Title: Contralateral delay activity during dynamic spatial updates in working memory

Authors: Charles Chernik[1*], Ronald van den Berg[2], Mikael Lundqvist[1, 3*]

1) Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.

2) Department of Psychology, Stockholm University, Stockholm, Sweden.

3) The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, MA, USA

*corresponding authors

Abstract: Working Memory (WM) enables us to maintain and directly manipulate mental representations, yet we know little about the neural implementation of this privileged online format. We recorded electroencephalography data as human subjects engaged in a task requiring continuous updates to the locations of objects retained in WM as well as in a visually identical task with WM demands removed. Analysis of neural data suggested WM-related contralateral delay activity reversed polarity as objects held in WM moved to the opposite hemifield. This was partially but not fully explained by visual attention demands. Thus, the cortical location of activity related to both attention and WM was updated to meet behavioral demands as the spatial location of remembered objects changed.

Introduction

A fundamental aspect of cognition is our ability to hold and mentally manipulate information in an online state in working memory (WM)[1,2]. Both the format of retention[3–5] and the actual manipulation of WM information[6–10] are highly studied topics. Despite this, we know relatively little about the transformations of memory representations that underlie such manipulations[11]. This knowledge gap may in part be due to WM representations typically being studied under relatively static settings, while in real life WM representations are continuously updated to support behavior[12,13].

Spiking and local-field potential data from monkeys suggest that the representation of visual objects transfers from one prefrontal hemisphere to the other following

saccades that move the location of an earlier presented stimulus to the opposite hemifield[12]. Likewise, gaze bias in humans, a behavioral correlate of the location of retained WM objects, shifts as subjects turn their heads[13]. These findings suggest that WM object representations transfer across cortex when humans mentally update their remembered locations, but electrophysiological evidence for this is lacking. Here, we aim to bridge that gap by using a novel experimental paradigm combined with electroencephalogram (EEG) measures.

We tracked the cortical locations of WM representations using Contralateral Delay Activity (CDA), an electrophysiological measure of WM representations measured via EEG[14,15]. This measure reflects the difference in brain activity between the electrodes contralateral and ipsilateral to the side where the remembered items were presented. It correlates with the number of relevant but not irrelevant visual WM objects, and is often used as a correlate of lateral WM load[14]. Prior work has shown that CDA is sensitive to the selection of a subset of particular items after initial encoding[16], as well as subsequent encoding of additional items[17]. Due to its lateralized nature, CDA offers a way to test whether memory representations are spatially transferred within the cortex when remembered stimuli are mentally relocated from one visual hemifield to the other. Here, we have tested this by recording EEG while participants performed a WM task that required them to mentally update the locations of two remembered items across visual hemifields. Our main hypothesis was that when the remembered locations of stimuli move across visual hemifields, the underlying neural representations likewise relocate across hemispheres, reversing the sign of the CDA signal.

Results

Experiment 1: Working memory

We recorded EEG from healthy humans during a delayed-estimation task[18] that involved mental updating of locations of memorized stimuli after they disappeared from the screen (Figure 1a). Each trial began with a central fixation, followed by a cue indicating on which side (left or right of the fixation point) the to-be-remembered objects would appear. Four differently colored discs (two on each side) were then

shown, with the two non-cued items serving to balance visual inputs in the two hemispheres and ensuring activity differences were due to memory-related processes. After 500 ms, the four discs were replaced by colorless outlines that started moving 1 s later along an (invisible) circle (180°/s, clockwise or counterclockwise, 0-360 degrees, with 13 possible distances; Figure 1b). Finally, 3 s after movement onset, one of the outlines was highlighted and the subject was prompted to provide their estimate of its remembered color through a mouse click on a color wheel. Importantly, since the direction and degree of movement were unpredictable, accurate performance was possible only if the subjects mentally updated item positions in WM in real time.

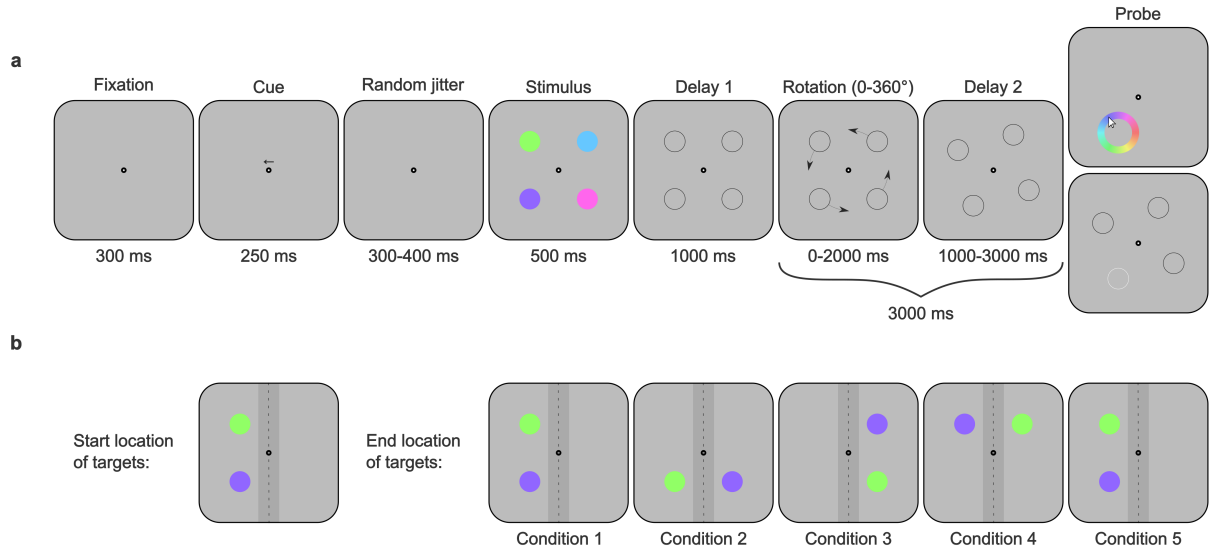

Figure 1. Task design of the experiments. a. Trial structure. Each trial started with a fixation period of 300 ms, followed by a cue indicating the to-be-attended side (left or right). After a random jitter of 300-400 ms, a memory array of four colored discs appeared, with two on each side of the fixation point. After 500 ms, the colored discs were replaced with black outlines to indicate their locations. After 1000 ms the outlines simultaneously started moving along an invisible circle, changing their positions relative to the midline by 0, 30, …, or 360 degrees (13 possible alternatives, with randomized sign, 180 deg/s). Once the movement was complete, the outlines remained in their new positions for at least 1 s. Then, in Experiment 1 (top), all the outlines disappeared, and a color wheel was presented at the post-movement location of one of the memorized discs, prompting recall. The response was submitted via a mouse click anywhere on the color wheel. In Experiment 2 (bottom), the outlines stayed in place, with one of them changing its color to white. The subject indicated through a keyboard button press whether the white outline corresponded to one of the tracked discs. **b.** Schematic illustration

of the five experimental conditions, defined by the laterality of the post-movement stimulus locations: Condition 1 – both discs remained in the original hemifield; Condition 2 – one of the discs transferred to the other hemifield; Condition 3 – both discs transferred to the other hemifield; Condition 4 – both discs transferred to the other hemifield and then one of them returned to the original hemifield; and Condition 5 – both discs transferred to the other hemifield and then returned to the original one. While there were 13 possible rotation angles, each of them fell in one of these five conditions. The dark gray region represents the area along the midline (width = 1.6°) where the discs could not end up after the movement, ensuring lateralization.

