the WM content (Figure 1A). This theory follows the logic that visual cortex is structurally optimized to represent fine-grained visual information, and an efficient system would repurpose the same cortical territory for multiple functions [36]. But the theory takes many shapes, and the role that early visual cortex plays in WM has proven vexing to define. The next section will summarize this theoretical landscape.

The form and function of WM activity in visual cortex

Although feature-specific WM activity is now routinely detected in early visual cortex, there are many explanations for what that activity might reflect. Even within the perspective that the activity reflects content storage (Figure 1A), there are various viewpoints as to how that storage is achieved (Figure 1D–H). By some initial accounts, the visual 'content' of WM is a sustained trace of perceived content and the same representations that support perception support WM as well [34,35,37] (Figure 1F). Thus, early visual WM representations would be held in an iconic-like format that has isomorphism with a corresponding perceptual representation, almost

Glossary

Association cortices: the regions of neocortex that are not involved in primary sensory or motor processing. These areas are spread throughout the lobes of the brain and they support higher-order cognitive function by receiving and integrating information from a variety of sources.
Attractors: stable states that a system converges toward over time. In the case of WM, the neural activity patterns representing certain stimuli may be more stable than others, forming attractors toward which the less stable patterns evolve, producing an apparent drift or distortion in the memory.
Domain-general: functions that are shared across several different kinds of tasks or types of information. For instance, the same region of prefrontal cortex may send control signals during both visual and auditory WM. Conversely, domain-specific functions are specialized for processing particular kinds of tasks or information.
Feature-specific: neural responses or activity patterns that are selective for certain stimuli and differentiate between specific instances of a feature. This term is sometimes used interchangeably with 'stimulus-specific' and the property is often considered a criterion for a region to store WM content. As an example, if a neural population produces reliably distinct activation patterns for different hues along the color wheel, we would consider the population to have color-specificity.
Feedback: processing where information flows from higher-order areas back down to lower-order areas. For instance, when the prefrontal cortex transmits control signals that modulate activity in visual cortex. This is sometimes referred to as top-down processing.
Feedforward: processing where information flows from lower-order areas up to higher-order areas. For instance, when input regions like the visual cortex transmit sensory information to the parietal or frontal cortex. This is sometimes referred to as bottom-up processing.
Functional MRI (fMRI): a non-invasive but indirect neuroimaging technique that approximates brain activity by detecting changes in blood flow. fMRI is sensitive to metabolic changes that can stem from neuronal spiking, but may also result from other signals such as sub-

like a photograph in your mind (Figure 1D). However, this perspective is undermined by the possibility that iconic-like WM representations would be perturbed by new sensory inputs [4,17], as well as the fact that WM-related sensory activity patterns can evolve over time [7,8]. Thus, several variations on the strictest interpretation of sensory recruitment have emerged – and these are not mutually exclusive. For instance, the raw activity may drift over the delay, while the feature information encoded in the activity pattern remains stable [17,38–40] (Figure 1E). WM may also engage sensory cortex alongside distributed representations across the brain, varying in specificity and abstraction, that help the relevant information to bridge disruptions and withstand perturbation [7,13,15] (Figure 1G). Or WM may recruit early visual cortex but employ distinct representations from perception; for instance, expressing a similar code in a distinct cortical layer, or interdigitated but distinct neural populations from perception [41,42] (Figure 1H). Thus, many theories assert that sensory cortex contributes to WM maintenance, but there is little consensus on its mechanistic function.

Likewise, other perspectives question whether apparent WM-related activity in visual cortex truly plays a storage role. For instance, sensory representations may anticipate an upcoming probe, serve as a template for visual comparison, or output goals for attention and eye movements [4,7] (Figure 1B). Alternatively, some are skeptical that sensory cortex is functionally involved in WM beyond encoding. According to one common refrain, apparent V1 WM representations may be an artifact of input from association regions, where the true storage occurs [3,4] (Figure 1C). In this case, higher-order representations send **feedback** signals that alter field potentials in visual cortex, and these sub-threshold changes are detected by macroscale imaging methods like fMRI despite no local spiking. Indeed, it has been rare to find sustained feature-selective spiking in V1 during WM. But it has also been rare to test for it. Electrophysiology studies of WM in PFC and other association areas outnumber those in primary sensory regions by at least an order of magnitude [3], hampering attempts to reconcile theoretical perspectives. Next, we describe new electrophysiological findings that corroborate a more nuanced take on early visual WM activity [43–46].

Mnemonic codes in visual cortex diverge from stimulus-evoked codes

Recent studies show that V1 spiking patterns do track visual WM contents, but with caveats [15,43,44]. For instance, in monkeys remembering natural objects, a corresponding V1 trace persisted into the delay period but decayed quickly, suggesting a fading sensory representation in local reverberating activity [43]. Such a trace emerged even for passive viewing, underscoring that apparent WM content representations may sometimes reflect only lingering sensory-evoked activity. In another case, however, V1 showed enduring stimulus-specific WM activity, but in spiking patterns that differed from those at perception [44]. Neurons showed distinct stimulus preferences and functional connectivity relationships from encoding to delay, and decoding failed to generalize across timepoints, consistent with a time-varying code. Therefore, V1 does appear to sustain feature-specific WM content, but via different sub-populations and patterns than the sensory-evoked trace.

These results are noteworthy in showing local V1 activity for non-spatial WM features (which has otherwise been limited). The results also echo human neuroimaging findings — namely, that stable WM information can persist in fluctuating sensory activity patterns [8,40]. That is, analyses that assume a steady activity pattern might train a stimulus classifier on data from a fixed timepoint (e.g., encoding), but this sort of classifier often does a poor job detecting stimulus-specific memory information at later timepoints in the trial [8]. If the classifier is instead both trained and tested on data from later in the trial (e.g., the delay), it can often reliably detect WM content information [39]. Indeed, we would expect such poor cross-temporal generalization if WM representations differed from the initial sensory-evoked code. These activity dynamics may contribute to apparent

threshold changes in local field potentials.

Multiplex: the idea that the same neurons or cortical regions can be repurposed to represent more than one type of information or to support more than one function.

Representational drift: the phenomenon whereby the neuronal response to the same stimulus gradually changes over time, even after learning plateaus and without any additional experimental manipulation. This could also be understood as a change in the tuning of individual neurons, as a neuron that is selective for a given stimulus at one timepoint may become selective for a different stimulus at another timepoint.

Spiking: neuronal firing, or propagating an action potential. Spikes are electrical impulses that transmit signals between neurons. They can be measured with electrophysiological techniques like electrodes that are inserted intracranially.

