Routledge
Taylor & Francis Group

Check for updates

Essential considerations for exploring visual working memory storage in the human brain

Polina Iamshchinina[a,b], Thomas B. Christophel[c,d], Surya Gayet[e] and Rosanne L. Rademaker[e,f,g]

[a]Department of Education and Psychology, Freie Universität Berlin, Germany; [b]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany; [c]Bernstein Center for Computational Neuroscience and Berlin Center for Advanced Neuroimaging and Clinic for Neurology, Charité Universitätsmedizin, Freie Universität Berlin, Humboldt Universität zu Berlin, Berlin Institute of Health, Berlin, Germany; [d]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany; [e]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands; [f]Department of Psychology, University of California, La Jolla, CA, USA; [g]Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, Frankfurt, Germany

ABSTRACT

Visual working memory (VWM) relies on a distributed cortical network. Yet, the extent to which individual cortical areas, like early visual cortex and intraparietal sulcus, are essential to VWM storage remains debated. Here, we reanalyze key datasets from two independent labs to address three topics at the forefront of current-day VWM research: Resiliency of mnemonic representations against visual distraction, the role of attentional priority in memory, and brain–behavior relationships. By utilizing different analysis approaches, each designed to test different aspects of mnemonic coding, our results provide a comprehensive perspective on the role of early visual and intraparietal areas. We emphasize the importance of analysis choices, and how a thorough understanding of the principles they test is crucial for unraveling the distributed mechanisms of VWM. Consequently, we caution against the idea of a singular essential storage area, which could limit our comprehension of the VWM system.

Introduction

Recent evidence points towards an understanding of visual working memory (VWM) as a distributed network of cortical regions that operate in concert to temporarily retain information about visual stimuli (Christophel et al., 2017; D'Esposito & Postle, 2015; Lee & Baker, 2016). Long standing theory (e.g., Fuster, 1995; 1997) suggests that for this purpose, distributed cortical stores provide a multitude of neural representations ranging from low-level sensory imprints in sensory regions, to abstract categorical contents and prospective action plans in more anterior regions. Contrary to this distributed view, a series of recent articles (Bettencourt & Xu, 2016; Xu, 2017; 2018; 2020) explores the role of two specific cortical sites, early visual cortex (EVC) and Intraparietal Sulcus (IPS), and their necessity for VWM storage. This work proposes that posterior parietal cortex (specifically, IPS) is the main and essential site of mnemonic retention and rejects the notion that EVC is essential for working memory storage.

Throughout the years, a large number of lesion studies in non-human primates and patient populations attempted to identify which region(s) of the brain are essential for VWM storage. Early work using this modular approach concluded that working memory was an exclusive attribute of the prefrontal cortex (Jacobsen, Elder, & Haslerud, 1936). Later work, however, implicated prefrontal, parietal, temporal and occipital areas (Farah, 1984, 1988; Fuster & Jervey, 1981; Ghent et al., 1962; Gross & Weiskrantz, 1964; Malmo, 1942; Petrides, 1995; Warrington & Shallice, 1969). Using single-cell recordings in non-human primates, several groups have asked whether mnemonic activity could be sustained under different types of sensory distraction (Miller et al., 1993; Miller et al., 1996; Woloszyn & Sheinberg, 2009). Initial studies found that such distracting inputs left behavioral reports relatively intact, but interrupted mnemonic activity in inferior temporal regions (Miller et al., 1993), implying the unimportance of these regions for information maintenance

per se. With more sensitive methods, however, it was later demonstrated that distractors did not fully disrupt, but merely diminished the amount of recoverable mnemonic activity in these regions (Woloszyn & Sheinberg, 2009), calling these earlier claims into question. The resiliency of behavioral performance in these distractor studies is a notable feature of VWM.

The latest iteration of this recurring debate surrounding essential storage sites, centers on whether mnemonic traces recovered from patterns of delay-period fMRI data are affected by distracting visual inputs (Bettencourt & Xu, 2016; Ester et al., 2016; Gayet et al., 2018; Kiyonaga et al., 2017a; Lorenc et al., 2018; Rademaker et al., 2019; Scimeca et al., 2018; Xu, 2017; 2018). In the target article (Xu, 2020), a case is made that early visual areas are nonessential to VWM storage, based on a number of criteria: To be considered essential, an area must represent the contents of working memory. Moreover, storage in an essential area will not be diminished as long as people's behavior is unaffected. By extension, neural representations in an essential area are expected to vary in unison with behavioral performance. In other words, factors that negatively impact behavior – such as visual distraction, impoverished stimulus encoding, or inter-subject variability – will also negatively affect mnemonic representations in an "essential" area (Xu, 2020).

The search for cortical areas essential to VWM storage has been a fruitful one, with lesion, single-cell, and human neuroimaging studies implicating multiple regions in multiple ways. Which leads to the inevitable question: Is there really *one* essential storage site? And – given the multitude of behavioral goals supported by VWM, combined with diverse strategies to achieve them – would it even make sense for VWM to rely on a single essential storage site? Instead of assuming a monolithic view of VWM and searching the brain for a single essential store, we consider VWM in light of the behavioral goals it subserves and from which its neural implementation follows (Boettcher et al., 2021). For the successful completion of behavioral tasks, VWM representations should be highly robust against external factors such as distraction (e.g., from visual inputs; Lorenc et al., 2021; Rademaker et al., 2015; Wildegger et al., 2015) and should be relatively stable over time (Rademaker et al., 2018; Shin et al., 2017). Having multiple representations of the same mnemonic contents distributed across multiple cortical locations allows to meet these functional demands – not only to ensure robust retention under many possible forms of interference, but also for the flexible application of the mnemonic contents at any moment in case of changing task demands.

Here, we reanalyze neuroimaging data from two independent labs (Christophel et al., 2018; Rademaker et al., 2019) for two distinct purposes. First, we evaluate several novel results against the criteria put forth in Xu (2020). These criteria address a strictly modular view on VWM storage by asking whether EVC or IPS is an "essential" storage site. Under this framework, our results by and large favor EVC, which proves largely robust against distraction, and covaries with behavioral performance. The second, and *most important purpose* of our reanalysis is to demonstrate the impact of analysis choices on the conclusions that might be drawn from any given data set. There are many different sensible ways of analyzing the same data, which all tap into different aspects of VWM-related activity. One approach isn't necessarily better than the other, but all interpretations are constrained by their own sets of underlying assumptions and limitations. As such, the unconstrained application of criteria to interpret findings across different studies is problematic, considering that different studies (often purposefully) use vastly different analysis approaches. In sum, before drawing broad conclusions about VWM, it is important to carefully consider what is being tested, how it is being tested, and what each approach *can* and *cannot* tell us.