The response error with respect to the target item gradually increased with larger movement angles (Figure 2a, green), while the error with respect to the non-target item (the relevant item not probed) decreased (Figure 2a, purple). This suggests that subjects tended to confuse the two retained colors (swapping errors[19]) more often with larger movements. This was corroborated by fitting the data with an equal-precision model with non-target responses[20], which explained the performance degradation through increasing swapping rates at larger movements, while memory precision of the two items remained stable (Figure 2b). Importantly, even at the largest movement angle subjects performed accurately: estimated swap rates were relatively low (around 10%) and memory precision was high (parameter log(κ) was on average around 2, which corresponds to an average memory error of 16.8° on non-swap trials). Hence, they must have been updating item position in WM in real time.

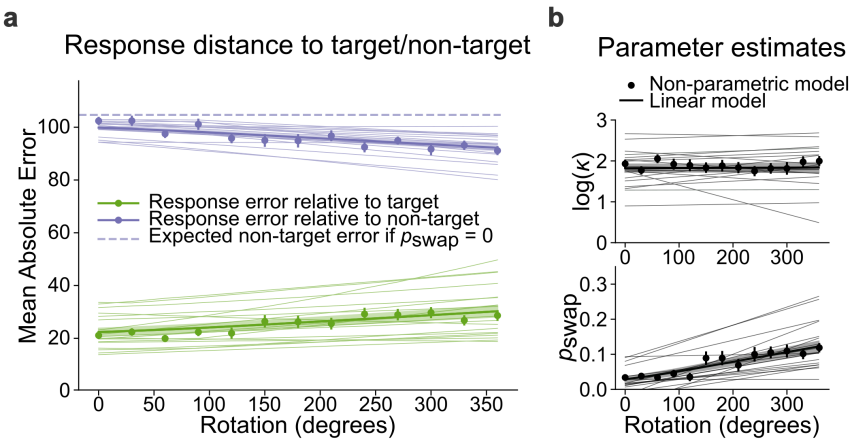

Figure 2. **Effect of stimulus movement on response accuracy. a.** Subject-averaged response errors as a function of movement angle. The error relative to the target (green markers) gradually increased with movement angle (Pearson's r=.88, p<.001), while the error relative to the non-target item (purple markers) decreased (r=−.89, p<.001). These trends were well accounted for by a model in which memory precision and swap rate are linearly related to motion angle (thin lines: estimates for individual subjects; thick lines: averages). Shaded areas and error bars represent one SEM. **b.** Estimated effect of movement angle on memory precision (top) and swap rates (bottom) in the linear model (lines) and in a non-parametric model (markers). While swap rate increased (r=.92, p<.001), there was no evidence that memory precision was affected by motion angle (r=−.02, p=.93).

Next, we computed the CDA from the EEG data to investigate the neural underpinning of this mental updating. Under the hypothesis that mental updating of item locations caused WM representations to physically transfer across the cortex (Model 1) both the magnitude and sign of the CDA were predicted to change in specific ways as the empty discs moved across visual hemifields (Figure 3a): right before motion onset, the CDA should be negative, as in previous studies using stationary items; in trials where both discs stayed in the original hemifield (Condition 1), the CDA should remain negative; if one disc moved over to the other hemifield (Condition 2), the CDA should disappear, because the two WM objects would be in opposite hemispheres, eliminating the lateralized imbalance that drives the CDA; if both discs crossed the midline (Condition 3), the CDA should change sign to positive; if one disc then moved back into the original hemifield, the CDA should again disappear (Condition 4); finally, in the trials where both discs moved full circle (Condition 5), the post-motion CDA should be negative again, as if no rotation had taken place. An alternative hypothesis is that the

neural representations of WM objects remain anchored to their original encoding locations, regardless of subsequent mental updates of spatial positions (Model 2). This model predicts that the sign of the CDA remains negative throughout any post-encoding updates of locations, with a magnitude that is either stable or may gradually decrease to reflect decaying memory quality. A final hypothesis is that mental movement turns the neural representations into a new format that is not tracked by CDA (Model 3). This model predicts that cross-hemispheric movements cause CDA (possibly apart from Condition 1) to disappear and not come back.

Qualitatively, our data closely aligned with the predictions of Model 1, with the CDA starting off negative, reversing sign as WM objects transitioned into the opposite hemifield and then again when they moved back into the original hemifield (Figure 3b). To test the predictions of Model 1 quantitatively, we performed t-tests on the time-averaged CDA in each of the five conditions, in a 700 ms window shortly after movement ended in each condition (Figure 3b, horizontal bars). For all 5 conditions, the Bayes Factors (BF) supported the prediction of Model 1 (Figure 3c, n=19 in all tests), though for Condition 5 the evidence was anecdotal (BF = 2.55). Similarly, the p-values were in line with the model's predictions for all conditions except Condition 5. These results were robust against changes in how the CDA time windows were defined (Supplemental Figure 1 and text). In the three conditions where Models 1 and 2 differed in their predictions, the data favored Model 1, with BFs of 3.57 (Condition 2), 471.7 (Condition 3), and 3.84 (Condition 4). In the comparison between Models 1 and 3, Model 1 was clearly favored in Condition 3 (BF = 32.1), but the evidence was weakly in favor of Model 3 in Condition 5 (BF = 0.60). Finally, we fitted Bayesian mixed-effects models to compare the three Models using all five conditions at once. This analysis again favored Model 1, capturing the predicted CDA across conditions with an estimated model weight of 59%, versus 25% for Model 3 and 16% for Model 2.