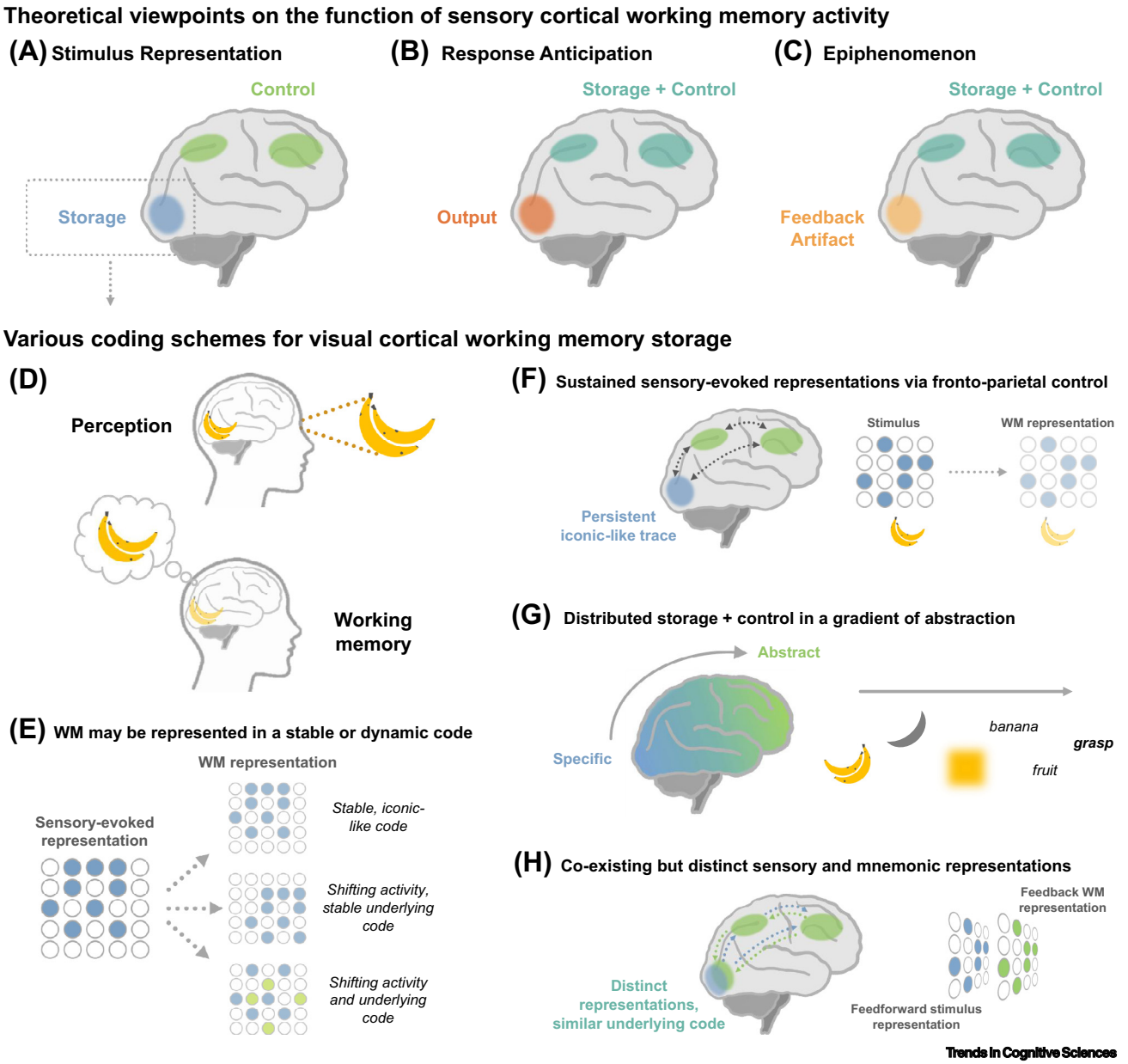

Figure 1. Theoretical perspectives on sensory cortical working memory (WM) activity. (A–C) Schematic depictions of three schools of thought regarding WM-related activity in visual cortex. (A) It may reflect precise content storage, while fronto-parietal regions exert flexible control over that content. (B) It may reflect a readout function to support a required response, while the content maintenance itself is supported by fronto-parietal regions. (C) It may be an artifact of input from higher order regions, where the activity in connected regions leads to epiphenomenal changes in visual cortical membrane potentials. (D–H) Within the broader perspective that WM-related visual cortex activity reflects content storage, there are various viewpoints about the coding scheme for storage. (D) Illustration of the iconic-like WM representation concept, where WM engages the same code as perception. (E) Schematics depicting stable and dynamic WM codes. Circles can be taken to represent neurons or voxels, and color represents the type of information being reflected in the neural activity. (F) A strict sensory recruitment perspective, where WM representations in visual cortex are sustained versions of the sensory-evoked code. (G) A distributed framework, where WM-related signals are spread across relevant brain regions. Low-level stimulus features might be represented in early visual areas, categorical codes in parietal or temporal cortex, and action plans in frontal cortex. (H) WM and sensory-evoked representations may also engage the same cortical regions, and similar coding schemes, but in distinct populations – one population that is excited by feedforward input (blue) and another that is excited by feedback (green).

discrepancies between human fMRI and other approaches that instead test for persistent firing in the same neurons that are selective at encoding (Box 1).

Stimulus activity in early visual cortex can be dynamic

There is now a compelling corpus of evidence that WM information is expressed in early sensory cortical activity, but that need not imply that it recapitulates the perceptual code or that its function is passive storage. Perception may impress a trace that briefly lingers after encoding, whereas higher-order WM feedback exploits sensory cortex to galvanize a unique spatio-temporal activity pattern.

Such time-varying activity may reflect processes that are distinct to WM, or it may reflect common neural processing motifs that are shared by functions like perception and long-term memory (LTM) as well. For instance, during sensory perception, early visual cortical representations can be biased by adaptation, attention, expectations, and numerous top-down factors [22,47–50]. And over a longer period of days and weeks, visual cortex activity associated with the same stimulus can gradually deviate from its initial pattern [51]. This long-term **representational drift** might support mnemonic codes that minimize interference with new inputs by morphing away from stimulus-evoked patterns. Alternatively, drift may simply reflect a random walk through different configurations of an overparameterized system that produce equivalently high performance [51,52]. Similarly, drift in WM activity patterns – over a much shorter timescale of seconds – may be functionally adaptive, or it may manifest neural codes that have many degrees of freedom and thus many equally viable configurations.

Box 1. Model systems shape working memory theory

The debate over sensory cortical WM has largely been waged on two fronts: one within the realm of human neuroimaging over how to interpret WM-related activity in sensory cortices [4,5,10], and another between human neuroimaging and non-human primate (NHP) electrophysiology research over whether the activity exists in the first place [3,9]. It therefore bears examining where methodological or species constraints may govern theoretical progress (Figure I).

Measurement sensitivity: electrophysiological recordings that detect spiking activity are prized for their temporal and spatial precision, but are necessarily limited to a fraction of the brain at a time. fMRI indirectly assesses responses across the whole brain, summing over populations to detect more distributed signals, including sub-threshold modulations that may not lead to spiking. These measurements lend themselves to different analysis approaches, and fMRI might more easily capture representations that are spatially diffuse or diverging from the sensory-evoked trace.

Subject experience: monkeys undergo months of training and thousands of trials to complete a basic memory task that humans learn in minutes [69]. Such experience may alter WM representations for repeated sensory stimuli, rendering them more categorical and reliant on anterior regions [105]. When NHP recordings are examined before and after training, anterior PFC regions show increased WM engagement and stimulus selectivity over time [101,106]. Likewise, when humans undergo months of WM training, PFC fMRI activity shows increased engagement and stimulus selectivity over time – similar to NHPs [107]. Visual cortical WM activity might appear more prevalent in humans because NHP recordings occur after training has altered their content representations.

Visual cortex function: the macaque visual system has been the primary model for human function, but NHPs may show different visual perception patterns than humans. For instance, macaques have shown opposite asymmetries from humans in their visual field preferences (e.g., lower vs. upper) and different effects of polar angle on perception [108]. Humans have a relatively expanded extrastriate cortex [109] and may also exhibit unique cytoarchitecture in V1, like distinct interneuron types and abundant non-neuronal cell types [110]. Human visual cortex anatomy and connections [111] may allow WM coding schemes of a timescale and complexity that are unattainable in macaques.

Reconciling approaches: NHP fMRI has begun to establish that electrophysiological activity patterns recapitulate in blood oxygenation level dependent (BOLD) imaging [31,112]. High-density electrophysiology and laminar fMRI may provide ever-better convergence across species. However, differences in methodological constraints can sometimes mask as species constraints and vice versa. For instance, differences in visual cortex function may be improperly attributed to measurement sensitivity, thereby biasing conclusions. We should therefore consider how our models determine what we consider possible.

