Title: Working memory representations are spatially transferred in cortex during mental manipulations

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 author

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 EEG data as human subjects engaged in a task requiring continuous updates to the locations of objects retained in WM. Analysis of contralateral delay activity (CDA) revealed that mental representations moved across cortex in real time as their remembered locations were updated.

Main: 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]. Furthermore, WM representations are typically studied under relatively static settings, while in real life WM representations are continuously updated to support behavior[12,13]. Consider playing football. You glance up the field to locate opponents, then refocus on the ball. Running up the field, you must remember both where your opponents were and their movements relative to your own position. Early studies have showed that the cortical location of spatial WM representations may shift as we move relative to remembered locations[12,13]. In this regard 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]. Based on these findings, it seems likely that WM object representations transfer across cortex when humans mentally update their remembered locations. Yet, electrophysiological evidence for this is lacking. Here, we aim to bridge that gap by using a novel experimental paradigm combined with electroencephalogram (EEG) measures.

Specifically, we utilized 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 objects, and can therefore be used as a correlate of WM load[14]. Moreover, it has been shown that CDA is sensitive to the selection of a subset of particular items after initial encoding[16], but a critical question that remains unaddressed is what happens to the CDA when subjects mentally manipulate the locations of remembered objects: does it track such movements, is it unaffected by them, or is it perhaps even eliminated by them? A secondary goal of our study was to resolve this issue.

We recorded EEG from healthy humans during a delayed-estimation task[17] 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 circles (two on each side) were then shown, balancing visual inputs and ensuring hemisphere activity differences were due to memory-related processes. After 500 ms, the four circles were replaced by colorless outlines which, 1s thereafter, started moving along an (invisible) circle (180°/s, clockwise or counterclockwise, 0-360 degrees, with 13 possible distances; Figure 1b). Finally, 3s after movement onset, one of the circles 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 subjects mentally updated object positions in WM in real time.