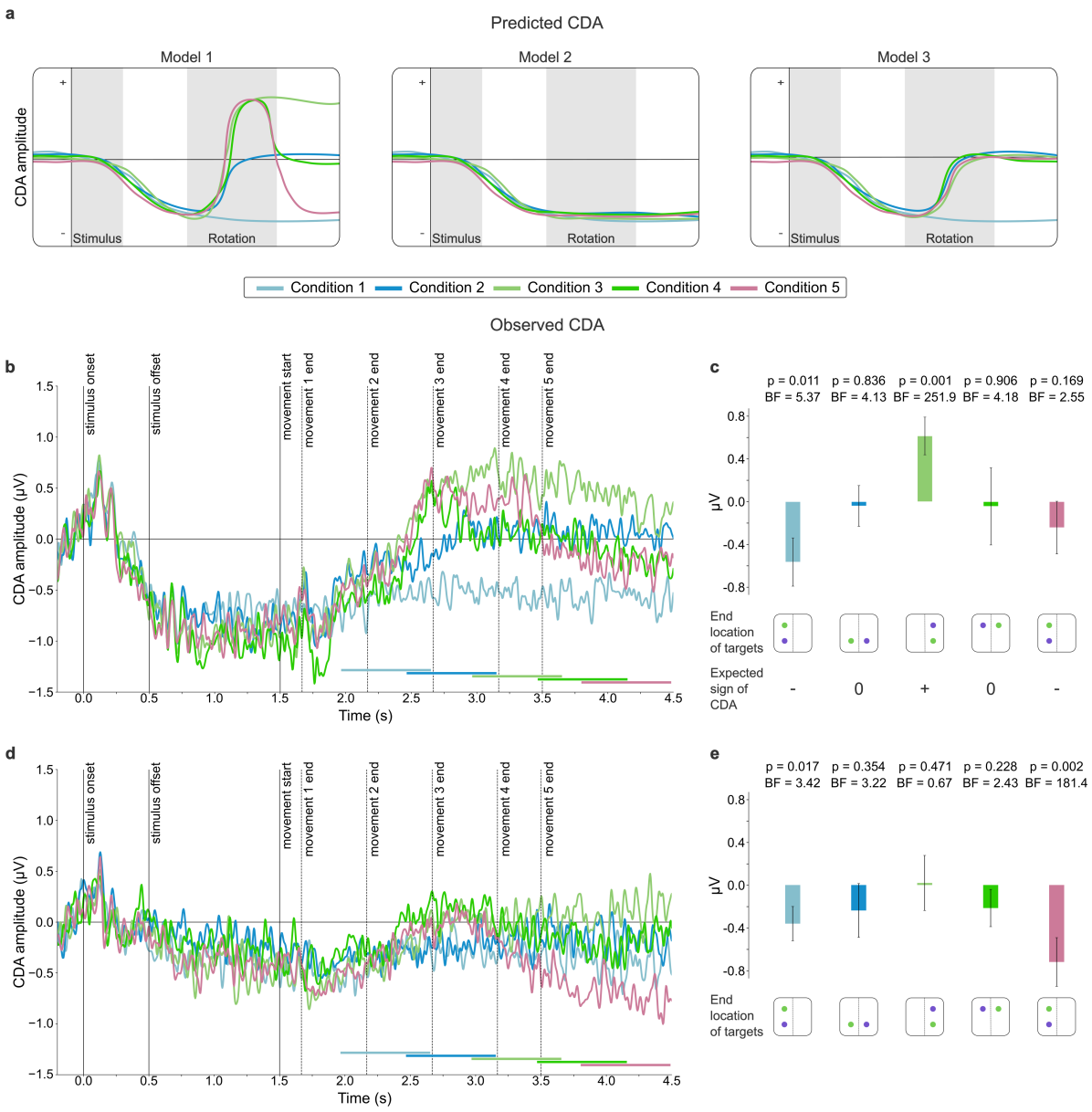

Figure 3. **CDA following movement of WM objects/attentional tracking. a.** Cartoon of model predictions of how the CDA changes following the mental relocation of the discs. In the main model, Model 1, the CDA changes depending on the hemifield the discs end up in post rotation (condition 1-5). For condition 1 and 5 the discs end up in the initial hemifield and the CDA should be negative. For condition 2 and 4, there is one disc in each hemifield and there should be no CDA. For condition 3, both discs end up in the opposite hemifield and the CDA should be positive. As alternatives, Model 2 predicts that the neural representations of the disks remain in their original locations (meaning CDA remains unchanged) and Model 3 predicts the CDA disappears completely after the update of WM contents occurs (meaning the CDA only remains when both discs stay in the initial hemifield). **b.** Measured CDA per

condition in Experiment 1. Dotted vertical lines denote the latest possible time when movement ended per condition. Colored horizontal bars denote 300-1000 ms post movement per condition, used to measure the CDA statistically. **c.** *Top:* average CDA post movement per condition in Experiment 1, taken from the time periods indicated in **b**. Error bars denote SEM. The p-values were obtained from t-tests on the null hypothesis of CDA being 0. In the conditions where Model 1 predicted a specific effect direction, one-sided tests were used. The BFs were obtained from Bayesian t-tests and quantify the evidence in favor of Model 1 (see Methods). *Middle:* Cartoon demonstrating the end location of the targets per condition (assuming both discs started on the left). *Bottom:* sign of the CDA predicted by Model 1. **d.** Measured CDA per condition in Experiment 2. **e.** Average CDA post movement per condition in Experiment 2.

Experiment 2: Attentional tracking

While the CDA is widely regarded as a neural marker of working memory (WM) storage [15,21,22], some studies suggest that it is also sensitive to attentional processes [23–25]. To test to what extent the patterns observed in Experiment 1 reflect attentional processes rather than WM retention, we conducted another experiment with the same sequence of stimuli as in Experiment 1, but with a task that did not require storing them in WM. In Experiment 2, participants were instructed to ignore the colors and simply keep track of the positions of the two cued discs. At probe, instead of recalling the color of the highlighted outline, they indicated whether it was one of the two tracked discs. If the cortical transfer suggested by the CDA patterns in Experiment 1 was primarily a marker of attentional tracking, we expected the pattern in Experiment 2 to be similar, both qualitatively and quantitatively. However, if cortical transfer of WM contents was an important driver of the patterns observed in Experiment 1, then the CDA and sign reversals should be less pronounced in Experiment 2.

As this was a simple task, the behavioral performance was unsurprisingly good (accuracy = 94.2±5.1%, M ± SD). Since the task in Experiment 2 did not involve WM encoding, we first checked if the stimuli evoked a CDA signal at all, by performing a Bayesian t-test on the CDA in the 700 ms window before the movement (this part of the trial was identical across the five conditions). Evidence for a negative CDA was observed (BF = 22.7), albeit of much weaker strength than in Experiment 1 (BF = 141.03). Hence, the CDA in our paradigm appears to reflect a mixture of attentional

and WM processes, with attention alone producing a weaker and less consistent signal.

Next, we computed and analyzed the post-movement CDA the same way as before. Qualitatively, the observed pattern was comparable to the one in Experiment 1, yet there were marked quantitative differences (Figure 3d). Notably, unlike in Experiment 1, no polarity reversal from negative to positive was observed in Condition 3 (Figure 3e). Bayesian mixed-effect modelling gave 47%, 45%, and 8% weight to Models 1, 2, and 3, respectively.

In summary, Experiment 2 showed that although attentional tracking alone can give rise to lateralized activity resembling a CDA, it does not reproduce the full pattern of sign reversals observed in Experiment 1. The absence of a polarity flip in Condition 3, together with the weaker overall evidence for Model 1, suggests that the robust CDA reversals in Experiment 1 cannot be explained solely by attentional selection.