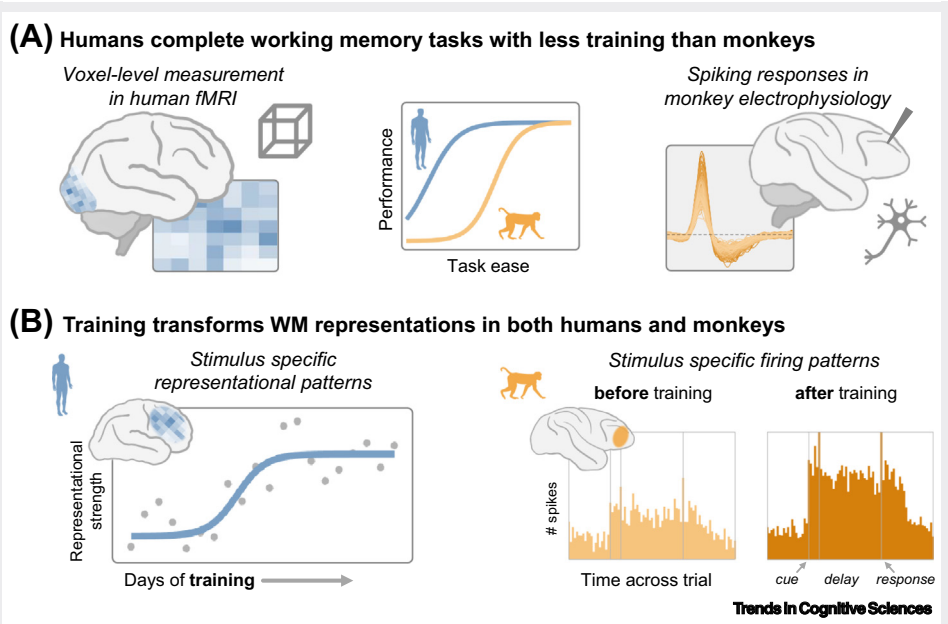

Figure I. Differences and similarities across human (left) and monkey (right) working memory (WM) research. (A) Typical human and macaque measurement approaches, as well as a schematic of psychometric functions to depict human WM performance after a small number of trials (blue) versus macaque performance after thousands of trials (beige; portraying findings from [69]). (B) Despite differences in measurement sensitivity, when either humans or macaques undergo extensive WM training, it induces parallel changes in prefrontal cortex WM representations. Schematic depictions (portraying findings from [106,107]) highlight the dynamism in WM activity patterns over longer timescales and the importance of subject experience to WM function.

In the preceding sections, we introduced basic evidence that WM representations in early sensory cortex deviate from the sensory-evoked code. Next, we will examine several different ways that the activity can morph and the purpose that those transformations might serve.

Feedback shapes context-sensitive mnemonic codes in sensory cortex

Unlike the **feedforward** response to sensory stimuli, an enduring WM signal relies on recurrent feedback to sustain the activity in early visual cortex [9,28]. Therefore, while information about specific WM features can be extracted from signals in visual cortex, feedback may shape the substance of those signals [18,33,42,53,54]. For one, feedback may constrain the precision that WM can achieve if the feedback emanates from later visual regions with coarser representations [55]. Further, feedback may modify the content and coding format of WM representations in visual cortex, if it originates from more anterior association regions with distinct and flexible coding properties [54,56,57]. Since behavioral goals can rapidly change, and WM traces may be subject to interference from new inputs, feedback may transform early representations into a task-optimized state.

Early mnemonic representations reflect the properties of later input regions

When visuo-spatial content is imagined or retrieved from memory, it is still organized retinotopically in visual regions, but can exhibit different tuning properties than it would during perception [58–60]. Namely, during visual perception, occipital regions show a processing hierarchy in that later units exhibit progressively larger receptive fields and are tuned to lower spatial frequencies [61]. However, during both spatial WM and LTM retrieval, responses at different

positions along this hierarchy become less differentiated from each other in their tuning and amplitude [58,60] (Figure 2A). For example, relatively later visual areas like V4 retain their broader tuning, whereas earlier areas go from being more narrowly to more broadly tuned – consistent

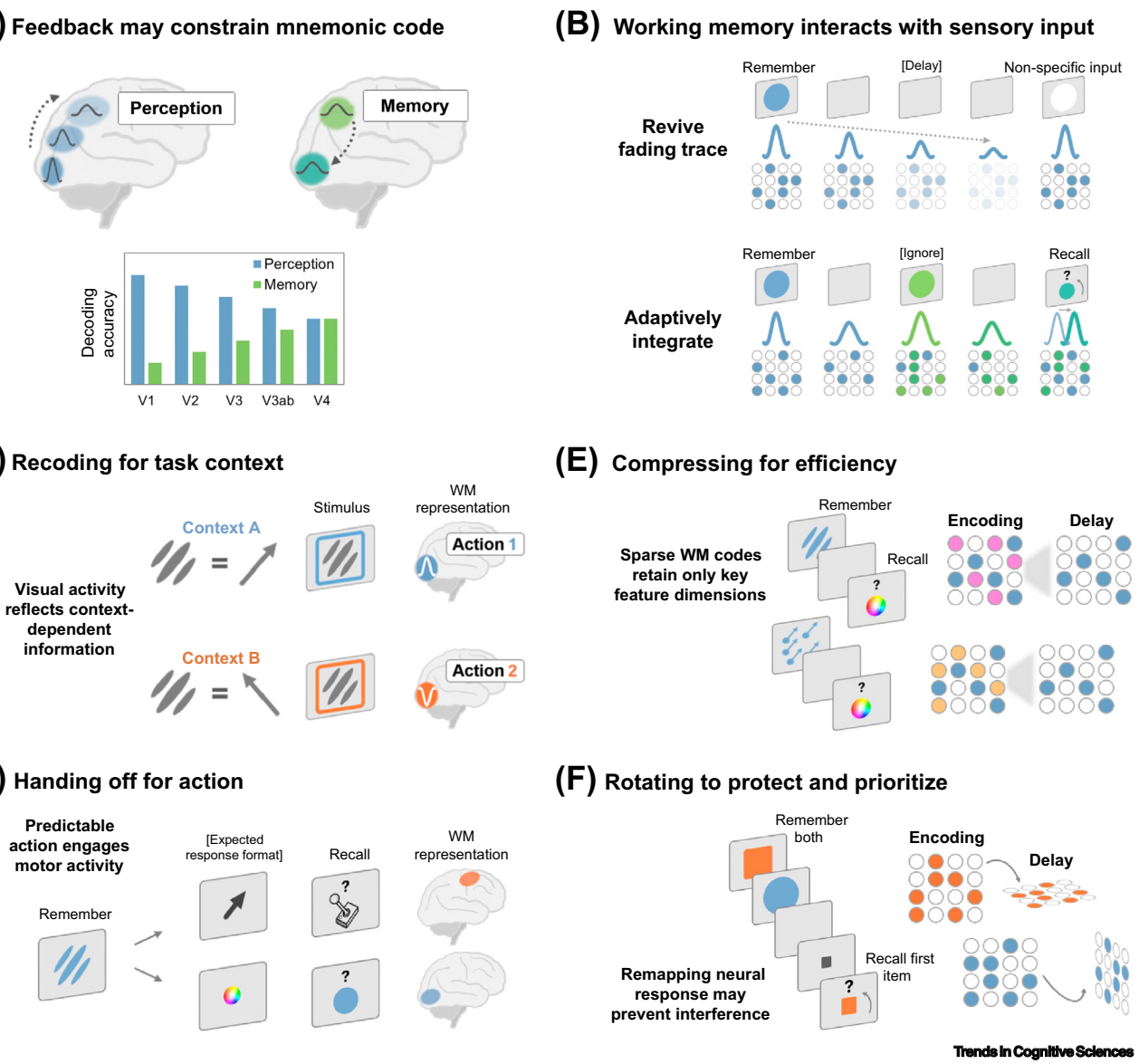

Figure 2. Context-sensitive transformations to visual cortical working memory (WM). (A) Brains illustrate that tuning in visual regions shows a processing hierarchy during perception (blue), but the WM representations in those regions (green) no longer show such differences in tuning [55,56,58–60]. Barplot schematically illustrates stimulus decoding accuracy across visual regions, during either perception or WM, when the model is trained on perception (after [41,55,56,58]). (B) Panels show schematic depictions of WM activity patterns that change across a trial and as WM interacts with feedforward input. A lingering sensory-evoked trace fades in the absence of input, but may be revived through external perturbation [43,71] (top). WM may also assimilate new input, manifesting as a subtle bias or distortion [90] (bottom). (C–F) The sensory-evoked representation may also reformat into a different coding scheme to support upcoming behavior, as depicted in these schematic illustrations. (C) Visual cortical activity may transform to prospectively represent a context-sensitive choice rather than the encoded sensory features [46]. (D) A visual stimulus may be offloaded to motor circuits when the associated response can be anticipated, but remain in a visual format when the response is unknown during the delay [74,75,79]. (E) A visual representation may compress by dropping information that is not behaviorally relevant [64,80]. As a result, two perceptually distinct stimuli may evoke similar WM representations if they share the same task-critical feature. (F) WM codes may reconfigure into an orthogonal activity subspace by remapping the neural response pattern while retaining the original representational structure (see also Box 2) [83,84].

with V4 acting as a conduit for feedback to earlier areas. Importantly, later occipital regions might still show flexible and time-varying WM activity, despite their stable tuning. And earlier areas can still represent stable memory feature information (in this case, spatial location), despite their shift in tuning. However, cross-generalization in activity patterns from perception to WM is often worse in earlier areas compared with later occipital areas like V3ab or V4 [41,55,56,60] (Figure 2A) – as we would expect if the response properties in earlier areas become shaped by feedback from later areas that have coarser feature and spatial tuning.