Methods and results: A reanalysis of data from two independent labs

For our first reanalysis, we investigate to what extent mnemonic representations in EVC and IPS are resistant to visually presented distractor stimuli. We reanalyzed data from Rademaker et al. (2019) where various visual distractors were shown while participants remembered an orientation (Figure 1(a,b)). In a first experiment, distractors were either randomly oriented gratings or Fourier filtered noise stimuli, phase-reversing on the screen for 11 s during the memory delay. In the original analysis, there was no consistent evidence that visual distractors reduced the amount of decodable memory information in EVC or IPS compared to a delay without distractors,
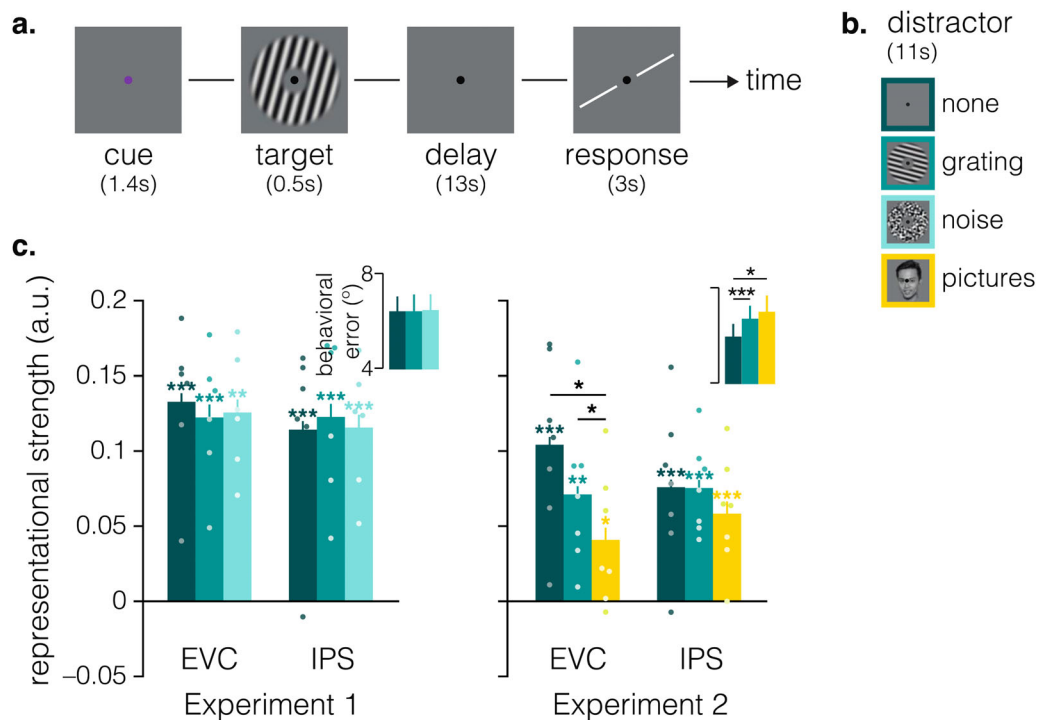
Figure 1. Task & reanalysis of Rademaker et al. (2019). (a) Schematic depiction of the memory task. Participants remembered a random target orientation over a 13-second delay, after which they rotated a white line (via button presses) to report the remembered orientation. Behavioral error was calculated as response orientation minus target orientation (in °). A color change at the start of each trial indicated one of three delay-period conditions. (b) During the delay of Experiment 1 there was either no distraction (dark-teal); an 11-second grating distractor (mid-teal); or an 11-second Fourier filtered noise distractor (light-teal). Experiment 2 was similar, except that pictures of faces and gazebos were used (yellow) instead of the filtered noise. Photo used with permission (S. Itthipuripat). (c) The strength of mnemonic representations is plotted for different distractor conditions, and was calculated using a model trained on data from the delay period (via a leave-one-out cross-validation approach). Across large retinotopic regions of EVC (V1–V4) and IPS (IPS0–IPS3), the strength of mnemonic representations (quantified as a projected vector mean of channel reconstructions, following an inverted encoding modeling approach – see Rademaker et al., 2019) did not differ as a function of distractor condition in Experiment 1 (left panel). This result dovetails participants' behavioral errors, which were also unaffected by distractors (see insert). In Experiment 2 (right panel), distractors caused a drop in the strength of mnemonic representations in EVC, but not in IPS. Results from EVC tracked behavior, with participants performing worse in the presence of distractors (see insert). Only visually responsive voxels were included in these analyses. Asterisks indicate significance at *$p \leq 0.05$, **$p \leq 0.01$, and ***$p \leq 0.001$ (uncorrected for multiple comparisons). Black asterisks indicate post-hoc differences in representational strength between distractor conditions, following a significant non-parametric one-way repeated-measures within-subjects ANOVA. Colored asterisks indicate significance according to a one-sided randomization test comparing representational strength in each condition and ROI to zero. Dots indicate individual participants (N=6 for Experiment 1; N=7 for Experiment 2). Error bars represent ±1 within-subject SEM around the mean. For more methodological details, see Rademaker et al. (2019).

nor did distractors impact behavioral performance (see Figure 1 in Rademaker et al., 2019). In a second experiment, distraction was increased by flickering distractors on and off, and by including pictures of faces and gazebos (instead of the noise distractor used in the first experiment). In the original analysis, visual distractors reduced the amount of decodable information in EVC (but not IPS), and negatively impacted behavioral performance (see Figure 3 in Rademaker et al., 2019). Notably, both EVC and IPS were analyzed on the basis of smaller retinotopically mapped regions of interest (or "ROIs"). It has been

argued by Xu (2020) that combining these smaller ROIs would allow for a better assessment of distraction effects per cortical area of interest. To test this, we combined retinotopic areas V1–V4 into a large EVC ROI. Similarly, we combined retinotopic areas IPS0–IPS3 into a large IPS ROI. We will use such combined EVC and IPS ROIs throughout.