Discussion

To gain insight into the neurophysiological processes underlying real-time manipulations of working memory representations in a dynamic world, the current study examined the CDA – an electrophysiological marker of WM load – in a novel experimental paradigm where subjects were required to encode objects (colored discs) into WM and then update the spatial location associated with them. In a separate experiment, subjects completed a visually identical task without having to remember the colors. Together, the results suggested that dynamic updates to the relative location of objects held in mind are reflected in qualitative changes in the CDA. These changes were related to both the updated WM information and attentional processes.

In Experiment 1, we observed that the CDA followed the mental movement in real time, reversing its sign when the objects moved into the opposite hemifield, and then again when they returned to the original hemifield. These results were consistent with our main hypothesis that mentally updating the location of a memorized item causes

230 a cortical transfer of the neural representation of that item, although it should be noted
231 that the evidence for Condition 5 was relatively weak. Interestingly, our findings stand
232 in contrast to those of an earlier study that found no CDA sign reversal when objects
233 were encoded in one hemifield and later tested in the opposite one[26]. We suspect the
234 effect observed here requires a task with strong emphasis on updating contents in real
235 time to prevent the updating during the test period itself. Consistently, including
236 subjects with poor accuracy on the behavioral task reduced the effect (Supplemental
237 Figure 2). However, the source of the effect could also be the continuous visuospatial
238 attention required in our task, if CDA is driven by attention rather than WM. The aim
239 of Experiment 2 was to examine to what extent attentional tracking can explain the
240 effect.

241 The results from Experiment 2 suggested that neither interpretation alone could fully
242 account for the dynamically updated CDA. If the CDA effects observed in Experiment
243 1 were solely due to attentional tracking of moving placeholders, then similar CDA
244 patterns should have emerged even in the absence of a WM requirement. Conversely,
245 if the CDA strictly reflected maintained and manipulated WM content, no sustained
246 lateralized activity would be expected when subjects attended to the placeholders
247 without encoding or storing color information[17,27]. Instead, the observed CDA response
248 in Experiment 2 showed partial activation, suggesting that attentional tracking alone
249 may elicit some CDA-like activity, but the full effect seen in Experiment 1 likely
250 depends on the presence of active WM representations. Crucially, unlike in
251 Experiment 1, the observed CDA did not reverse polarity in any condition at any time
252 point, indicating that polarity reversals could be a distinctive marker of active updating
253 of spatial WM information.

254 Another interesting observation was the strong negative post-movement CDA in
255 Condition 5 of the attention task. This strong negative CDA, compared with the
256 relatively weak and inconclusive effect in the same condition in the WM task,
257 represents a notable divergence between the two experiments. One tentative
258 explanation is that this difference relates to the absence of a polarity reversal in
259 Condition 3 of the attention task. If attentional tracking remained anchored to the
260 original hemifield even when both discs crossed the midline, then when they returned
261 in Condition 5 the CDA would re-emerge strongly negative, unlike in the WM task

where intermediate remapping attenuated the effect. Further work will be needed to clarify the mechanisms behind the divergence in Condition 5.

Previous research has shown CDA to reflect filtering (tracking the number of relevant items, and not distractors[14,21]) and to dynamically track various transformations of the WM load – such as grouping[28,29], splitting[30], and sequential encoding[17], – but its response to spatial manipulations, up until very recently[32], has been explored only in the context of multiple-object tracking (MOT)[24]. Notably, Drew et al.[24] observed CDA reversing polarity in their MOT task, which they refer to as 'handoff of attention'. However, the lack of such a reversal in our second experiment contradicts this purely 'attentional' interpretation. Additionally, previous findings suggest that object-tracking related CDA fades approximately 350 milliseconds after the objects stop moving[23], whereas we observed sustained CDA throughout the post-movement delay periods with objects held fixed, suggesting WM- rather than attention-related activity. What is more, a recent study by Wang et al.[31] suggests that the classic MOT tasks may include an implicit WM demand by requiring participants to maintain the spatial representations of the tracked stimuli in WM, meaning that past observations from MOT might have reflected more than just attentional processes too.

In both WM and MOT literature, CDA has widely been considered to reflect item-based processes, be it WM maintenance or attentional tracking. Recently, a refined view has emerged, suggesting it to reflect a common process of assigning spatiotemporal 'pointers'[33]. In a combined WM and MOT task, Styrkowiec et al.[32] observed CDA reflecting tracking load rather than WM load when both were simultaneously tested. The authors interpreted this as evidence that CDA reflects an indexing mechanism rather than content-related activity. However, this view cannot account for the observed differences between our two experiments. Even if the two disks in Experiment 2 were assigned a single pointer (which could explain a weaker pre-movement CDA compared to Experiment 1), it would not explain the lack of polarity reversal during or after the movement. Moreover, the pointer-based accounts fail to explain the growing body of evidence showing CDA to reflect more than discrete item-based processing[34], as it has been observed to be sensitive to perceptual complexity[35,36], semantics[37,38], and familiarity[39]. Our results seem in line with these less discrete accounts of the nature of CDA.

In conclusion, our results suggest that the cortical location of activity related to both attention and WM was updated to meet behavioral demands as the spatial location of remembered objects changed. This would put them in line with single neuron activity recorded from non-human primates suggesting that neural information retaining WM object identity is transferred across hemispheres as item locations are updated[12]. However, since the nature of the CDA is still debated, whether the observed activity reflects pointers (or a similar indexing mechanism) or the WM content itself remains an open question. This could be further tested by using methods directly decoding object identity from EEG[40].

Methods

Subjects

A total of 26 (Experiment 1) and 31 (Experiment 2) subjects were recruited via Karolinska Institutet's online recruitment system. All subjects provided informed written consent before the start of the experiment and were reimbursed 100 SEK per hour. The study was approved by the Swedish Ethical Review Authority and was performed according to the Declaration of Helsinki. All subjects reported normal or corrected-to-normal visual acuity and no neurological disorders. Prior to participation in Experiment 1, color vision was tested using the short (6-plate) version of the Ishihara's Test for Color Deficiency[41]. The test was considered passed if all plates were read normally. All subjects passed the test.

Experiment 1: Data from one subject was excluded from analysis due to excessive noise in the EEG and data from an additional six subjects were excluded due to poor behavioral performance (see below for details). The analysis reported in the main text was performed on the data from the remaining 19 subjects (M_{age} = 28.2, SD_{age} = 5.2; 11 female, 8 male; 3 left-handed), who each performed 480 trials across 5 conditions (see below). According to a recent meta-analysis[42], this was sufficient to test our main hypothesis. We verified that our conclusions did not critically depend on the behavior-based participant exclusions (Supplemental Material).