Visual cortex representations during WM can also appear more categorical than they do during perception [62–64]. This could stem from drift toward stable **attractors** [65] or could arise due to feedback from parietal and frontal regions that become engaged during WM. Representations across these fronto-parietal regions are broadly found to be flexible, goal-oriented, and often categorical in nature [25–27,33,66–69], and they may bias WM representations in visual cortex accordingly. Indeed, visual cortex regions that encode precise stimulus features during perception (like particular colors or orientations) can instead reflect categorical groupings and biases toward category prototypes during WM [62,63,70]. Moreover, visual cortex patterns during WM can look more like the parietal patterns measured at perception, as compared with the corresponding visual patterns measured at perception [56]. Thus, across colors [62], spatial locations [58,60], orientations [63,64], real-world objects [56], and natural scenes [59], feedback may shape sensory cortical WM representations to higher-order codes rather than sustain precise sensory-evoked representations.

However, WM does not occur in a vacuum. When we use visual WM to cross the street, we do not check the traffic once, then close our eyes and hope for the best. Instead, we continuously sample from the sensory environment to update our mental picture. Such interactions between top-down feedback and feedforward input may alter the information content and achievable resolution of sensory cortical WM representations [48,49]. Even nonspecific input may boost coarse or weak sensory cortical WM representations [71], and early visual WM representations may be less bounded by feedback properties when they interact with sensory input [8,72,73] (Figure 2B). The sensory cortical WM resolution may thus depend on whether WM content is purely top-down or continuously interacting with bottom-up input, stressing the importance of task context in shaping WM function.

Prospective demands mold the mnemonic coding format

When the task context informs how WM will be used, WM representations can be tailored toward upcoming behavior [1,74–76]. Prioritizing certain WM content may reformat its representation into a task-optimized state [77], and the same WM information may be maintained in different 'use-dependent' codes [12,66]. Such prospective behavioral contingencies can be conveyed in higher order (e.g., PFC) representations [77,78], and motor WM codes may activate alongside precise visual information [75]. However, a growing body of evidence suggests that behavioral intentions can shape WM representations in visual cortex as well (Figure 2C–F).

For example, perceptually identical WM stimuli are remembered as more dissimilar when they are paired with distinct response actions, suggesting that motor intentions distort sensory memories to align them with upcoming behavior [76]. In monkeys trained to alternate between stimulus-response mappings in a visuo-spatial WM task, the information encoded in V4 delay activity also differed with the context-dependent mapping rule [46]. That is, V4 signaled the relevant information for an upcoming response, rather than sensory features of the WM sample (Figure 2C). Further, WM delay activity for the same visuo-spatial stimulus can exhibit a more motor-like pattern when an upcoming manual response is known, but a more sensory-like pattern when the

response is unknown [74,79]. For features like color and spatial location, the sensory information appears to be re-coded into a motor format if an upcoming action can be prepared (Figure 2D). Such motor WM codes can also emerge immediately in a WM delay, and relate to better performance [74,75], as if the modus operandi of WM were fast translation of sensory input into prospective codes for action.

Perceptually distinct stimuli can also evoke interchangeable visual cortical WM representations when they share an action-relevant feature [64] (Figure 2E). For instance, WM sample stimuli that are presented as moving dot arrays versus oriented gratings both share the relevant dimension of 'angle'. Fittingly, although the encoded stimuli physically differ, their visual WM representations appear to share a line-like format, suggesting that the visual cortical signal conveys an abstract or categorical code [64,80]. While several mechanisms might undergird that apparent abstraction, it could reflect a form of compression from a dense to a sparse code, given that unneeded feature dimensions can be abandoned without sacrificing behavioral performance (Box 2). Likewise, some of the earliest WM decoding from visual cortex showed that delay activity patterns for multi-feature stimuli (e.g., color + orientation) amplified just the relevant feature for an upcoming memory test [35]. Simplifying codes in this way might support more efficient information processing and protect from interference caused by new sensory inputs. This would echo theoretical LTM functions whereby representations grow increasingly sparse over time to incorporate new memories without overwriting old ones [81].

Collectively, the evidence suggests that perceptually similar WM stimuli are represented distinctly when they precede distinct actions, whereas perceptually distinct stimuli are represented similarly when they precede similar actions. How the WM content will be used seems to supersede what the encoded stimulus looks like. Even early visual cortical WM patterns may reflect compressed low-dimensional codes rather than high-resolution simulacra of physical stimuli.

Rotating WM representations may save them for later

When multiple items are maintained simultaneously, or in the face of concurrent sensory input, the pattern supporting mnemonic codes can seemingly reconfigure in a way that minimizes overlap between representations [53,57,82–85]. For instance, in mice learning sound sequences, a subset of auditory cortex neurons reverse their selectivity for a given stimulus from perception to memory – effectively remapping the sensory code into a designated memory subspace [84]. In this case, the underlying information remains stable and the same neurons generally participate in the code, but some of the neurons flexibly re-map their tuning, generating a representation that is nearly orthogonal to the sensory-evoked response.

Relatedly, in humans, alternating attention between two visual WM stimuli, attended and unattended (i.e., deprioritized) WM items can evoke opposing patterns of multivariate fMRI activity [83,85]. In simplified terms, a 45° oriented grating may evoke a pattern of neural responses that looks like 45° when actively maintained in WM, but that looks like 135° when that same stimulus is maintained in the background. Like a rigid-body transformation or simple cipher, the underlying representational structure is seemingly retained by these 'rotational dynamics', but in different neural response patterns (Figure 2F). This general phenomenon has been observed across a range of WM stimulus classes and tasks, especially when similar stimuli vie for attention or must be ordered in sequences [53,57,82–88]. Such partitioning may therefore limit conflict between sensory and mnemonic content, untangle object features, or tag WM representations at different levels of temporal immediacy. However, the mechanisms supporting apparent rotations are still unresolved. For instance, they may emerge from cells with conjunctive coding [78], nonlinear mixed selectivity (Box 2), or be explained by propagating activity in travelling waves [89].

Box 2. Mixed selectivity supports flexible sensory working memory codes

The ability to hold information in memory has previously been attributed to stereotyped neural responses evoked by external stimuli, such as orientation selective cells in V1 [113]. In contrast to this relatively rigid architecture, recent work focuses on the need to continuously balance energy efficiency, capacity, and resistance to interference. This balance can be viewed from an information-theoretic perspective by considering the number of possible states that a code can take and the probability of each state (i.e., the entropy of a code) [114]. For example, sparse codes are generally low entropy because their small size restricts the number of possible configurations and thus limits their capacity to represent information. Despite the low capacity, sparse codes are inherently energy efficient and, by recruiting a small proportion of all active units, multiple co-active sparse codes are less likely to mutually interfere [17,115]. By contrast, a dense code with units tuned to different stimulus attributes can achieve many possible states and will generally have a higher entropy. However, a dense code can also be low entropy if many units encode the same stimulus attributes to sacrifice encoding capacity in favor of redundancy and error tolerance [114].

Given the range of possible coding strategies, when memory requires lower precision, codes could be compressed into a sparse (or dense-redundant) code with fewer degrees of freedom to retain only the essential elements for behavior [64,74]. By contrast, when memory requires higher precision, it may rely on dense, high-entropy codes in early sensory areas [16,34,35]. However, sensory areas often need to precisely encode memories along with new inputs, raising the possibility of interference. One solution may be to rotate WM codes away from the sensory evoked response, which may mitigate interference while still supporting high precision representations [17] (Figure I). These rotations are likely supported – at least in part – by neurons that flexibly change their tuning selectivity (termed non-linear mixed-selectivity). For example, a neuron might respond maximally to stimulus A during perception, but to stimulus B during memory [29]. The ability of the same neuron to represent multiple sensory features in the context of different tasks can further increase the degrees of freedom and overall encoding capacity of even a dense code [29]. Thus, codes could be sparse or dense depending on encoding demands, and mixed-selectivity may support flexible, high-entropy codes that have enough degrees of freedom to jointly represent memory and sensory related information in different sub-spaces, thereby avoiding destructive interference [84].