In addition to a memory task, the Rademaker et al. (2019) study included an independent "mapping" task where participants directly viewed orientation stimuli presented on the screen. This provides us with a choice between two widely employed model training

approaches: We can train on the memory-delay data itself using a leave-one-out cross-validation approach. This model training approach capitalizes on *any* signal differentiating remembered orientation representations during the delay. Alternatively, we can train on sensory-driven responses, which only captures mnemonic signals that are similar to the activity measured in the "mapping" task. The results of Rademaker et al. (2019) already illustrated that this distinction matters. Mnemonic representations in IPS were successfully retrieved by a model trained on delay-period data, but not by a model trained on the independent mapping data (implying a lack of generalizability from sensory-driven to mnemonic patterns of activity).

We start by training our model on data from the memory delay. This model training approach closely mimics that of an earlier study with a similar paradigm, where strong interference effects in EVC were observed with distraction (Bettencourt & Xu, 2016). This approach also allows us to evaluate interference effects in IPS – an area that yields little decodable information under a sensory-driven training regime (see Supplementary Figure 13 in Rademaker et al., 2019). In Experiment 1, where *no behavioral differences* were observed between distractor conditions ($F_{(2,10)} = 0.044$; $p = 0.943$), neither EVC ($F_{(2,10)} = 0.52$; $p = 0.62$) nor IPS ($F_{(2,10)} = 0.155$; $p = 0.843$) showed a drop in representational strength with distractors during the delay (Figure 1(c), left panel). This is a clear example of the robustness of VWM, which appears unaffected by the 11-second-long influx of visual input. In Experiment 2, where we did observe a *behavioral deficit* with distraction during the delay ($F_{(2,12)} = 10.154$; $p < 0.001$), a corresponding drop in representational strength was observed in EVC ($F_{(2,12)} = 5.936$; $p = 0.008$), but not in IPS ($F_{(2,12)} = 0.919$; $p = 0.418$) (Figure 1(c), right panel). Considering Experiments 1 and 2 together, we conclude that both EVC and IPS represent mnemonic information during visual distraction. While IPS demonstrates the strongest resiliency against distraction, EVC best tracks recall performance. Importantly, these conclusions rest on an analysis approach that differentiates between remembered orientations in any representational format.

What happens when we narrow our analysis approach to include only sensory-like VWM representations, by using independent "mapping" data for model training? In Experiment 1, we now observe a drop in representational strength with distraction in EVC ($F_{(2,10)} = 7.226$; $p = 0.007$) despite the absence of a measurable behavioral deficit. IPS showed no difference between distractor conditions ($F_{(2,10)} = 2.639$; $p = 0.108$), which is unsurprising given the lack of sensory-like information in IPS (the IPS representation was significantly above chance only in the no-distractor condition, $p = 0.008$). In Experiment 2, IPS representations also failed to reach the threshold for significance (all $p > 0.072$), with no differences between distractor conditions ($F_{(2,12)} = 0.921$; $p = 0.436$). In EVC a drop in representational strength with distraction was still observed ($F_{(2,12)} = 17.828$; $p < 0.001$), akin to the drop observed in behavior. Considering Experiments 1 and 2 together, we conclude that only EVC represents mnemonic information in a "sensory-like" format during visual distraction. These EVC representations are diminished during visual distraction, even in the absence of a measurable behavioral deficit. Importantly, these conclusions hold for the specific representational format that was evoked by the sensory responses our model was trained on. In this case, those are responses evoked by directly viewed full-contrast orientations that were actively attended (via an orthogonal task).

Which analysis is right? Should we only consider representations in a format similar to that of a sensory driven response? If so, our analysis will likely favor areas specialized for processing sensory inputs, such as EVC. Or should we instead remain agnostic to the different ways in which memoranda could be represented throughout the brain? These questions have no unequivocal answer and depend on the assumptions and definitions that experimenters adhere to.