Experiment 2: One subject was excluded from analysis due to not completing the full experimental session. Two subjects were excluded due to excessive noise in the EEG. Additionally, two more were excluded due to poor behavioral performance. The final sample size was 26 subjects (M_{age} = 28.2, SD_{age} = 5.1; 14 female, 12 male; 2 left-handed, 1 ambidextrous). As in Experiment 1, all subjects performed 480 trials across 5 conditions.

Stimuli and procedure

All tasks were completed on a Windows PC with a Dell E2422H monitor (24”, 1920 × 1080 pixels, 60 Hz refresh rate). The viewing distance was approximately 58 cm, maintained by positioning subjects’ chin on a chin rest. The tasks and stimuli were generated and displayed using MATLAB with the Psychophysics Toolbox Version 3 (PTB-3) extension[43–45]. During the experiment, the subjects’ pupil size and eye movements were recorded using a Gazepoint GP3 eye tracker. Trials in which participants broke fixation by moving their gaze were automatically aborted, but the experimental session continued until they had successfully completed a total of 480 trials. A break in fixation was defined as the average gaze position straying more than 5.4° from the center of the screen.

Experiments 1 and 2 were structurally and visually identical, except for the format of the probe. The experimental session consisted of two parts, with a 5-10-minute rest period in between. Each part consisted of 240 trials, divided into 12 blocks of 20 trials each. Blocks were separated by 10-second breaks. Before the start of the experimental session, subjects completed a training block of 20 trials. EEG was not recorded during the training and the behavioral data were not included in the analysis.

On each trial subjects were presented with an array of four colored discs on a gray (RGB = 128 128 128) background. The colors of the discs were chosen from an isoluminant color wheel (CIE L*a*b* color space, [L = 70, a = −6, b = 14, radius = 49]), discretized to 180 colors and calibrated to the testing monitor. The colors were drawn from a uniform distribution on the color wheel but with the constraint that no two colors were separated by less than 30 degrees. In Experiment 1, subjects were asked to memorize two of the four presented colors, located in a particular hemifield. The relevant hemifield was randomized across trials and indicated by a pre-cue. In Experiment 2, subjects were instructed to ignore the colors.

Trial structure: A trial started with a fixation period, during which a fixation point (two concentric circles: black, radius of 0.3°, and white, radius of 0.1°) was presented at the center of the screen. The subjects' focus on the fixation point was monitored by the eye tracker to ensure 300 ms of uninterrupted fixation, meaning that moving their gaze away from the screen center extended the fixation period rather than aborting the trial. Breaking fixation at any later point, from the appearance of the cue and until the appearance of the probe, however, aborted the trial. A cue in the form of an arrow, indicating the to-be-attended side, was presented for 250 ms following the fixation period. Then, after a variable delay of 300-400 ms with only the fixation point visible, four colored discs (radius=1.3°) were presented for 500 ms, located at the corners of an invisible square centered on the fixation point. The distance between the fixation point and each stimulus was 8.1°. This was followed by a delay period of 1000 ms, during which only four identical black outlines of the discs remained on the screen, with their colors having disappeared at the start of the delay. Then, these outlines started moving along an invisible orbit centered on the fixation point, in identical direction and with identical speed, so that both the distances between the discs and the distance between the discs and the fixation point remained the same. The direction and angle of the movement was randomly selected from the range of [0°, 360°] in steps of 30°, meaning there were 13 possible angles in total. The movement was performed at a constant speed of 180°/s, meaning that its duration depended on the angle, with the shortest movement lasting 0 s (angle of 0°) and the longest movement lasting 2 s (angle of 360°). Another delay followed the movement, during which the discs' outlines remained visible in their new locations. The duration of the delay was set to 3 s minus the duration of the movement, so that the total duration of movement and delay was always exactly 3 s, of which at least 1 s was delay. After the delay, in Experiment 1, the outlines of the discs disappeared and a memory probe – a color wheel – appeared at the new location of one of the previously cued discs. Subjects were asked to recall the color of that disc by picking it on the color wheel. Subjects had unlimited time to respond. After their response was submitted, they received feedback showing both the color they had chosen and the correct color for 500 ms, which concluded the trial. In Experiment 2, all outlines stayed on the screen, with one – the probe – changing its color to white. Subjects were asked to indicate whether the probe was one of the two discs they were tracking by pressing the "F" or "L" key. The keys were labeled with the words "YES" and "NO", with the label assignment

randomized between participants. The probe was a tracked disc in exactly half of the trials. After the response was submitted, the fixation point changed color to reflect whether it was correct (green) or incorrect (red). This feedback was shown for 500 ms. The intertrial interval was 1000 ms.

Conditions: We defined five experimental conditions, based on the final locations of the two cued discs: Condition 1 (0°, 30°), in which both discs remained in the original hemifield; Condition 2 (60°, 90°, 120°), in which one of the discs transferred to the other hemifield; Condition 3 (150°, 180°, 210°), in which both discs transferred to the other hemifield; Condition 4 (240°, 270°, 300°), in which both discs transferred to the other hemifield and then one of them returned to the original hemifield; and Condition 5 (330°, 360°), in which both discs transferred to the other hemifield and then returned to the original one.

Trial balancing: The attended hemifields were set to have equal numbers of trials (240) and within each attended hemifield, the numbers of trials per movement condition were likewise set to be equal (48). The numbers of trials per angle within each condition were balanced as well (24 or 16 for conditions with 2 or 3 angles, respectively). The sign of the angle, which determined whether the movement would be clockwise or counterclockwise, and the probed item were randomized across trials without enforcing equal numbers. Trial order was randomized, separately for each subject.

Participant exclusion based on poor behavioral accuracy

A pivotal aspect of our task design for Experiment 1 was that accurate performance was possible only if subjects mentally updated locations of stimuli held in WM as the disc outlines moved on the screen (see main text). Subjects with large response errors likely failed to perform this mental updating. Since our main interest was to study neural correlates of mental updating, data of such subjects were expected to be uninformative, essentially only adding noise. Therefore, we excluded subjects with unusually large errors from the main analyses. To this end, we imposed a performance criterion that excluded all subjects with an average response error of at least two standard deviations above the group mean, which was enforced iteratively until no subjects fulfilled this criterion. This resulted in a total of 6 exclusions (compare Figure 2a with Supplemental Figure 3 to see the performance of the excluded subjects relative to the rest). To make sure that our main findings do not critically depend on

this exclusion, we reran the analysis with these subjects included. As anticipated, the results were highly similar, albeit with reduced statistical strength due to a lowered signal-to-noise ratio in the EEG data (see Supplemental Figure 2), suggesting that accurate performance was indeed pivotal. For Experiment 2, the same criterion was applied but only once, rather than iteratively, leading to 2 exclusions.