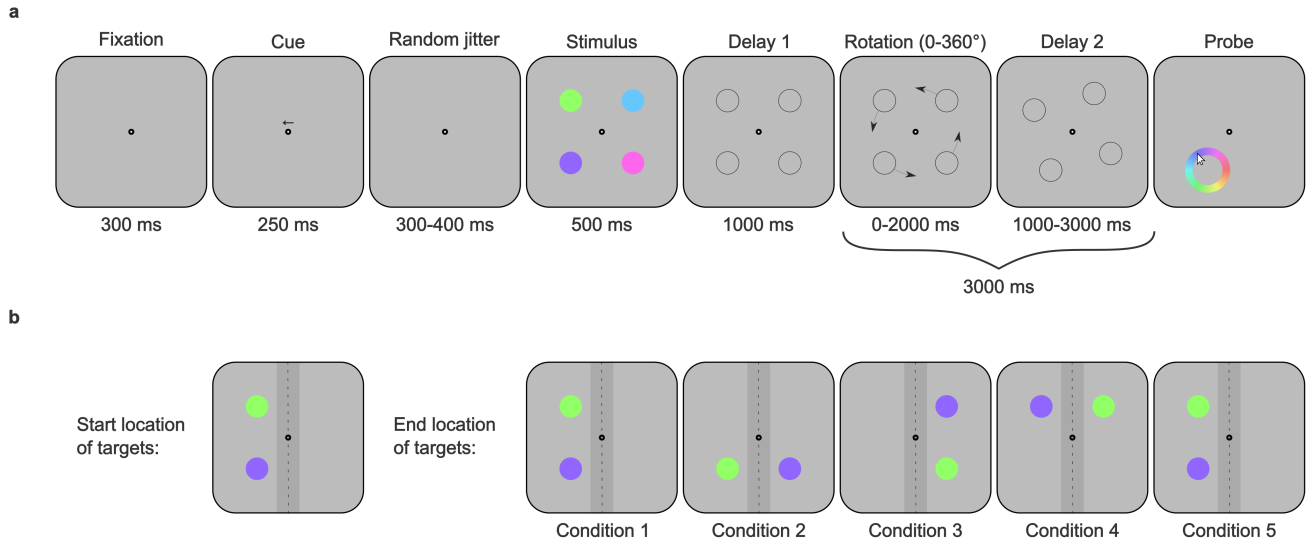

Figure 1. Task design of the experiment. a. Trial structure. Each trial started with a fixation period of 300 ms, followed by a cue indicating the to-be-attended side (left or right). After a random jitter of 300-400 ms, a memory array of four colored circles appeared, with two circles on each side of the fixation point. After 500 ms, the colored circles 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 $K \in \{0,\ 30,\ ...,\ 360\}$ degrees (13 possible alternatives, with randomized sign, 180 deg/s). Once movement was complete, the outlines remained in their new positions for at least 1000 ms. Then, all the outlines disappeared, and a color wheel was presented at the post-movement location of one of the memorized circles, prompting recall. Response was submitted via mouse click anywhere on the color wheel. **b.** Schematic illustration of the five experimental conditions, defined by the laterality of the post-movement stimulus locations: Condition 1 – both circles remained in the original hemifield; Condition 2 – one of the circles transferred to the other hemifield; Condition 3 – both circles transferred to the other hemifield; Condition 4 – both circles transferred to the other hemifield and then one of them returned to the original hemifield; and Condition 5 – both circles 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 dva) where the circles 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 circles (swapping errors[18]) more often with larger movements. In fact, fitting the data with an equal-precision model with non-target responses[19] revealed that the performance degradation could be entirely explained as a result of increasing swapping rates at larger movements, while memory precision remained stable (Figure 2b). Importantly, even at the largest movement angle subjects showed accurate performance: estimated swap rates were relatively low (around 10%) and memory fidelity was high. Hence, they must have been updating object position in WM in real time.

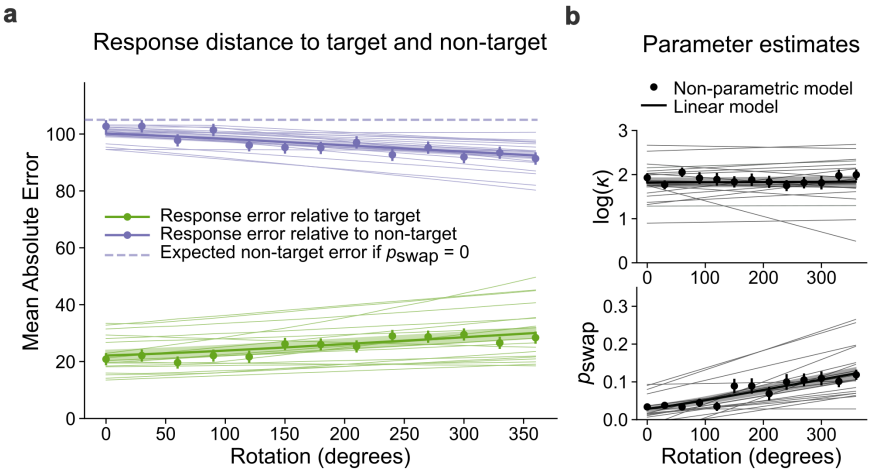

Figure 2. **Effect of stimulus movement on response accuracy. a.** Subject-averaged response errors as a function of movement angle. The error relative to the target (green markers) gradually increased with movement angle (Pearson's r=.88, p<.001), while the error relative to the non-target item (purple 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 s.e.m. **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=0.92, p<.001), there was no evidence that memory precision was affected by motion angle (r=−.02, p=.93).

Next, we computed CDA from the EEG data to investigate the neural underpinning of this mental updating. Under the hypothesis that the movement manipulation 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 circles moved from one hemisphere into the other (Figure 3a): right before motion onset, the CDA was expected to be negative, as in previous studies using static objects; in trials where both circles stayed in the original hemifield (0 or 30° rotation, Condition 1), CDA should remain negative; if one circle moved over to the other hemifield (60 – 120° rotation, Condition 2), CDA should disappear, because the two WM objects would be in opposite hemispheres and cancel each other out; if both circles crossed the midline (150 – 210° rotation, Condition 3) CDA should change sign to positive; if one circle then moved back into the original hemifield, CDA should again disappear (240 – 300° rotation, Condition 4); finally, in trials where both circles moved into the opposite hemifield and then back into the original one (330 or 360° rotation, Condition 5), CDA should become negative again, as if no rotation had taken place. An alternative hypothesis is that CDA reflects only initial encoding and does not track mental updates of WM representations (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), and p-values, supported the prediction of Model 1 (Figure 3c, n=19 in all tests). 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 4.59 (Condition 2), 44.5 (Condition 3), and 3.56 (Condition 4). Likewise, Model 1 was favored in both conditions where its prediction differed from that of Model 3, with BFs of 3.79 (Condition 3), and 1.69 (Condition 5). Thus, the CDA data support the notion that the WM representations moved across cortex in real time as the mental movement progressed. This is, as far as we know, the first direct demonstration of this phenomenon.

Since CDA tracks WM load rather than spatial attention[20,21], the results are unlikely to be explained by the tracking of the circles per se. Moreover, since we found no evidence for Models 2 and 3, where CDA no longer tracks WM representations after they have been updated, our findings provide support for CDA as a correlate of WM representations beyond static short-term memory traces. While CDA has been shown to track the number of relevant items, and not distractors[14,22], it has not been demonstrated to track transformed WM contents before. In fact, an earlier study indicated that CDA does not change sign when objects are encoded in one hemifield and later tested in the opposite[23]. We suspect the effect observed here requires a task with strong emphasis on updating contents in real time to prevent updating of WM contents during the test period itself. Consistently, including subjects with poor accuracy on the behavioral task reduced the effect (Supplemental Figure 2). Most importantly, our findings demonstrate that individual WM representations are not stored in fixed cortical locations. Instead, the location, and thus participating neurons, change over time as the environment moves relative to the subject. Relatedly, we recently provided evidence from macaques that the spatial dimensions of cortex are used to assign the status of WM objects in other circumstances, such as their serial order[24]. The brain therefore appears to deal with dynamic environments with equally dynamic memory representations[3].