For our second reanalysis, we investigated to what extent EVC and IPS represent memories that are not immediately relevant to behavior and might therefore be held in mind in an unattended state. For this purpose, we reanalyzed data from Christophel et al. (2018), where the amount of attention allocated to different items in memory was manipulated (see Figure 2(a)): After sequentially presenting two different orientations, one gained immediate relevance as participants were cued to report it after a 7-second delay – this was the attended memory item (AMI). The uncued item could still be probed at a later timepoint in the trial and was therefore not
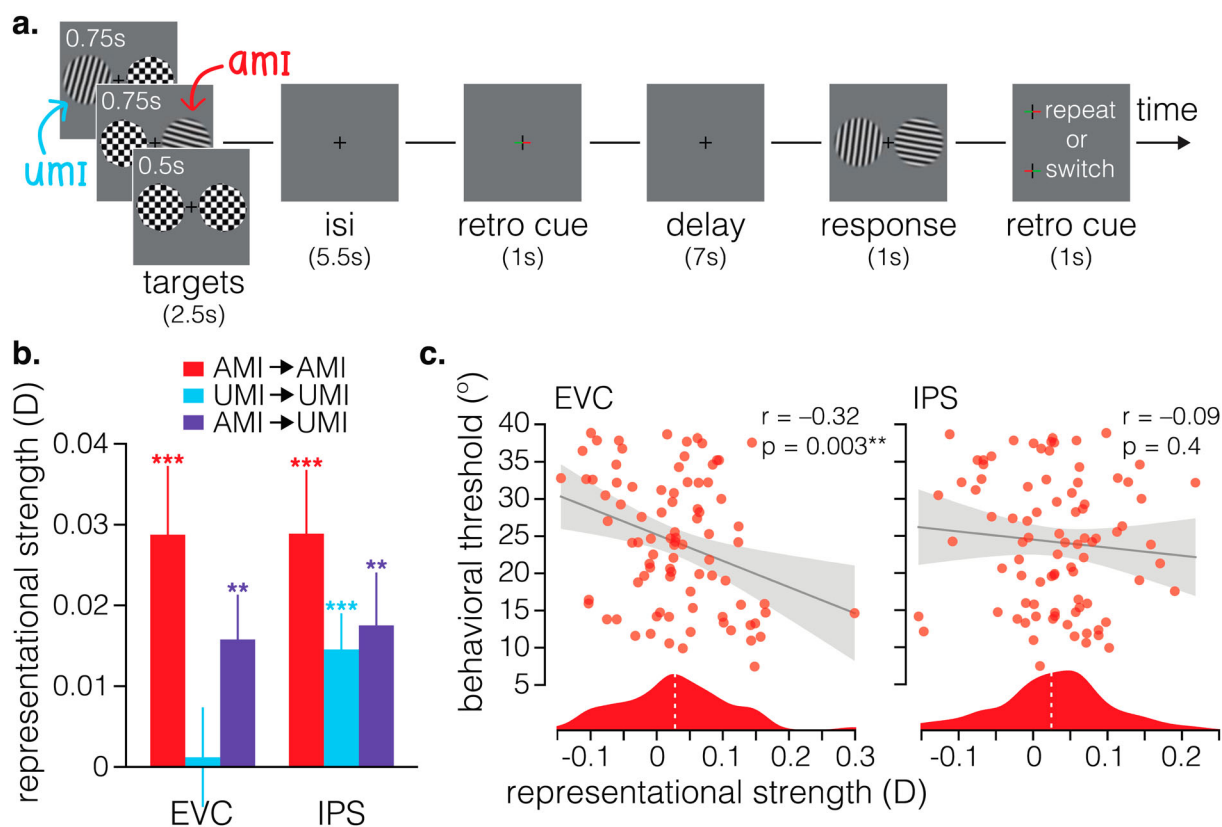
Figure 2. Task & reanalysis of Christophel et al. (2018). (a) Schematic depiction of the memory task. Participants are shown two subsequently presented oriented gratings flanked by masks, followed by a set of two more masks (all 0.25s apart). Both target orientations are initially remembered, and a first retro-cue indicates (via a red horizontal line) which of the two needs to be reported after a 7-second delay. This cued target item is prioritized in memory (AMI = Attended Memory Item) until participants report a clockwise or counterclockwise orientation change after the delay. Following this response, participants see a second retro-cue indicating which of the two target orientations will need to be reported next (the rest of the trial is not shown here for simplicity). Because there is a 50% chance that participants will need to "switch" and report the orientation of the previously uncued item, this item cannot be fully discarded. Rather, this item is assumed to receive less priority during the delay (UMI = Unattended Memory Item). (b) Results from the original paper, using within-condition model training, showed AMI representations (red) in both EVC and IPS, and UMI representations (light blue) only in IPS. Note: these results (in red and light blue) are replotted from Figure 2b in Christophel et al. (2018). To further investigate UMI representations, our reanalysis used activity patterns evoked by the AMI to "train" the multivariate model, as these could carry more robust information than activity patterns evoked by the UMI. Now, activity patterns measured during the delay *do* reveal a UMI representation (purple) in both EVC and IPS. Hence, this approach results in more reliable model estimation (for more analysis details, see Allefeld & Haynes, 2014). Asterisks indicate a representational strength that was significantly above zero as indicated by one-tailed one-sample t-tests, with **$p \leq 0.01$, and ***$p \leq 0.001$ (uncorrected for multiple comparisons). Error bars indicate ±1 SEM (c) Across participants, there is a relationship between behavioral performance on the memory task, and the strength with which the AMI is represented in EVC (left panel), but not in IPS (right panel). Behavioral performance (on the y-axes) is quantified as the threshold (in degree) resulting from an adaptive staircase procedure that pins participants' performance level at ~80% correct. A larger behavioral threshold (i.e., larger difference between the target orientation and probe orientation) indicates poorer performance. Each dot indicates an individual subject. Representational strength of the AMI (on the x-axes) is shown here for all N=87 individual participants. To better visualize how representational strength is distributed across participants, a kernel density plot (normal distribution with bandwidth of 0.02) is added to the x-axis. The white dotted line is the mean representational strength (matching the height of the red bars in (b)). The grey line and grey error area indicate the slope and the 95% CI of the brain-behavior relationship. For more methodological details, see Christophel et al. (2018). Note: When excluding the single participant with exceptionally high EVC representational strength (D ~ = 0.03), the brain-behavior correlation in EVC remains (r = −0.297; $p = 0.005$).

completely discarded – this was the unattended memory item (UMI). In the original paper, it was shown that patterns of delay activity in IPS contained information about both the AMI and the UMI. By contrast, patterns of delay activity in EVC only contained

information about the AMI, but not the UMI – despite participants' ability to successfully report the orientation of the UMI (see Figure 2(b) in Christophel et al., 2018; replotted here in red and blue in Figure 2(b)). In other words, the representation of a successfully

remembered item was not detectable in EVC when it was not imminently relevant. The authors nevertheless cautioned against prematurely interpreting null results, and already noted how analysis choices in prior work (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; LaRocque, Riggall, Emrich, & Postle, 2017) had limited the potential to find representations for unattended items.

As with any dataset, the Christophel et al. (2018) study lends itself to multiple analysis approaches. The original work used data from within each condition to train and test their multivariate model. This means that AMI representations were uncovered by training on patterns of activity evoked by attended mnemonic contents ("AMI→AMI"). To test for UMI representations, the model was trained on patterns of activity evoked by unattended mnemonic contents ("UMI→UMI"). However, it is reasonable to assume that UMI signals are weaker than AMI signals, and that information about the UMI went undetected when training the model on UMI signals. This would be particularly likely in areas where UMI responses were much weaker, but still evoked patterns similar to AMI responses. In the current analysis, we looked for UMI representations using the more robust AMI representation to "train" the model. Importantly, our new analysis uncovered UMI information in both EVC ($t_{(86)} = 2.844$; $p = 0.006$) and IPS ($t_{(86)} = 2.671$; $p = 0.009$) (Figure 2(b), purple). On the basis of the present reanalysis, we therefore conclude that both EVC and IPS maintain VWM representations for items that are unattended. The take home message from contrasting different model training approaches as we do here, is that absence of evidence does not equal evidence of absence. When it comes to uncovering VWM representations, some analysis approaches may be more or less powerful than others – depending on specifics of the experiment.

For our third reanalysis, we wanted to investigate the relationship between representational strength in the brain, and behavioral performance. In Figure 1, we already allude to such a brain–behavior relationship, with changes in behavioral performance mirrored by representational changes in EVC across two experiments. In this next step, we look at brain–behavior relationships at two additional levels of analysis, namely, across participants and within participants (across trials). In search of a brain–behavior relationship, it is important to consider factors that

might impact both neural and behavioral measures in a similar direction. For example, some participants might be more engaged (or caffeinated) than others. Individual participants might be intermittently drowsy or make eye blinks, potentially affecting neural representations and behavior in unison from trial-to-trial. Importantly, factors mediating the brain–behavior relationship at one level of analysis may not be able to do so at another level of analysis, and vice versa. Thus, using multiple levels of analysis can be a powerful approach to obtain converging lines of evidence in support of a brain–behavior relationship.