Behavioral analysis

Experiment 1: For each participant we calculated the mean response error for each of the 13 motion angles. We calculated this error both with respect to the color of the target item and the color of the non-target (the cued but not probed) item. To assess how these two types of errors related to the motion angle, we used a Pearson correlation test, applied to participant-averaged errors (hence, 19 data points per correlation test). To estimate how the swap rate and memory precision related to the motion angle, we fitted an equal-precision model with non-target responses[20] to the raw response data, separately for each participant. In this model, the response error on each trial was drawn from a Von Mises distribution with a concentration parameter κ. Moreover, there was a certain probability, p_{swap}, that the response was based on the memory of the non-target item instead of the target. To investigate potential effects of motion angle α on memory precision and swap rate, we modelled both these variables as linear functions, through $\kappa(\alpha) = a + b\alpha$ and $p_{\text{swap}}(\alpha) = c + d\alpha$, where a, b, c and d were fitted as free parameters. The best-fitting parameters for each participant were found through maximum-likelihood estimation, that is, by finding the values of a, b, c, and d that maximized

$$\sum_{i=1}^{n_{\text{trials}}} \log\left[\left(1-p_{\text{swap}}\left(\alpha_i\right)\right)\text{VM}\left(\varepsilon_{\text{target},i},\kappa\left(\alpha_i\right)\right)+p_{\text{swap}}\left(\alpha_i\right)\text{VM}\left(\varepsilon_{\text{non-target},i},\kappa\left(\alpha_i\right)\right)\right],$$

where VM(ε, κ) is a zero-mean Von Mises distribution with concentration parameter κ, evaluated at ε; $\varepsilon_{\text{target},i}$ and $\varepsilon_{\text{non-target},i}$ denote the response errors relative to the target and non-target item, respectively, on trial i; α_i denotes the motion angle on trial i. The optimization was performed in Python, using the Bayesian Adaptive Direct Search (BADS) package[46].

Experiment 2: Since the purpose of the task in Experiment 2 was to ensure that participants were actively tracking rather than passively viewing the movement, we did

not perform any behavioral analyses apart from computing the performance accuracy defined as the ratio of correct responses to the total number of trials.

EEG recording

EEG was acquired at 2048 Hz from 64 active Ag/AgCl electrodes using BioSemi ActiveTwo system. The electrodes were placed on an elastic cap according to the International 10-20 system. Two additional electrodes were placed on the left and right mastoids. The vertical electrooculogram (VEOG), used to detect eye blinks, was recorded from electrodes located 2 cm above and below the right eye. The horizontal electrooculogram (HEOG), used to measure horizontal eye movements, was recorded from electrodes located 1 cm lateral to the external canthi.

EEG preprocessing

Preprocessing and analysis of the EEG data was carried out using MNE-Python[47]. For the initial visual inspection, all electrodes were re-referenced offline to the average of the mastoids. If faulty electrodes were (manually) detected during or after the recording, they were interpolated using the MNE interpolate_bads() function with default parameters. This was done for 2 out of 190 (10 electrodes x 19 participants) electrodes kept in the main CDA analysis for Experiment 1, and for 9 out of 260 electrodes for Experiment 2. As excessive noise was discovered in the mastoid electrodes of several participants, EEG data were then re-referenced to the average of all 64 electrodes. The continuous data were bandpass-filtered using a finite impulse response (FIR) filter to 0.1-30 Hz and down-sampled to 256 Hz. A Picard[48] Independent Component Analysis (ICA), as implemented in MNE, was then run on the filtered and down-sampled data and the obtained components were automatically tested (using MNE function find_bads_eog() that utilizes correlation with EOG channels) and manually inspected in order to detect and remove ocular artifacts (blinks and horizontal eye movements). The results of automatic detection aided the process, but the final decision was made based on manual inspection. EEG was epoched between -200 and 4500 ms time-locked to stimulus onset. Each epoch, therefore, included 200 ms of pre-stimulus fixation, 500 ms of stimulus presentation, 1000 ms of pre-movement delay, 0-2000 ms of stimuli movement, and 1000-3000 ms of post-movement delay.

Automatic trial rejection was used to remove trials contaminated with muscle artifacts, channels blocking, slow drifts, and step-like voltage changes in the electrodes used for the ERP analyses. The detection algorithms and rejection criteria were implemented as described in[49]. The remaining trials were manually inspected to make sure no obvious artifacts had been missed. To aid reproducibility, manual rejection was substituted with an additional criterion of peak-to-peak amplitudes exceeding 150 μV, as the result of applying it was virtually identical. Subjects with fewer than 360 (75%) clean trials were excluded from analyses. Thus, for the remaining 19 subjects in Experiment 1, we rejected on average 1.6% (SD = 1.4%) of trials, resulting in an average of 472 trials per participant. For the 26 subjects in Experiment 2, an average of 2.4% (SD = 3.1%) trials were rejected, leaving an average of 468 clean trials per participant.

EEG analysis

CDA extraction: Artifact-free epochs were baseline-corrected using the 200 ms of pre-stimulus fixation as baseline and trial-averaged separately for left and right attended sides. All ERP analysis was restricted to the 10 posterior channels reported to show maximal CDA in memory tasks[50]: P5/P6, P7/P8, PO3/PO4, PO7/PO8, and O1/O2. The difference (CDA) waves were computed per participant, by subtracting their average activity recorded at ipsilateral sites from the average activity recorded at contralateral sites, with laterality defined relative to the attended (cued) side.

For statistical analysis, we measured CDA as the mean amplitude between 300 and 1000 ms post stimulus offset for pre-movement delay (analyzed in the Supplemental Material) and the mean amplitude between 300 and 1000 ms post movement end per condition in post-movement delay (main analysis). The movement end per condition was defined as the time point at which the longest movement included in the condition ended, e.g. 30° movement for Condition 1 and 210° movement for Condition 3.

Frequentist t-tests: To test the predictions of our main model against the data, we performed Student's t-tests on the time-averaged CDA measures. The null hypothesis was specified as H_0: μ=0. The model predicted H_0 to be true in Conditions 2 and 4 and false in Conditions 1, 3, and 5. Since the model predicted a specific effect direction in each of the latter three conditions, we used one-tailed tests for those conditions. For

the two conditions where the null hypothesis was predicted to be true, we used a two-tailed test.