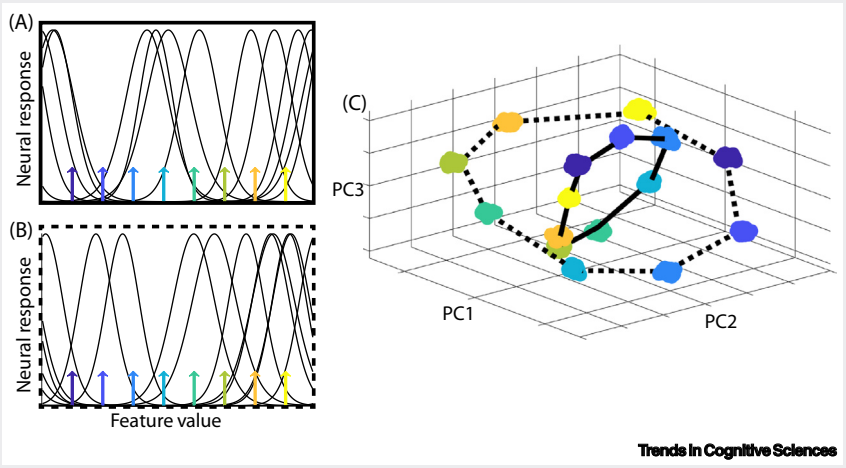

Figure I. Mixed selectivity gives rise to rotations in feature-selective representations. (A) A set of simulated neurons that are tuned to different values in a feature space. A circular stimulus space is depicted for simplicity (e.g., motion direction, orientation). (B) The same set of neurons, but with shifted tuning functions as observed under non-linear mixed selectivity (here depicted as extreme shifts for illustrative purposes). (C) Population-level representation of eight different feature values (colored vertical arrows) based on tuning functions in A (unbroken line) and B (broken line) after passing the stimuli through the tuning functions, adding noise, and performing dimensionality reduction. In both cases the stimulus space is represented with a ring-like geometry, but rotation renders the two representations nearly orthogonal.

Taken together, across model systems and task domains, mnemonic content often activates a distinct pattern from initial perception, even when it engages overlapping sensory populations. Given that sensory WM traces would likely be vulnerable to various forms of interference, morphing representations might circumvent that fate by configuring into a more robust or task-adaptive code.

Sensory cortical working memories are queued up for use

Some sensory WM transformations retain a stable (if sparser) representation, as dictated by top-down feedback. These raise questions about what purpose early sensory WM representations serve, if they largely recapitulate the higher-order code. Other transformations may degrade or alter WM content away from the initially encoded features. For instance, in addition to the categorical biases and abstractions described earlier, sensory cortical WM representations can sometimes become biased by concurrent perceptual input [41,72,73,90] (Figure 2B). Such biases are typically considered corruptions, but they spur questions about whether there might be advantages to a system with mutable or coarse-grained sensory WM representations.

WM is for (near) future use – whether that be attentional deployment, other cognitive operation, or motor action – and WM usage often requires abstraction to a few critical dimensions. For instance, when crossing the street, you need not recall that the oncoming car is a pearlescent burnt sienna, nor memorize its license plate. Instead, basic speed and color attributes are sufficient for the task at hand. From this perspective, WM should often operate by a 'good enough' principle, expending only the necessary metabolic resources to meet current goals [91]. Indeed, in more naturalistic WM settings, humans tend to evade unnecessary effort [92,93]; there is little need to hold a high-resolution stimulus in mind if it can be sampled from the environment or a simpler representation will suffice [94,95]. Thus, the malleable nature of sensory WM representations may reveal an efficient system to activate the most ecologically relevant signals.

One natural advantage to using visual cortex for WM is that it **multiplexes** specialized cortical territory, but a reciprocal advantage is that it renders the activated content behaviorally potent. Content-specific sensorimotor engagement may potentiate processing in favor of WM, facilitating attention and action for goal-relevant information in the environment [6,96,97]. Thus, compressed codes and broader WM feature tuning would quickly capture an effective range of potentially relevant external signals to prioritize. And a pliable, early representation space would enable WM to promptly adapt to a changing environment. That is, on balance, it may be beneficial for WM to assimilate goal-adjacent perceptual input, encouraging seamless information flow and continuously updating which external signals to amplify. The fast sensorimotor time-scale would facilitate such updating, as well as comparing WM content with perceptual targets and responsively translating WM representations into action.

While WM has been closely linked with action and response preparation since delay activity was discovered [2,98], such prospective functions have typically been attributed to motor circuits and PFC coding. Newer findings (reviewed earlier) now show that occipital WM representations can also favor action-oriented codes [46,64]. This may enable visual cortical WM to serve as a junction with the oculomotor system, where abstracted sensory codes can most efficiently steer motor plans. In this framework, bottom-up visual cortex activity tracks eye movements that change the retinal input, while top-down templates in visual WM guide the eyes toward goal-matching targets in turn. Visual cortical WM representations may be routinely read out as oculomotor commands – in a continuous loop between frontal cortical and ocular signaling – in which case they would best reflect prospective goals rather than retrospective sensory features (Box 3). Indeed, behavioral precision corresponds with WM representations in early visual cortex better than other areas [8,73], suggesting that those sensory representations are most proximal to action outcomes [37,99].

Early visual cortex receives feedback projections from later visual regions, other primary sensory regions, and more anterior association regions [49]. As a result of this convergence, visual cortex may record intermediate computations from later regions in a format that is accessible and easily read out for action. Rather than solely supporting retrospective storage, visual cortex

Box 3. Oculomotor working memory parallels flexible visual cortex function

WM content information can now be detected in evolutionarily earlier and more primary structures than previously thought – like the cerebellum [116], thalamus [117], and superior colliculus [118]. Beyond simple spatial coding, moreover, these areas may express flexible and abstract WM codes.

For instance, the superior colliculus is a midbrain structure typically associated with visuo-spatial orienting and oculomotor control. However, it also appears to represent visual category abstractions during WM [119]. In monkeys completing a delayed matching task – where they had to categorize visual motion to answer a later probe – delay activity in the superior colliculus represented category information independent of physical stimulus properties. This early oculomotor structure may therefore be involved in more complex cognitive coding than previously realized.

Mounting evidence now shows that ocular indices like gaze position and microsaccade frequency can also reflect WM feature content (Figure I). For instance, during a WM delay, small gaze biases veer toward locations in memorized visual space [120]. These biases extend beyond location coding to also reflect geometric shapes, object-specific fixation patterns, and the relational structure in a continuous feature space [121,122]. As a WM delay progresses, such gaze biases reflect less object-specific content and more abstraction to the response-relevant dimension, prospectively anticipating an upcoming stimulus or probe [122–124]. Thus, like flexible visual cortical WM signatures, feature-specific WM gaze biases appear aligned with future behavior more than reflective of what was encoded.

This ocular WM modulation is not limited to eye movements or spatial features, as pupil size also rises and falls with the remembered brightness of a WM stimulus [125,126]. This effect mirrors the well-known pupillary light response, except it emerges endogenously when sensory input is matched between conditions. The effect is further magnified when the WM content is more behaviorally relevant, and it appears to ramp up in anticipation of a probe – suggesting that it is also prospective in nature [127].

Thus, WM content information can be read out from subcortical and peripheral sensorimotor structures, which may play a more cognitively complex role than previously assumed – potentially manifesting the prospective codes that are now observed in visual cortex. Still, it is unknown what underlying activity these oculomotor signatures reflect. There is some evidence that saccades can causally impact WM, but the direction of influence between ocular and cortical WM signals is unclear [128,129]. The relationship is likely reciprocal, whereby WM activation in visual cortex triggers a corresponding motor adaptation to align perceptual processing and action with visual goals – thereby updating and refining WM representations in return [130].