First, we asked whether participants who were worse at a VWM task had weaker neural representations of memory items. While such a relationship has been reported previously (Ester, Anderson, Serences, & Awh, 2013), the data from Christophel et al. (2018) lends itself particularly well to investigate this question across N=87 individual participants (Figure 2(c)). Indeed, performance on the working memory task was correlated with the representational strength of the attended memory item in EVC ($r_{(85)} = -0.32$, $p = 0.003$, left panel), but not in IPS ($r_{(85)} = -0.09$, $p = 0.4$, right panel). Second, we asked whether participants' neural representations displayed larger biases on trials with larger behavioral errors. The Rademaker et al. data (2019) from Experiment 1 are well suited to address this question, as large numbers of trials were collected from each participant. For this analysis each individual participant's data was split into three bins based on the magnitude of their behavioral errors (Figure 3(a)). For trials with counterclockwise behavioral errors, we flipped both the sign of the behavioral error, as well as the corresponding channel response measured from the brain (Figure 3 (a), insert). This resulted in a metric of behavioral errors that were all $\geq 0°$. If neural biases track behavioral errors, then channel response biases (calculated as the channel response circular mean) should be $>0°$ also. Indeed, this analysis reveals that larger behavioral errors were accompanied by stronger biases in the memory representations in EVC ($t_{(5)} = 2.257$; $p = 0.023$), but not in IPS ($t_{(5)} = 0.116$; $p = 0.422$) (Figure 3b).

Complementary levels of analysis (across- and within-participants) of two independent datasets revealed a brain–behavior relationship during VWM maintenance in EVC, but not in IPS. To illustrate why
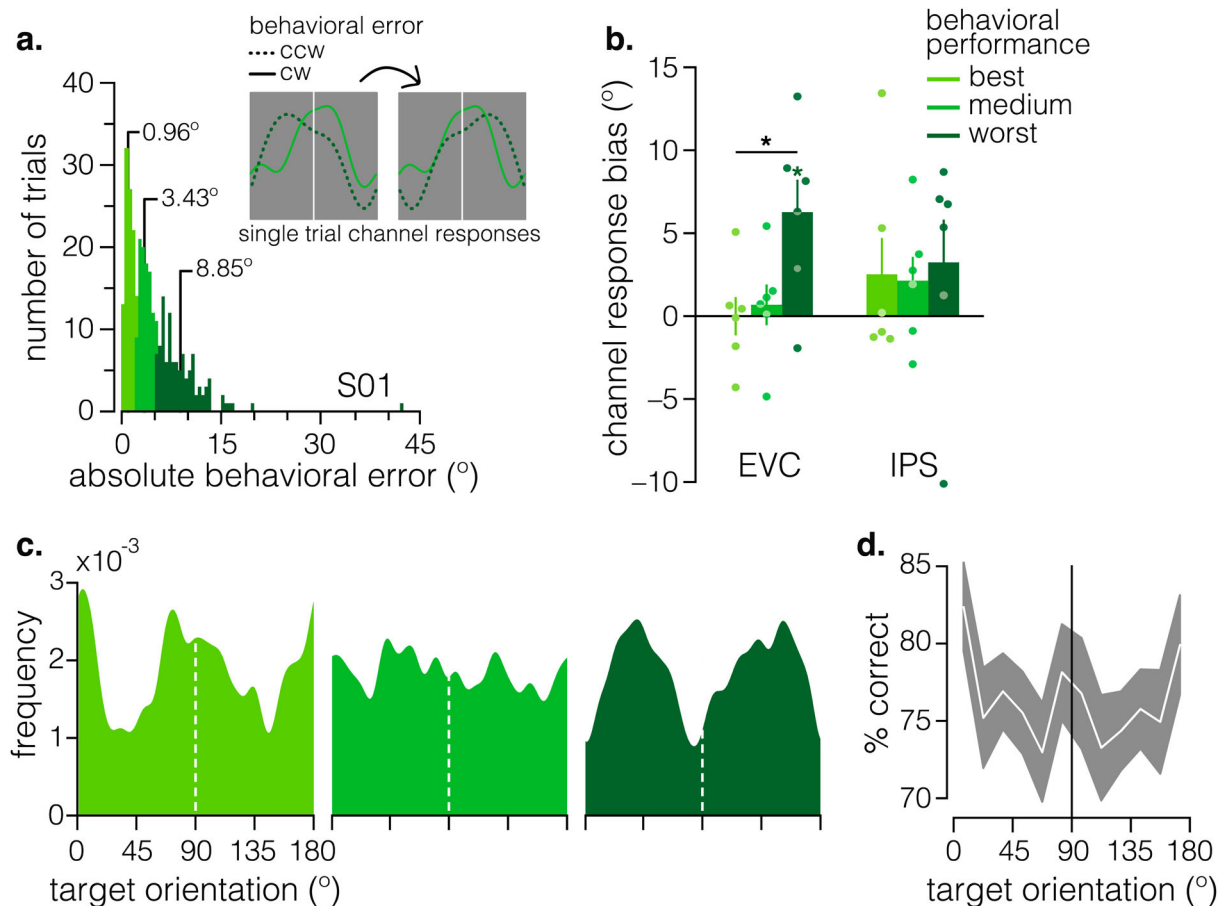
Figure 3. Behavioral responses are linked to neural biases. (a) For each participant in Experiment 1 (Rademaker et al., 2019) data were split into 3 equally sized bins of 108 trials each, based on the magnitude of behavioral errors. Thus, the 3 bins represent best, medium, and worst behavioral performance quantiles, shown here in three shades of green. The main panel shows the frequency of absolute behavioral errors for an example participant, split into 3 performance bins, with numerical values corresponding to the mean error in each bin. In addition to measuring a behavioral error, on each trial we estimate the brain's mnemonic representation by applying a multivariate model that yields a channel response function (for more details, see Rademaker et al., 2019). Channel response functions from 2 example trials are shown in the insert. If the mean (or "bias") of a channel response function changes in accordance with the behavioral error, then trials with larger behavioral errors are expected to have more strongly biased channel responses. To test this, for all trials with counterclockwise behavioral errors, the sign of both the behavioral error and the corresponding channel response function were flipped. In our example (insert), the dark-green channel response comes from a trial with a behavioral error of −16° and is therefore flipped (compare the left and right panels of the insert). The medium-green channel response comes from a trial with behavioral error of 3.5° and is not flipped. In this example, the channel response bias on the trial in dark green is indeed larger than that of the trial in medium green (where the behavioral error was also smaller). (b) There is a relationship between behavioral performance and mnemonic representations in EVC, but not in IPS. Specifically, in EVC the channel response bias (y-axis) was larger with poorer performance, as supported by a permutation test against a null model. In this test, bin assignment was shuffled on every permutation (of 1000), slopes were calculated (positive slopes indicate a larger channel response bias with larger behavioral error), and a t-statistic across participants was calculated, resulting in a null-distribution of t-statistics. A second permutation test was constructed to test for a brain-behavior relationship *within* each performance bin. Here, we broke the correspondence between the sign of the behavioral error and the flip of the channel response function, after which we calculated a t-statistic across subjects on every permutation. We only observed a significant channel response bias in EVC for the worst-performance bin. (c) The oblique effect in VWM for orientation. In the best performance bin (leftmost panel), trials with cardinal memory targets (around 0° and 90°) were more frequent. In other words, cardinal orientations tended to result in smaller behavioral errors, which in turn were associated with a smaller neural bias. Conversely, trials in the worst performance bin (rightmost panel) were frequently associated with oblique memory targets (around 45° and 135°). In other words, oblique orientations tended to result in larger behavioral errors, which were associated with a larger neural bias. All three plots use kernel density estimation (normal distribution with bandwidth of 5°) (d) To verify the generality of the stimulus-based effects shown in (c), we split all behavioral responses from the Christophel et al. (2018) task by target orientation. For each of the 12 possible target orientations (x-axis) the percentage correct was calculated (y-axis). A clear oblique effect is shown, with highest performance for targets around cardinal orientations, and lowest performance for oblique targets. Dashed lines indicate the 95% CI.