Bayesian t-tests: A limitation of frequentist tests is that they can only provide evidence *against* the null hypothesis. A high *p*-value means that the data do not provide evidence against the null hypothesis, but it would be a statistical fallacy to interpret such a finding as evidence *in favor* of the null[51]. Therefore, we also analyzed the data using Bayesian t-tests. This type of test quantifies evidence by means of a Bayes Factor (BF), expressing how strongly the data support one hypothesis relative to another hypothesis. BFs above 1 provide support for a given hypothesis, below 1 evidence against. We use the interpretation scale from [52] to map Bayes factors to verbal labels such as 'weak' and 'strong' evidence. An advantage of BFs is that they can quantify evidence both for and against any well-defined hypotheses, including the null[53]. In the Bayesian evaluation of the model predictions, we computed for each condition the BF for the hypothesis following from our main model, denoted H_M, relative to an alternative hypothesis H_A, as listed in Table 1. Note that in Condition 1 we compared the model prediction, H_M: $\mu<0$, only against "$\mu=0$", not against "$\mu>0$". Our motivation for this was that published literature on CDA consistently reports the amplitude to be negative or non-existent and rarely or never as positive. Therefore, "$\mu>0$" has near-zero prior plausibility and should, therefore, not be included when doing Bayesian testing (including it would have increased the evidence for our model, so the reported value can be considered as a conservative estimate). For the other conditions we had no strong prior expectations, because there is no prior literature on the effect of mental updating on CDA. Therefore, we specified the alternative hypothesis in those conditions as the complement of the model's hypothesis (e.g., when the model predicts ">0", the alternative would be defined as "≤0"). We computed BFs using the Bayesian t-test in JASP[54], which allows to test four different types of hypotheses ($\mu<0$, $\mu=0$, $\mu>0$, $\mu\neq0$) against each other. A limitation is that it does not allow to directly test the "compound" alternative hypotheses that we specified for Conditions 3 and 5. Instead, we computed these BFs as the average BF obtained from testing H_M separately against each component in the compound hypothesis. Hence, in Condition 3, we took the average of the BF obtained from testing H_M: $\mu>0$ against H_A: $\mu=0$ and the BF obtained from testing H_M: $\mu>0$ against H_A: $\mu<0$. The latter BF (for H_M: $\mu>0$ against H_A: $\mu<0$) could also not be computed directly in JASP. Instead, we computed

it as the ratio of BF_{+0} and BF_{-0}: $\mathrm{BF}_{+-} = \frac{p(D \mid \mu > 0)}{p(D \mid \mu < 0)} = \frac{\left(\frac{p(D \mid \mu > 0)}{p(D \mid \mu = 0)}\right)}{\left(\frac{p(D \mid \mu < 0)}{p(D \mid \mu = 0)}\right)} = \frac{\mathrm{BF}_{+0}}{\mathrm{BF}_{-0}}$. The BF for Condition 5 was computed in a similar way. In all Bayesian t-tests, the default prior was used (a Cauchy distribution with a scale parameter of .707).

Condition	H_M	H_A	Computation of BF_{MA}
1	$\mu < 0$	$\mu = 0$	$BF_{MA} = BF_{-0}$
2	$\mu = 0$	$\mu \neq 0$	$BF_{MA} = BF_{01}$
3	$\mu > 0$	$\mu \leq 0$	$BF_{MA} = \frac{1}{2} BF_{+0} + \frac{1}{2} \frac{BF_{+0}}{BF_{-0}}$
4	$\mu = 0$	$\mu \neq 0$	$BF_{MA} = BF_{01}$
5	$\mu < 0$	$\mu \geq 0$	$BF_{MA} = \frac{1}{2} BF_{-0} + \frac{1}{2} \frac{BF_{-0}}{BF_{+0}}$

Table 1. Overview of hypotheses tested in the Bayesian t-tests. For each condition, we calculated a Bayes Factor BF_{MA}, quantifying the evidence in favor of the hypothesis H_M relative to H_A. Bayes Factors larger than 1 indicate support for the model prediction, while Bayes Factors smaller than 1 indicate evidence against the model prediction.

Bayesian mixed-effects modeling

To evaluate the three competing hypotheses about CDA dynamics on all 5 conditions at once, we fitted Bayesian mixed-effects models to post-movement CDA amplitudes (300 – 1000 ms). Analyses were performed in Python using the *bambi* package with MCMC sampling (4 chains, 2000 draws after 1000 warm-up, target acceptance 0.9). CDA values were expressed in microvolts, and all models included subject-level random intercepts; random slopes were added when condition-specific effects were estimated.

We tested three models that aligned with the theoretical accounts. In all three models, the intercept was fixed to 0, which is motivated by the baselining procedure. In Model

1 (hemispheric transfer), the beta coefficients for Conditions 1 and 5 were fitted as free parameters with a negative prior; the beta coefficient for Condition 3 was fitted as a free parameter with a positive prior; and the beta coefficients in Conditions 2 and 4 were fixed to 0. In Model 2 (uniform CDA), a single beta coefficient with negative prior was fitted across the mean across all five conditions. In Model 3, a beta coefficient with negative prior was fitted to the first condition, while the coefficients for the other four conditions were fixed to 0.

All priors were Half-Normal distributions with $\sigma = 1.5$. Model fits were compared using leave-one-out cross-validation (LOO) with pseudo-BMA model weights; WAIC was used when LOO was unstable.

Robustness checks: To verify that our statistical results did not critically depend on the specific time windows used in the CDA calculations, we also ran the analyses with time windows locked to the movement end for each individual motion angle (resulting in 13 separate time windows instead of 5) and with a single window that was time-locked to the end of post-movement delay. The results (see Supplemental Figure 1 and text) were consistent with the results reported in the main text.

Data availability

All preprocessed data will be made available upon publication.

Code availability

All code will be made available upon publication.

Acknowledgements

This work was funded by ERC starting grant 949131 and Swedish research council project grant VR 2022-02328.

References

1. D'Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* **66**, 115–142 (2015).

2. Baddeley, A. D. Working memory. *Science* **255**, 556–559 (1992).

3. Stokes, M. G. "Activity-silent" working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci (Regul Ed)* **19**, 394–405 (2015).

4. Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).

5. Lundqvist, M. *et al.* Gamma and beta bursts underlie working memory. *Neuron* **90**, 152–164 (2016).

6. Rose, N. S. *et al.* Reactivation of latent working memories with transcranial magnetic stimulation. *Science* **354**, 1136–1139 (2016).

7. Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J.-D. Cortical specialization for attended versus unattended working memory. *Nat. Neurosci.* **21**, 494–496 (2018).

8. Sprague, T. C., Ester, E. F. & Serences, J. T. Restoring latent visual working memory representations in human cortex. *Neuron* **91**, 694–707 (2016).

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

10. Sauseng, P., Griesmayr, B., Freunberger, R. & Klimesch, W. Control mechanisms in working memory: a possible function of EEG theta oscillations. *Neurosci. Biobehav. Rev.* **34**, 1015–1022 (2010).

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

619 12. Brincat, S. L. *et al.* Interhemispheric transfer of working memories. *Neuron*
620 **109**, 1055-1066.e4 (2021).

621 13. Draschkow, D., Nobre, A. C. & van Ede, F. Multiple spatial frames for
622 immersive working memory. *Nat. Hum. Behav.* **6**, 536–544 (2022).

623 14. Vogel, E. K., McCollough, A. W. & Machizawa, M. G. Neural measures reveal
624 individual differences in controlling access to working memory. *Nature* **438**,
625 500–503 (2005).

626 15. Vogel, E. K. & Machizawa, M. G. Neural activity predicts individual
627 differences in visual working memory capacity. *Nature* **428**, 748–751 (2004).