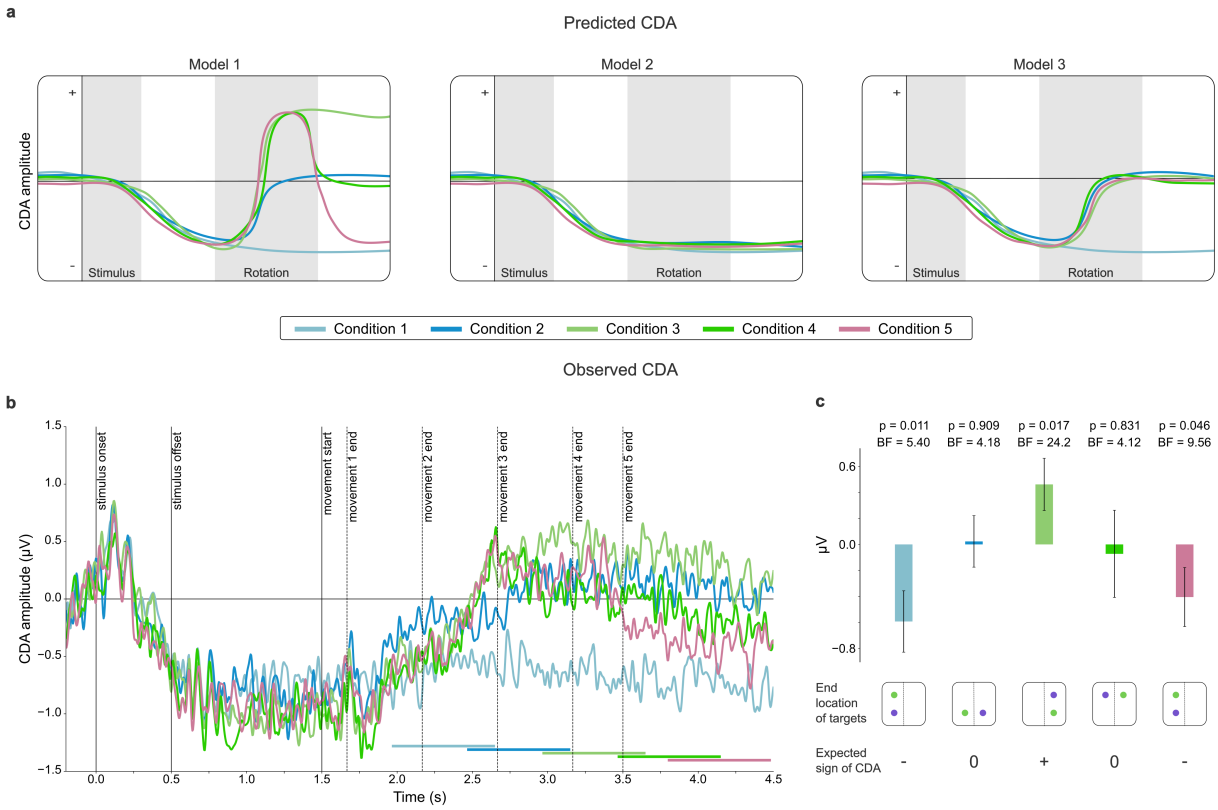

Figure 3. **CDA following movement of WM objects. a.** Cartoon of model predictions of how CDA changes following the mental relocation of the circles. In the main model, Model 1, CDA changes depending on the hemifield the circles end up in post rotation (condition 1-5). For condition 1 and 5 the circles end up in the initial hemifield and CDA should be negative. For condition 2 and 4, there is one circle each hemifield and there should be no CDA. For condition 3, both circles end up in the opposite hemifield and CDA should be positive. As alternatives, Model 2 predict that CDA tracks the initial memory trace and not the updated one (meaning CDA remain unchanged) and Model 3 predicts CDA disappears completely after the update of WM contents occur (meaning CDA only remains when both circles stay in the initial hemifield). **b.** Measured CDA per condition. Dotted vertical lines denote the longest possible time when movement ended per condition. Colored horizontal bars denote 300-1000 ms post movement per condition, used to measure CDA statistically. **c.** *Top:* average CDA post movement per condition, taken from the time periods indicated in **b**. Error bars denote SEM. The p-values were obtained based on t-tests against the null hypothesis of CDA being 0. 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 circles started to the left). *Bottom:* sign of CDA predicted by Model 1.