having converging lines of evidence is important, we will consider one factor that might facilitate analogous changes in brain and behavior: When looking at VWM for orientation within individual participants (Rademaker et al., 2019), trials that yielded the highest behavioral performance were trials with more cardinal orientations as targets, whereas trials that yielded the lowest behavioral performance were trials with more oblique target orientations (Figure 3(c)). This phenomenon, known as the oblique effect (Appelle, 1972), might contribute to the brain–behavior relationship shown in Figure 3 (b). Importantly, while an oblique effect was also present in the across-participant data from Christophel et al. (2018) (Figure 3(d)), it cannot contribute to the brain–behavior correlation at this level of analysis because trial-by-trial differences in orientation are averaged out. Vice versa, factors that might contribute to brain–behavior correlations across participants may not play a role when analyzing data within individual participants.

This third reanalysis demonstrates how a thorough understanding of analysis choices, and their implications at different levels, can further our understanding of the VWM system via converging lines of evidence. At the same time, it highlights how results across different studies cannot always be directly compared when the same question is approached at different levels of analysis.

Discussion

From the results presented here it becomes clear that analysis choices have a non-trivial impact on the sensitivity to retrieve mnemonic content from delay activity patterns. Different analysis approaches tap into different aspects of VWM-related activity, and the advantage of one analysis choice over the other often comes down to the specific research question being tested.

In our first reanalysis, we explored VWM representations in EVC and IPS during visual distraction, using two widely employed model training approaches. First, training the model on the responses from the memory delay revealed mnemonic information in EVC and IPS, with IPS demonstrating the strongest resilience against distraction, and EVC best tracking changes in behavioral recall across two different experiments. Second, when training the model on the sensory-driven responses evoked by an independent "mapping" task, mnemonic representations were found in EVC but not in IPS. Moreover, representations in EVC were reduced with distraction, even in the absence of a behavioral deficit. If we define VWM as including *any* representation pertaining to delay period activity (i.e., we remain agnostic to the type of signals contributing to VWM storage), then training a model on data from the memory delay is the most appropriate approach. If we restrict our definition of VWM to representations that generalize from sensory-driven responses, then training a model on sensory-driven responses is the more appropriate choice. Neither analysis approach is inherently better than the other; and each approach can be purposefully applied to address different research questions (e.g., to test different representational formats).

In our second reanalysis, we investigated VWM representations during attentional disengagement, and demonstrated how results can depend on the robustness of the signals used for model training. In this case, training on stronger mnemonic representations of currently prioritized ("attended") memory items helped uncover weaker mnemonic representations of "unattended" memory items. Using this approach, unattended memory items were uncovered in both EVC and IPS. By contrast, when training the model on signals evoked by the unattended items themselves, such items could only be detected in IPS. This result showcases how a previous null finding in fact depended on analysis choice and cautions against prematurely interpreting null results as evidence of absence (note: This caution extends more generally beyond analysis choices, as there are many additional reasons why existing neural representations can remain undetected from VWM delay activity (Chota & Van der Stigchel, 2021)).

In our third reanalysis, we showed that mnemonic representations in EVC (but not IPS) covaried with behavioral performance measured in two different experiments and at two different levels of analysis: across participants and within participants (i.e., across trials). Many factors might contribute to brain–behavior relationships. Here we used the oblique effect to demonstrate how one such factor can have different contributions at different levels of analysis. Evaluating results across multiple complementary levels of analysis can strengthen our understanding of the brain–behavior relationship.

The results of the reanalyses presented here provide important additional detail to the already existing studies in which the data were originally published (Christophel et al., 2018; Rademaker et al., 2019). They demonstrate that both EVC and IPS are involved in VWM storage but in different ways. One possible interpretation is that these brain regions might store mnemonic contents in different formats, as a result of which particular analysis choices might differentially affect the results from these areas. Our findings are in line with the view that the same mnemonic contents (e.g., memory for orientation) can undergo multiple neural implementations distributed across multiple locations (Lee, Kravitz, & Baker, 2013). Thereby, these reanalyses support the idea of the VWM system as a cortical network where no single brain area plays the role of monolithic store under every circumstance (Lorenc & Sreenivasan, 2021; Teng & Postle, 2021). However, involvement of some areas could be critical for specific aspects of different tasks. For instance, it's been suggested that visual cortex might serve as a local comparison circuit, ideally suited for matching mnemonic contents to the sensory environment (Rademaker et al., 2019; Xu, 2020). At the same time, it is unlikely that memory for visual stimuli solely relies on parts of cortex specialized in visual processing. Some visual stimuli are readily verbalized, such that little visual trace may persist (e.g., Yan et al., 2021), and representations of visual stimuli outside of visual cortex are abundant (Brissenden et al., 2021; Christophel et al., 2017; Ester et al., 2015; Lee & Baker, 2016). VWM is likely mediated through a larger number of regions, where each region is functionally distinct and essential in its own right. Such an architecture allows VWM to best serve behavior, by ensuring robust maintenance of visual information, and allowing for the flexible application of mnemonic contents under varying task requirements.