628 16. Kuo, B.-C., Stokes, M. G. & Nobre, A. C. Attention modulates maintenance of
629 representations in visual short-term memory. *J. Cogn. Neurosci.* **24**, 51–60
630 (2012).

631 17. Feldmann-Wüstefeld, T., Vogel, E. K. & Awh, E. Contralateral delay activity
632 indexes working memory storage, not the current focus of spatial attention. *J.*
633 *Cogn. Neurosci.* **30**, 1185–1196 (2018).

634 18. Wilken, P. & Ma, W. J. A detection theory account of change detection. *J. Vis.*
635 **4**, 1120–1135 (2004).

636 19. Bays, P. M., Catalao, R. F. G. & Husain, M. The precision of visual working
637 memory is set by allocation of a shared resource. *J. Vis.* **9**, 7.1-11 (2009).

638 20. van den Berg, R., Awh, E. & Ma, W. J. Factorial comparison of working
639 memory models. *Psychol. Rev.* **121**, 124–149 (2014).

640 21. Luria, R., Balaban, H., Awh, E. & Vogel, E. K. The contralateral delay activity
641 as a neural measure of visual working memory. *Neurosci. Biobehav. Rev.* **62**,
642 100–108 (2016).

643 22. Günseli, E. *et al.* EEG dynamics reveal a dissociation between storage and
644 selective attention within working memory. *Sci. Rep.* **9**, 13499 (2019).

645 23. Drew, T., Horowitz, T. S., Wolfe, J. M. & Vogel, E. K. Delineating the neural
646 signatures of tracking spatial position and working memory during attentive
647 tracking. *J. Neurosci.* **31**, 659–668 (2011).

648 24. Drew, T., Mance, I., Horowitz, T. S., Wolfe, J. M. & Vogel, E. K. A soft handoff
649 of attention between cerebral hemispheres. *Curr. Biol.* **24**, 1133–1137
650 (2014).

651 25. Berggren, N. & Eimer, M. Does Contralateral Delay Activity Reflect Working
652 Memory Storage or the Current Focus of Spatial Attention within Visual
653 Working Memory? *J. Cogn. Neurosci.* **28**, 2003–2020 (2016).

654 26. Grubert, A. & Eimer, M. Does visual working memory represent the predicted
655 locations of future target objects? An event-related brain potential study.
656 *Brain Res.* **1626**, 258–266 (2015).

657 27. Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E. & Vogel, E. K. Dissecting the
658 Neural Focus of Attention Reveals Distinct Processes for Spatial Attention
659 and Object-Based Storage in Visual Working Memory. *Psychol. Sci.* **30**, 526–
660 540 (2019).

661 28. Luria, R. & Vogel, E. K. Come together, right now: dynamic overwriting of an
662 object's history through common fate. *J. Cogn. Neurosci.* **26**, 1819–1828
663 (2014).

664 29. Balaban, H. & Luria, R. The number of objects determines visual working
665 memory capacity allocation for complex items. *Neuroimage* **119**, 54–62
666 (2015).

667 30. Balaban, H. & Luria, R. Neural and behavioral evidence for an online
668 resetting process in visual working memory. *J. Neurosci.* **37**, 1225–1239
669 (2017).

670 31. Wang, Y., Lin, X., Ye, C., Parviainen, T. & Liu, Q. Contralateral Delay Activity
671 Tracks Storage Load in Visual Working Memory: Evidence from the Multiple

Object Tracking Task. *J. Cogn. Neurosci.* 1–10 (2025) doi:10.1162/jocn_a_02325.

32. Styrkowiec, P. P. *et al.* Item-based Parsing of Dynamic Scenes in a Combined Attentional Tracking and Working Memory Task. *J. Cogn. Neurosci.* 1–17 (2025) doi:10.1162/jocn.a.96.

33. Awh, E. & Vogel, E. K. Working memory needs pointers. *Trends Cogn Sci (Regul Ed)* **29**, 230–241 (2025).

34. Emrich, S. M., Salahub, C. & Katus, T. Sensory Delay Activity: More than an Electrophysiological Index of Working Memory Load. *J. Cogn. Neurosci.* **35**, 135–148 (2022).

35. Gao, Z. *et al.* Storing fine detailed information in visual working memory--evidence from event-related potentials. *J. Vis.* **9**, 17 (2009).

36. Luria, R., Sessa, P., Gotler, A., Jolicoeur, P. & Dell'Acqua, R. Visual short-term memory capacity for simple and complex objects. *J. Cogn. Neurosci.* **22**, 496–512 (2010).

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

38. Thibeault, A. M. L., Stojanoski, B. & Emrich, S. M. Investigating the effects of perceptual complexity versus conceptual meaning on the object benefit in visual working memory. *Cogn. Affect. Behav. Neurosci.* **24**, 453–468 (2024).

39. Xie, W. & Zhang, W. Familiarity Speeds Up Visual Short-term Memory Consolidation: Electrophysiological Evidence from Contralateral Delay Activities. *J. Cogn. Neurosci.* **30**, 1–13 (2018).

40. Kandemir, G., Wilhelm, S. A., Axmacher, N. & Akyürek, E. G. Maintenance of color memoranda in activity-quiescent working memory states: Evidence from impulse perturbation. *iScience* **27**, 109565 (2024).

41. Ishihara, S. Tests for colour-blindness. (1960).

42. Ngiam, W. X. Q., Adam, K. C. S., Quirk, C., Vogel, E. K. & Awh, E. Estimating the statistical power to detect set-size effects in contralateral delay activity. *Psychophysiology* **58**, e13791 (2021).

43. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).

44. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).

45. Kleiner, M. *et al.* What's new in psychtoolbox-3. *Perception* (2007).

46. Acerbi, L. & Ma, W. J. Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. *Advances in Neural Information Processing Systems* (2017).

47. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).

48. Ablin, P., Cardoso, J.-F. & Gramfort, A. Faster independent component analysis by preconditioning with hessian approximations. *IEEE Trans. Signal Process.* **66**, 4040–4049 (2018).

49. Hakim, N., Feldmann-Wüstefeld, T., Awh, E. & Vogel, E. K. Controlling the flow of distracting information in working memory. *Cereb. Cortex* **31**, 3323–3337 (2021).

50. Emrich, S. M., Al-Aidroos, N., Pratt, J. & Ferber, S. Visual search elicits the electrophysiological marker of visual working memory. *PLoS ONE* **4**, e8042 (2009).

51. Goodman, S. A dirty dozen: twelve p-value misconceptions. *Semin. Hematol.* **45**, 135–140 (2008).

52. Wagenmakers, E.-J. *et al.* Bayesian inference for psychology. Part II: Example applications with JASP. *Psychon. Bull. Rev.* **25**, 58–76 (2018).

53. Wagenmakers, E.-J. *et al.* Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* **25**, 35–57 (2018).

54. JASP Team. JASP (Version 0.18.3)[Computer software]. (2024).