Methods

Subjects

A total of 26 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, their color vision was tested using the short (6-plate) version of the Ishihara's Test for Color Deficiency[25]. The test was considered passed if all plates were read normally. All subjects passed the test.

Data from three subjects were excluded from analysis due to excessive noise in the EEG and data from an additional four 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, 8 male, 3 left-handed), who each performed 480 trials across 5 conditions (see below). According to a recent meta-analysis[26], 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).

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 task and stimuli were generated and displayed using MATLAB with the Psychophysics Toolbox Version 3 (PTB-3) extension[27–29]. 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.

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 circles on a gray (RGB = 128 128 128) background. The colors of the circles 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. 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.

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. 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 circles (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 circles 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 circles and the distance between the circles 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 circles' 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, the outlines of the circles disappeared and a memory probe – a color wheel – appeared at the new location of one of the previously cued circles. Subjects were asked to recall the color of that circle 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. The intertrial interval was 1000 ms.

Conditions: We defined five experimental conditions, based on the final locations of the two cued circles: Condition 1 (0°, 30°), in which both circles remained in the original hemifield; Condition 2 (60°, 90°, 120°), in which one of the circles transferred to the other hemifield; Condition 3 (150°, 180°, 210°), in which both circles transferred to the other hemifield; Condition 4 (240°, 270°, 300°), in which both circles transferred to the other hemifield and then one of them returned to the original hemifield; and Condition 5 (330°, 360°), in which both circles 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 was that accurate performance was possible only if subjects mentally updated locations of stimuli held in WM as the circle 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 4 exclusions (compare Figure 2a with Supplemental Figure 3 to see the performance of the excluded subjects relative the rest). To make sure that our main findings do not critically depend on this exclusion, we reran the analysis with these four 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.

Behavioral analysis

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[19] 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[30].

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[31]. 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 electrodes kept in the main CDA analysis. All electrodes were then re-referenced offline to the average of the mastoids. 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[32] 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 (memory period 1), 0-2000 ms of stimuli movement, and 1000-3000 ms of post-movement delay (memory period 2).

Automatic trial rejection was used to remove trials contaminated with muscle artifacts, channels blocking, slow drifts, and step-like voltage changes. The detection algorithms and rejection criteria were implemented as described in[33]. The remaining trials were manually inspected to make sure no obvious artifacts had been missed. Subjects with fewerthan 300 clean trials were excluded from analyses. For the remaining 19 subjects, we rejected on average 16.1% (SD = 8.6%) of trials, resulting in an average of 403 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[34]: 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 memory period 1 (analyzed in the Supplemental Material) and the mean amplitude between 300 and 1000 ms post movement end per condition in memory period 2 (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 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[35]. 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. An advantage of BFs is that they can quantify evidence both for and against any well-defined hypotheses, including the null[36]. 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 "μ=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[37], which allows to test four different types of hypotheses ($\mu<0$, $\mu=0$, $\mu>0$, $\mu\neq 0$) 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: μ>0 against H_A: μ=0 and the BF obtained from testing H_M: μ>0 against H_A: μ<0. The latter BF (for H_M: μ>0 against H_A: μ<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.

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 memory period 2. 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).

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

419 7. Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C. & Haynes, J.-D. Cortical
420 specialization for attended versus unattended working memory. *Nat. Neurosci.* **21**, 494–
421 496 (2018).

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

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

426 10. Sauseng, P., Griesmayr, B., Freunberger, R. & Klimesch, W. Control mechanisms in
427 working memory: a possible function of EEG theta oscillations. *Neurosci. Biobehav. Rev.*
428 **34**, 1015–1022 (2010).

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

431 12. Brincat, S. L. *et al.* Interhemispheric transfer of working memories. *Neuron* **109**, 1055-
432 1066.e4 (2021).

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

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

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

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

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

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

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

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

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

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

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

24. Lundqvist, M. *et al.* Working memory control dynamics follow principles of spatial computing. *Nat. Commun.* **14**, 1429 (2023).

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

461 26. Ngiam, W. X. Q., Adam, K. C. S., Quirk, C., Vogel, E. K. & Awh, E. Estimating the
462 statistical power to detect set-size effects in contralateral delay activity. *Psychophysiology*
463 **58**, e13791 (2021).

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

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

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

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

470 31. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**,
471 267 (2013).

472 32. Ablin, P., Cardoso, J.-F. & Gramfort, A. Faster independent component analysis by
473 preconditioning with hessian approximations. *IEEE Trans. Signal Process.* **66**, 4040–
474 4049 (2018).

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

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

479 35. Goodman, S. A dirty dozen: twelve p-value misconceptions. *Semin. Hematol.* **45**, 135–
480 140 (2008).

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

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