The above shows how our results can be interpreted in the context of a distributed network of cortical regions. Instead, if we assume a singular essential memory store, and interpret our findings according to the criteria put forth in Xu (2020), interpretation becomes fraught. To illustrate we will first look at brain–behavior relationships – deemed a necessary, but not sufficient criterion for an area to be considered "essential" (Xu, 2020). Our first reanalysis showed EVC, but not IPS, tracking behavior across

two experiments when training a model on memory delay period data. Based on this, one could conclude that EVC is essential for VWM storage and, by extension, that IPS is not. But when using sensory-driven responses for model training instead, EVC no longer tracked behavior, and IPS no longer represented mnemonic information at all. Based on this, neither EVC nor IPS are essential for VWM storage. In our third reanalysis we again showed that EVC, and not IPS, tracked behavior (both across and within participants), re-implicating EVC as the essential store. Through this exercise, one inevitably notices a volatility in interpretation, depending on analysis choice. Conclusions based on the other criteria proposed by Xu (2020) suffer from a similar volatility when evaluated against our results: Essential areas must represent VWM contents, which IPS does in one analysis but not another. Essential areas must also not show diminished storage as long as behavioral performance is unaffected, which EVC does in one analysis but not another. Thus, a strictly modular interpretation of our results based on these criteria cannot account for the complexity of brain mechanisms involved in VWM.

Beyond being too narrow in scope, we see another major concern in the unrestrained application of these criteria (Xu, 2020) to the VWM literature. Namely, the criteria themselves seem to be irrefutable. To illustrate let's again look at the brain–behavior relationship. While our results would by and large favor EVC as the "essential" store, we must still address the "sufficiency" clause: According to Xu (2020), mnemonic contents in early sensory areas may covary with behavioral performance as a result of *feedback* from a higher-order "essential" area. Thus brain–behavior correlations in EVC are not *sufficient* to prove its "essentialness". Few would argue that mnemonic signals in EVC are anything other than feedback driven, which means we should entirely disregard brain–behavior relationships in EVC when searching for an essential store. By extension of the Xu (2020) argument, however we could also posit that mnemonic contents in IPS may covary with behavioral performance as a result of *feedforward* signals from a lower-order "essential" area. In sum, whether or not the brain–behavior criterion is fulfilled seems to matter little, as one could always disregard these findings using this argument. More importantly, using the potential of feedback or

feedforward inputs as a means to disregard findings in any one area distracts from the very real possibility that the interplay between areas might be an essential way by which memory is upheld in the human brain.

Taken together, considering VWM as supported by a distributed cortical network has the potential to account for both a flexible storage-reconfiguration based on task demands, and stability of mnemonic contents in the face of various potential sources of interference. A strictly modular viewpoint on VWM accounts for neither. Moreover, evaluating our results against criteria used to determine a single "essential" cortical storage site yielded ambiguous conclusions at best. Such a narrow definition of VWM may limit our comprehension of the VWM system as a whole. Instead, from a distributed VWM viewpoint, memoranda are stored within multiple interconnected cortical sites, in multiple formats. Thus, rather than asking which area is essential to VWM, the question now becomes under what circumstances an area is of particular importance to VWM storage. Here we demonstrated that through careful consideration of analysis choices and the principles they test, we can gain insight into potential representational formats, their generalizability, and relationship with behavior.

Diversity Statement

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minorities are under-cited relative to the number of such papers in the field (Caplar et al., 2017; Dion et al., 2018; Dworkin et al., 2020; Maliniak et al., 2013; Mitchell et al., 2013; Zhou et al., 2020). We seek to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, and other factors. We obtained predicted gender of the first and last author of each reference by using databases that store the probability of a name being carried by a woman (Dworkin et al., 2020; Zhou et al., 2020). By this measure (and excluding self-citations), our references contain the percentages of (first/last authors): 1. woman/woman 2. man/woman 3. woman/man 4. man/man 5. unknown categorization. This method is limited in that (a) names, pronouns, and social media profiles used to construct the databases

may not, in every case, be indicative of gender identity and (b) it cannot account for intersex, nonbinary, or transgender people. We look forward to future work that could help us to better understand how to support equitable practices in science. Excluding self-citations to the authors of this paper, our references contain (first/last authors): 10% woman/woman, 22% woman/man, 10% man/woman and 36% man/man. Single author papers accounted for 10% (man) and 12% (woman) of citations.

Target article

Y. Xu (2020). Revisit once more the sensory storage account of visual working memory, *Visual Cognition, 28*: 5-8, 433-446. DOI: 10.1080/13506285.2020.1818659

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

PI and RLR performed the data reanalysis, PI, RLR, TBC, and SG wrote the manuscript.

Data availability

Behavioral and fMRI data from Rademaker et al. (2019) are available at https://osf.io/dkx6y. Behavioral and fMRI data from Christophel et al. (2018) are available from TBC upon request.

Code availability

All reanalysis code from this paper is available at https://osf.io/dcrf6/.

References

Allefeld, C., & Haynes, J. D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage*, *89*, 345–357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The" oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278. https://doi.org/10.1037/h0033117

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157. https://doi.org/10.1038/nn.4174

Boettcher, S. E., Gresch, D., Nobre, A. C., & van Ede, F. (2021). Output planning at the input stage in visual working memory. *Science Advances*, *7*(13), eabe8212. https://doi.org/10.1126/sciadv.abe8212

Brissenden, J. A., Tobyne, S. M., Halko, M. A., & Somers, D. C. (2021). Stimulus-specific visual working memory representations in human cerebellar lobule VIIb/VIIIa. *The Journal of Neuroscience*, *41*(5), 1033–1045. https://doi.org/10.1523/JNEUROSCI.1253-20.2020

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*(6), 1–5. https://doi.org/10.1038/s41550-017-0141

Chota, S., & Van der Stigchel, S. (2021). Dynamic and flexible transformation and reallocation of visual working memory representations. *Visual Cognition*, 1–7. https://doi.org/10.1080/13506285.2021.1891168

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

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, *26*(3), 312–327. https://doi.org/10.1017/pan.2018.12

Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. arXiv preprint arXiv:2001.01002.