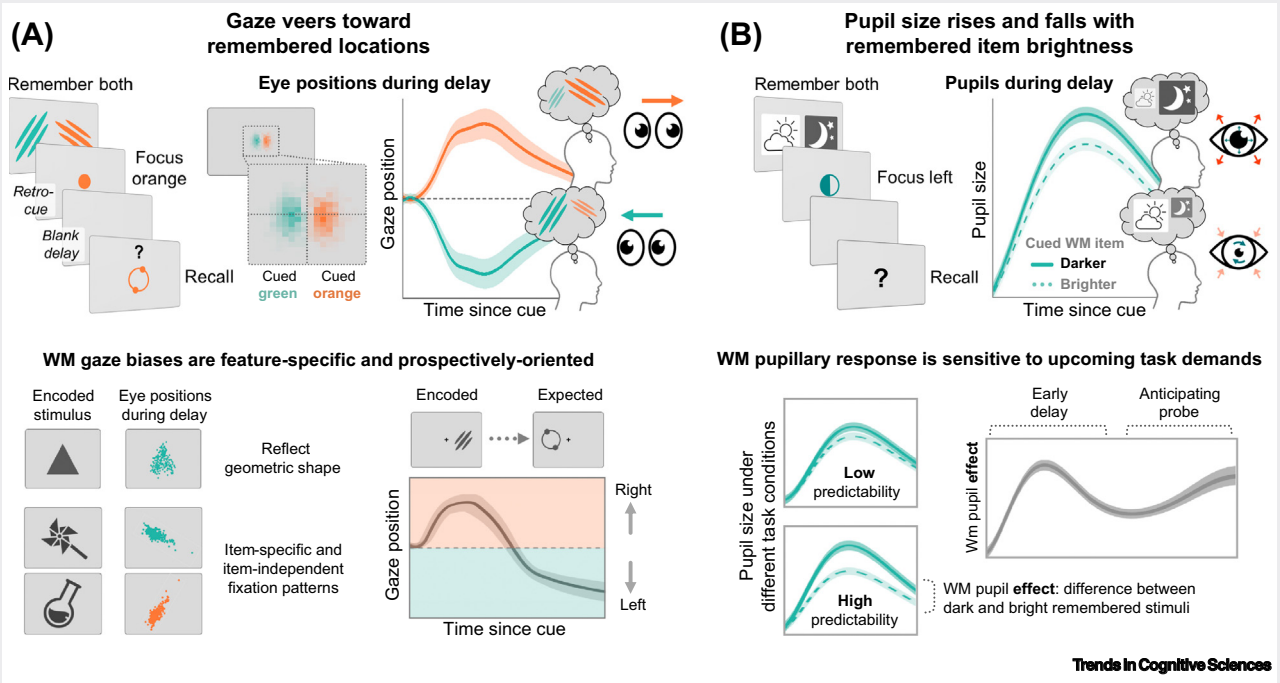

Figure I. Saccadic and pupillary working memory (WM) signatures under various task conditions. (A) Schematic depiction of a visual WM task and associated gaze effects (top). Eye positions during visual imagery or WM track remembered locations, specific object features, and abstract stimulus dimensions that are relevant to an upcoming probe (bottom) [120–122]. (B) Schematic depiction of a visual WM task and associated pupil effects. When one of two items is cued as relevant after encoding, remembering a darker item elicits a larger pupil dilation than a brighter item (top) [125,126]. This pupil effect is greater when the WM content is more likely to be tested with an upcoming probe (bottom) [127]. Both gaze and pupillary signatures reflect expectations about future task demands [122–124,127].

may be an active workspace [48,49,100] – both the earliest and latest stage of cortical visual WM processing – which receives visual input and outputs it in a format that is ready to support behavior.

Concluding remarks

While it has long been clear that WM engages flexible fronto-parietal function [2,4,9,13,18,26,29,32,33,38,57,67,101], a corresponding elasticity in visual cortical WM is now coming to light. Visual cortex was previously considered a passive, feedforward processing stage that parsed complex inputs into component parts. However, this view has been challenged in recent decades, first by evidence for gain modulations related to the behavioral relevance of stimuli during perception [102], and later by evidence that WM stimulus information is maintained in visual cortex when ongoing sensory stimulation is absent [34,35]. Now, converging evidence indicates that WM activity in visual cortex can meaningfully depart from iconic-like, sensory-evoked representations, suggesting an even more complex role in cognitive functions.

An intriguing possibility is that WM recruits whatever relevant apparatus is available across the nervous system to efficiently distribute the workload among structures that are equipped to achieve current goals [103]. Adaptive decision-making often sits at the precipice of sensation and WM [104], so visual WM may capitalize on sensory cortical representations because they are best-positioned to quickly interface with perception and action. Rather than sustain or weakly reinstate a faithful imprint of encoded content, however, WM may conjure transformed representations in line with behavioral goals and sensory conditions. In that sense, apparent distortions in sensory WM may not reflect corruptions but adaptive shifts to align WM with task demands. That does not mean that a veridical visual WM representation is impossible – especially in typical lab settings with no disruptions or interfering input. But there may be an upper limit on the precision such representations can achieve, how long they can be sustained, and how applicable they are in natural settings (see Outstanding questions). The growing appreciation for representational flexibility in sensory cortex heralds a paradigm shift away from asking where or how WM is stored, and toward testable theories of how behavioral imperatives modify sensory information content.

Outstanding questions

Are different sensory WM transformations supported by distinct association region functions? What are those functions?

To what extent is the fading sensory-evoked trace sustainable via either feedback or exogenous input? Is there an upper bound on WM resolution due to feedback?

Is feedback essential to all sensory WM reformatting, or might some transformations be intrinsic to the local sensory cortical representation?

Which apparent transformations reflect distinct mechanisms from each other versus the same phenomenon being measured in different ways? What are those mechanisms, and how might different analysis techniques lead to different conclusions about the same data?

Are apparent transformations to WM strategic, provoked by exogenous factors, or inherent to the generative nature of WM? To what extent are such transformations governed by traits that vary across individuals?

Which transformations reflect reformatting of sensory content versus transfer to different systems (e.g., long-term memory or motor planning)?

Is visual WM possible without visual cortical delay activity? Is visual cortex special when it comes to WM, or just one of many areas that can be engaged flexibly, depending on individual and situational factors?

Acknowledgments

This work was supported by National Institutes of Health awards R01EY036843 to A.K. and R01EY025872 to J.T.S. We thank members of the Kiyonaga lab for providing feedback on a draft of the manuscript – Ana Chkhaidze, Yueying Dong, Lana Gaspariani, Connie Xie, Sihan Yang, and Zhuojun Ying.

Declaration of interests

The authors declare no competing interests.

References

1. van Ede, F. and Nobre, A.C. (2023) Turning attention inside out: how working memory serves behavior. *Annu. Rev. Psychol.* 74, 137–165
2. Fuster, J.M. (2004) Upper processing stages of the perception–action cycle. *Trends Cogn. Sci.* 8, 143–145
3. Leavitt, M.L. *et al.* (2017) Sustained activity encoding working memories: not fully distributed. *Trends Neurosci.* 40, 328–346
4. Xu, Y. (2017) Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* 21, 794–815
5. Scimeca, J.M. *et al.* (2018) Reaffirming the sensory recruitment account of working memory. *Trends Cogn. Sci.* 22, 190–192
6. Gayet, S. *et al.* (2018) Visual working memory storage recruits sensory processing areas. *Trends Cogn. Sci.* 22, 189–190
7. Lorenc, E.S. and Sreenivasan, K.K. (2021) Reframing the debate: the distributed systems view of working memory. *Vis. Cogn.* 29, 416–424
8. Iamshchinina, P. *et al.* (2021) Essential considerations for exploring visual working memory storage in the human brain. *Vis. Cogn.* 29, 425–436
9. Roussy, M. *et al.* (2021) Neural substrates of visual perception and working memory: two sides of the same coin or two different coins? *Front. Neural Circuits* 15, 764177
10. Adam, K.C.S. *et al.* (2022) Evidence for, and challenges to, sensory recruitment models of visual working memory. In *Visual Memory* (Brady, T.F. and Bainbridge, W.A., eds), pp. 5–25, Routledge
11. Gomez-Lavin, J. (2021) Working memory is not a natural kind and cannot explain central cognition. *Rev. Philos. Psychol.* 12, 199–225
12. Nobre, A.C. and Stokes, M.G. (2019) Premembering experience: a hierarchy of time-scales for proactive attention. *Neuron* 104, 132–146
13. Christophel, T.B. *et al.* (2017) The distributed nature of working memory. *Trends Cogn. Sci.* 21, 111–124