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761.

Ester, E. F., Rademaker, R. L., & Sprague, T. C. (2016). How do visual and parietal cortex contribute to visual short-term memory? *eNeuro*, *3*(2), 1–3. https://doi.org/10.1523/ENEURO.0041-16.2016

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, *18*(1-3), 245–272. https://doi.org/10.1016/0010-0277(84)90026-X

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*(3), 307–317.

Fuster, J. M. (1995). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. The MIT Press.

Fuster, J. M. (1997). Network memory. *Trends in Neurosciences*, *20*(10), 451–459. https://doi.org/10.1016/S0166-2236(97)01128-4

Fuster, J. M., & Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*, *212*(4497), 952–955. https://doi.org/10.1126/science.7233192

Gayet, S., Paffen, C. L., & Van der Stigchel, S. (2018). Visual working memory storage recruits sensory processing areas. *Trends in Cognitive Sciences*, *22*(3), 189–190. https://doi.org/10.1016/j.tics.2017.09.011

Ghent, L., Mishkin, M., & Teuber, H. L. (1962). Short-term memory after frontal-lobe injury in man. *Journal of Comparative and Physiological Psychology*, *55*(5), 705–709. https://doi.org/10.1037/h0047520

Gross, C. G., & Weiskrantz, L. (1964). In J. M. Warren, & K. Akert (Eds.), *Some changes in behavior produced by lateral frontal lesions in the macaque, in The frontal granular cortex and behavior*. McGraw-Hill.

Jacobsen, C. F., Elder, J. H., & Haslerud, G. M. (1936). Studies of cerebral function in primates.

Kiyonaga, A., Dowd, E. W., & Egner, T. (2017a). Neural representation of working memory content is modulated by visual attentional demand. *Journal of Cognitive Neuroscience*, *29*(12), 2011–2024. https://doi.org/10.1162/jocn_a_01174

Larocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-category decoding of information in different attentional states in short-term memory. *Cerebral Cortex*, *27*(10), 4881–4890.

Lee, S. H., & Baker, C. I. (2016). Multi-voxel decoding and the topography of maintained information during visual working memory. *Frontiers in Systems Neuroscience*, *10*, 2. https://doi.org/10.3389/fnsys.2016.00002

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*(8), 997–999.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & 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.

Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in visual working memory: Resistance is Not futile. *Trends in Cognitive Sciences*, *25*(3), 228-239. https://doi.org/10.1016/j.tics.2020.12.004

Lorenc, E. S., & Sreenivasan, K. K. (2021). Reframing the debate: The distributed systems view on working memory. *Visual Cognition*, 1–9. https://doi.org/10.1080/13506285.2021.1899091

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R., & D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. *The Journal of Neuroscience*, *38*(23), 5267–5276. https://doi.org/10.1523/JNEUROSCI.3061-17.2018

Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, *67*(4), 889–922. https://doi.org/10.1017/S0020818313000209

Malmo, R. B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. *Journal of Neurophysiology*, *5*(4), 295–308. https://doi.org/10.1152/jn.1942.5.4.295

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience*, *16*(16), 5154–5167. https://doi.org/10.1523/JNEUROSCI.16-16-05154.1996

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience*, *13*(4), 1460–1478. https://doi.org/10.1523/JNEUROSCI.13-04-01460.1993

Mitchell, S. M., Lange, S., & Brus, H. (2013). Gendered citation patterns in international relations journals. *International Studies Perspectives*, *14*(4), 485–492. https://doi.org/10.1111/insp.12026

Petrides, M. (1995). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *The Journal of Neuroscience*, *15*(1), 359–375. https://doi.org/10.1523/JNEUROSCI.15-01-00359.1995

Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1650–1665. https://doi.org/10.1037/xhp0000110

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336–1344. https://doi.org/10.1038/s41593-019-0428-x

Rademaker, R. L., Park, Y., Sack, A. T., & 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. https://doi.org/10.1037/xhp0000491

Scimeca, J. M., Kiyonaga, A., & D'Esposito, M. (2018). Reaffirming the sensory recruitment account of working memory. *Trends in Cognitive Sciences*, *22*(3), 190–192. https://doi.org/10.1016/j.tics.2017.12.007

Shin, H., Zou, Q., & Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, *17*(14), 10–10. https://doi.org/10.1167/17.14.10

Teng, C., & Postle, B. R. (2021). Understanding occipital and parietal contributions to visual working memory: Commentary on Xu (2020). *Visual Cognition*, 1–8. https://doi.org/10.1080/13506285.2021.1883171

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain*, *92*(4), 885–896. https://doi.org/10.1093/brain/92.4.885

Wildegger, T., Myers, N. E., Humphreys, G., & Nobre, A. C. (2015). Supraliminal but not subliminal distracters bias working memory recall. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 826–839. https://doi.org/10.1037/xhp0000052

Woloszyn, L., & Sheinberg, D. L. (2009). Neural dynamics in inferior temporal cortex during a visual working memory task. *Journal of Neuroscience*, *29*(17), 5494–5507. https://doi.org/10.1523/JNEUROSCI.5785-08.2009

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences*, *21*(10), 794–815. https://doi.org/10.1016/j.tics.2017.06.013

Xu, Y. (2018). Sensory cortex is nonessential in working memory storage. *Trends in Cognitive Sciences*, *22*(3), 192–193. https://doi.org/10.1016/j.tics.2017.12.008

Xu, Y. (2020). Revisit once more the sensory storage account of visual working memory. *Visual Cognition*, *28*(5-8), 433–446. https://doi.org/10.1080/13506285.2020.1818659

Yan, C., Christophel, T. B., Allefeld, C., & Haynes, J. D. (2021). Decoding verbal working memory representations of Chinese characters from broca's area. *NeuroImage*, *226*, 117595. https://doi.org/10.1016/j.neuroimage.2020.117595

Zhou, D., Cornblath, E. J., Stiso, J., Teich, E. G., Dworkin, J. D., Blevins, A. S., & Bassett, D. S. (2020). Gender Diversity Statement and Code Note-Book v1. 0.