14. Courtney, S.M. (2022) Working memory is a distributed dynamic process. *Cogn. Neurosci.* 13, 208–209
15. Dotson, N.M. *et al.* (2018) Feature-based visual short-term memory is widely distributed and hierarchically organized. *Neuron* 99, 215–226.e4
16. Serences, J.T. (2016) Neural mechanisms of information storage in visual short-term memory. *Vis. Res.* 128, 53–67
17. Buschman, T.J. (2021) Balancing flexibility and interference in working memory. *Annu. Rev. Vis. Sci.* 7, 367–388
18. D'Esposito, M. and Postle, B.R. (2015) The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142
19. Miller, E.K. *et al.* (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167
20. Suzuki, M. and Gottlieb, J. (2013) Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat. Neurosci.* 16, 98–104
21. Bettencourt, K.C. and Xu, Y. (2015) Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat. Neurosci.* 19, 150–157
22. Sheehan, T.C. and Serences, J.T. (2022) Attractive serial dependence overcomes repulsive neuronal adaptation. *PLoS Biol.* 20, e3001711
23. Chunharas, C. *et al.* (2022) An adaptive perspective on visual working memory distortions. *J. Exp. Psychol. Gen.* 151, 2300–2323
24. Hogendoorn, H. (2022) Perception in real-time: predicting the present, reconstructing the past. *Trends Cogn. Sci.* 26, 128–141
25. Woolgar, A. *et al.* (2011) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage* 56, 744–752
26. Sreenivasan, K.K. *et al.* (2014) Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89
27. Miller, E.K. *et al.* (2002) The prefrontal cortex: categories, concepts and cognition. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 357, 1123–1136
28. Riley, M.R. and Constantinidis, C. (2016) Role of prefrontal persistent activity in working memory. *Front. Syst. Neurosci.* 9, 181
29. Fusi, S. *et al.* (2016) Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74
30. Wang, Y. *et al.* (2023) Long-range functional connections mirror and link microarchitectural and cognitive hierarchies in the human brain. *Cereb. Cortex* 33, 1782–1798
31. Manea, A.M. *et al.* (2022) Intrinsic timescales as an organizational principle of neural processing across the whole rhesus macaque brain. *eLife* 11, e75540
32. Parthasarathy, A. *et al.* (2017) Mixed selectivity morphs population codes in prefrontal cortex. *Nat. Neurosci.* 20, 1770–1779
33. Lara, A.H. and Wallis, J.D. (2015) The role of prefrontal cortex in working memory: a mini review. *Front. Syst. Neurosci.* 9, 173
34. Harrison, S.A. and Tong, F. (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635
35. Serences, J.T. *et al.* (2009) Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207–214
36. Pasternak, T. and Greenlee, M.W. (2005) Working memory in primate sensory systems. *Nat. Rev. Neurosci.* 6, 97–107
37. Ester, E.F. *et al.* (2013) A neural measure of precision in visual working memory. *J. Cogn. Neurosci.* 25, 754–761
38. Murray, J.D. *et al.* (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 114, 394–399
39. Myers, N.E. *et al.* (2015) Testing sensory evidence against mnemonic templates. *eLife* 4, e09000
40. Wolff, M.J. *et al.* (2020) Drifting codes within a stable coding scheme for working memory. *PLoS Biol.* 18, e3000625
41. Rademaker, R.L. *et al.* (2019) Coexisting representations of sensory and mnemonic information in human visual cortex. *Nat. Neurosci.* 22, 1336–1344
42. van Kerkoerle, T. *et al.* (2017) Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* 8, 13804
43. Yiling, Y. *et al.* (2024) Dynamic fading memory and expectancy effects in the monkey primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2314855121
44. Huang, J. *et al.* (2024) Neuronal representation of visual working memory content in the primate primary visual cortex. *Sci. Adv.* 10, eadk3953
45. Yiling, Y. *et al.* (2024) Joint encoding of stimulus and decision in monkey primary visual cortex. *Cereb. Cortex* 34, bhad420
46. Jonikaitis, D. *et al.* (2025) Robust encoding of stimulus–response mapping by neurons in visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 122, e2408079122
47. Kok, P. *et al.* (2013) Prior expectations bias sensory representations in visual cortex. *J. Neurosci.* 33, 16275–16284
48. Roelfsema, P.R. and de Lange, F.P. (2016) Early visual cortex as a multiscale cognitive blackboard. *Annu. Rev. Vis. Sci.* 2, 131–151
49. Zhang, N. and Xu, N. (2022) Reshaping sensory representations by task-specific brain states: toward cortical circuit mechanisms. *Curr. Opin. Neurobiol.* 77, 102628
50. Henderson, M.M. *et al.* (2025) Dynamic categorization rules alter representations in human visual cortex. *Nat. Commun.* 16, 3459
51. Micou, C. and O'Leary, T. (2023) Representational drift as a window into neural and behavioural plasticity. *Curr. Opin. Neurobiol.* 81, 102746
52. Ratzon, A. *et al.* (2024) Representational drift as a result of implicit regularization. *eLife* 12, RP90069
53. Panichello, M.F. and Buschman, T.J. (2021) Shared mechanisms underlie the control of working memory and attention. *Nature* 592, 601–605
54. Merrikhi, Y. *et al.* (2017) Spatial working memory alters the efficacy of input to visual cortex. *Nat. Commun.* 8, 15041
55. Park, S. and Serences, J.T. (2022) Relative precision of top-down attentional modulations is lower in early visual cortex compared to mid- and high-level visual areas. *J. Neurophysiol.* 127, 504–518
56. Xu, Y. (2023) Parietal-driven visual working memory representation in occipito-temporal cortex. *Curr. Biol.* 33, 4516–4523.e5
57. Xu, Y. (2024) The human posterior parietal cortices orthogonalize the representation of different streams of information concurrently coded in visual working memory. *PLoS Biol.* 22, e3002915
58. Favila, S.E. *et al.* (2022) Perception and memory have distinct spatial tuning properties in human visual cortex. *Nat. Commun.* 13, 5864
59. Breedlove, J.L. *et al.* (2020) Generative feedback explains distinct brain activity codes for seen and mental images. *Curr. Biol.* 30, 2211–2224.e6
60. Woodry, R. *et al.* (2025) Feedback scales the spatial tuning of cortical responses during both visual working memory and long-term memory. *J. Neurosci.* 45, e0681242025
61. Wandell, B.A. and Winawer, J. (2015) Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* 19, 349–357
62. Yan, C. *et al.* (2023) Categorical working memory codes in human visual cortex. *NeuroImage* 274, 120149
63. Chunharas, C. *et al.* (2025) A gradual transition toward categorical representations along the visual hierarchy during working memory, but not perception. *eLife* 14, RP103347
64. Kwak, Y. and Curtis, C.E. (2022) Unveiling the abstract format of mnemonic representations. *Neuron* 110, 1822–1828.e5
65. Panichello, M.F. *et al.* (2019) Error-correcting dynamics in visual working memory. *Nat. Commun.* 10, 3366
66. Lee, S.-H. *et al.* (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat. Neurosci.* 16, 997–999
67. Badre, D. and Nee, D.E. (2018) Frontal cortex and the hierarchical control of behavior. *Trends Cogn. Sci.* 22, 170–188
68. Wutz, A. *et al.* (2018) Different levels of category abstraction by different dynamics in different prefrontal areas. *Neuron* 97, 716–726.e8
69. Birman, D. and Gardner, J.L. (2016) Parietal and prefrontal: categorical differences? *Nat. Neurosci.* 19, 5–7

70. Bae, G.-Y. (2021) Neural evidence for categorical biases in location and orientation representations in a working memory task. *NeuroImage* 240, 118366
71. Wolff, M.J. *et al.* (2017) Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871
72. Lorenc, E.S. *et al.* (2018) Flexible coding of visual working memory representations during distraction. *J. Neurosci.* 38, 5267–5276
73. Hallenbeck, G.E. *et al.* (2021) Working memory representations in visual cortex mediate distraction effects. *Nat. Commun.* 12, 4714
74. Henderson, M.M. *et al.* (2022) Flexible utilization of spatial- and motor-based codes for the storage of visuo-spatial information. *eLife* 11, e75688
75. Boettcher, S.E.P. *et al.* (2021) Output planning at the input stage in visual working memory. *Sci. Adv.* 7, eabe8212
76. Trentin, C. *et al.* (2024) Action similarity warps visual feature space in working memory. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2413433121
77. Myers, N.E. *et al.* (2017) Prioritizing information during working memory: beyond sustained internal attention. *Trends Cogn. Sci.* 21, 449–461
78. Ehrlich, D.B. and Murray, J.D. (2022) Geometry of neural computation unifies working memory and planning. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115610119
79. Bae, G.-Y. and Chen, K.-W. (2024) EEG decoding reveals task-dependent recoding of sensory information in working memory. *NeuroImage* 297, 120710
80. Duan, Z. and Curtis, C.E. (2024) Visual working memories are abstractions of percepts. *eLife* 13, RP94191
81. Norman, K.A. and O'Reilly, R.C. (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol. Rev.* 110, 611–646
82. van Loon, A.M. *et al.* (2018) Current and future goals are represented in opposite patterns in object-selective cortex. *eLife* 7, e38677
83. Yu, Q. *et al.* (2020) Different states of priority recruit different neural representations in visual working memory. *PLoS Biol.* 18, e3000769
84. Libby, A. and Buschman, T.J. (2021) Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* 24, 715–726
85. Wan, Q. *et al.* (2022) Priority-based transformations of stimulus representation in visual working memory. *PLoS Comput. Biol.* 18, e1009062
86. Tian, Z. *et al.* (2024) Mental programming of spatial sequences in working memory in the macaque frontal cortex. *Science* 385, eadp6091
87. Zaksas, D. and Pasternak, T. (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742
88. Lara, A.H. and Wallis, J.D. (2014) Executive control processes underlying multi-item working memory. *Nat. Neurosci.* 17, 876–883
89. Kuzmina, E. *et al.* (2024) Neuronal travelling waves explain rotational dynamics in experimental datasets and modelling. *Sci. Rep.* 14, 3566
90. Lorenc, E.S. *et al.* (2021) Distraction in visual working memory: resistance is not futile. *Trends Cogn. Sci.* 25, 228–239
91. Yu, X. *et al.* (2023) Good-enough attentional guidance. *Trends Cogn. Sci.* 27, 391–403
92. Ballard, D.H. *et al.* (1995) Memory representations in natural tasks. *J. Cogn. Neurosci.* 7, 66–80
93. Draschkow, D. *et al.* (2021) When natural behavior engages working memory. *Curr. Biol.* 31, 869–874.e5
94. Chota, S. *et al.* (2023) A matter of availability: sharper tuning for memorized than for perceived stimulus features. *Cereb. Cortex* 33, 7608–7618
95. Somai, R.S. *et al.* (2020) Evidence for the world as an external memory: a trade-off between internal and external visual memory storage. *Cortex* 122, 108–114
96. Miller, J.A. *et al.* (2020) Prioritized verbal working memory content biases ongoing action. *J. Exp. Psychol. Hum. Percept. Perform.* 46, 1443–1457
97. Gayet, S. *et al.* (2013) Information matching the content of visual working memory is prioritized for conscious access. *Psychol. Sci.* 24, 2472–2480
98. Jonikaitis, D. *et al.* (2023) Dissociating the contributions of frontal eye field activity to spatial working memory and motor preparation. *J. Neurosci.* 43, 8681–8689
99. Weber, S. *et al.* (2024) Working memory signals in early visual cortex are present in weak and strong imagers. *Hum. Brain Mapp.* 45, e26590
100. Logie, R.H. (2003) Spatial and visual working memory: a mental workspace. In *Psychology of Learning and Motivation* (Vol. 42) (Irwin, D.E. and Ross, B.H., eds), pp. 37–78, Academic Press
101. Miller, J.A. and Constantinidis, C. (2024) Timescales of learning in prefrontal cortex. *Nat. Rev. Neurosci.* 25, 597–610
102. Gandhi, S.P. *et al.* (1999) Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3314–3319
103. Postle, B.R. (2006) Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38
104. Kumle, L. *et al.* (2025) Sensorimnemonic decisions: choosing memories versus sensory information. *Trends Cogn. Sci.* 29, 311–313
105. Liu, Z. *et al.* (2025) Emergence of categorical representations in parietal and ventromedial prefrontal cortex across extended training. *J. Neurosci.* 45, e1315242024
106. Riley, M.R. *et al.* (2018) Anterior-posterior gradient of plasticity in primate prefrontal cortex. *Nat. Commun.* 9, 3790
107. Miller, J.A. *et al.* (2022) Long-term learning transforms prefrontal cortex representations during working memory. *Neuron* 110, 3805–3819.e6
108. Tünçok, E. *et al.* (2025) Opposite asymmetry in visual perception of humans and macaques. *Curr. Biol.* 35, 681–687.e4
109. Van Essen, D.C. *et al.* (2001) Mapping visual cortex in monkeys and humans using surface-based atlases. *Vis. Res.* 41, 1359–1378
110. Jorstad, N.L. *et al.* (2023) Transcriptomic cytoarchitecture reveals principles of human neocortex organization. *Science* 382, eadf6812
111. Deco, G. *et al.* (2021) Rare long-range cortical connections enhance human information processing. *Curr. Biol.* 31, 4436–4448.e5
112. Klink, P.C. *et al.* (2021) Population receptive fields in nonhuman primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. *eLife* 10, e67304
113. Hubel, D.H. and Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154
114. Barlow, H. (2001) Redundancy reduction revisited. *Netw. Bristol Engl.* 12, 241–253
115. Duncker, L. *et al.* (2020) Organizing recurrent network dynamics by task-computation to enable continual learning. In *Advances in Neural Information Processing Systems* (33), pp. 14387–14397, NeurIPS
116. Brissenden, J.A. *et al.* (2021) Stimulus-specific visual working memory representations in human cerebellar lobule VIIb/VIIIa. *J. Neurosci.* 41, 1033–1045
117. Rahmati, M. *et al.* (2023) Mnemonic representations in human lateral geniculate nucleus. *Front. Behav. Neurosci.* 17, 1094226
118. Rahmati, M. *et al.* (2020) Spatially specific working memory activity in the human superior colliculus. *J. Neurosci.* 40, 9487–9495
119. Peysakhovich, B. *et al.* (2024) Primate superior colliculus is causally engaged in abstract higher-order cognition. *Nat. Neurosci.* 27, 1999–2008
120. van Ede, F. *et al.* (2019) Human gaze tracks attentional focusing in memorized visual space. *Nat. Hum. Behav.* 3, 462–470
121. Laeng, B. *et al.* (2014) Scrutinizing visual images: the role of gaze in mental imagery and memory. *Cognition* 131, 263–283
122. Linde-Domingo, J. and Spitzer, B. (2024) Geometry of visuo-spatial working memory information in miniature gaze patterns. *Nat. Hum. Behav.* 8, 336–348

123. van Ede, F. *et al.* (2021) Looking ahead in working memory to guide sequential behaviour. *Curr. Biol.* 31, R779–R780
124. Liu, B. *et al.* (2024) Jointly looking to the past and the future in visual working memory. *eLife* 12, RP90874
125. Hustá, C. *et al.* (2019) The pupillary light response reflects visual working memory content. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 1522–1528
126. Zokaei, N. *et al.* (2019) Modulation of the pupillary response by the content of visual working memory. *Proc. Natl. Acad. Sci. U. S. A.* 116, 22802–22810
127. Dong, Y. and Kiyonaga, A. (2024) Ocular working memory signals are flexible to behavioral priority and subjective imagery strength. *J. Neurophysiol.* 132, 162–176
128. Mostert, P. *et al.* (2018) Eye movement-related confounds in neural decoding of visual working memory representations. *eNeuro* 5 ENEURO.0401-17.2018
129. Postle, B.R. *et al.* (2006) The selective disruption of spatial working memory by eye movements. *Q. J. Exp. Psychol.* 59, 100–120
130. Ebitz, R.B. and Moore, T. (2019) Both a gauge and a filter: cognitive modulations of pupil size. *Front. Neurol.* 9, 1190