Flexible Working Memory in the Peripheral Nervous System

Sihan Yang[1], Yueying Dong[1], Anastasia Kiyonaga[1*]

[1]Dept of Cognitive Science, University of California San Diego, La Jolla, CA 92093, USA
[*]akiyonaga@ucsd.edu

SUMMARY

Working memory (WM) representations that are distributed across the brain can be flexibly recruited to best guide behavior[1–4]. For instance, information may be represented relatively more strongly in visual cortex when a WM task requires fine visual detail, or more strongly in motor cortex when a specific response can be prepared[5–10]. If WM is geared to prospectively guide actions, then we might also expect such task-oriented neural signals to propagate to the peripheral sensors and effectors that realize WM goals. Likewise, there is now evidence that oculomotor signals like saccade biases can track simple visuo-spatial WM features[11,12]. However, it is unclear if such signatures are functionally meaningful, and how much information is contained in them. Here, we test the idea that WM content is adaptively distributed across the nervous system according to behavioral demands. Namely, we test whether visual WM stimulus features are expressed in patterns of both eye and hand movements during a WM delay, and whether the distribution of such peripheral motor activity shifts with the task context. In a delayed recall task, we manipulated how human participants reported their memory by having them either draw a line or adjust a wheel to match a remembered orientation. Via eye- and stylus-tracking, we found that remembered orientations were decodable from small inflections in both gaze and hand movements during the blank WM delay. Moreover, this decoding strength varied by response format: gaze patterns tracked memorized features relatively better in the *wheel* condition (vs. *draw*), while hand movements were better in the *draw* condition (vs. *wheel*). Individuals who showed greater *wheel* benefits in gaze-based decoding also showed greater *draw* benefits in hand-based decoding, suggesting a strategic processing shift to the more relevant system. Therefore, visually encoded WM contents may be adaptively allocated to the most task-relevant motor effectors, balancing WM representations across peripheral activity according to behavioral needs.

RESULTS AND DISCUSSION

The crux of working memory (WM) lies in its flexibility to apply mental representations toward myriad current goals. Accordingly, WM content may be held in different neural response patterns depending on how it will be used[4,6,9,13,14]. Rather than sustain a precise copy of encoded sensory features, an adaptive WM code may incorporate critical features of the task context. Therefore, while the quality of visual WM is often tested by having participants adjust a visible prompt to match their memory (e.g., click locations on a wheel, continuously adjust the angle of a bar)[15–18], this test format may shape the memory itself. Here, to examine how motor response demands might influence visual WM maintenance, we had participants report their memory using a

digital stylus and tablet. The critical manipulation was whether they were asked to freely recall the remembered stimulus by drawing it from memory, or whether they used a more conventional response wheel to make a continuous adjustment report. We recorded both hand and eye movements throughout the task, to test whether WM content would be flexibly expressed across relevant motor effectors.

Drawing enables precise free recall in working memory

Participants (n=36) completed the delayed report task (Figure 1a). They saw two sequentially presented Gabor patches and were cued to remember the orientation of either one or both patches over a WM delay of several seconds. They then reported the remembered angle(s) of cued stimuli by either drawing a line (in *draw* blocks) or adjusting a wheel to match their memory (in *wheel* blocks; see Methods & Stats for all trial conditions).

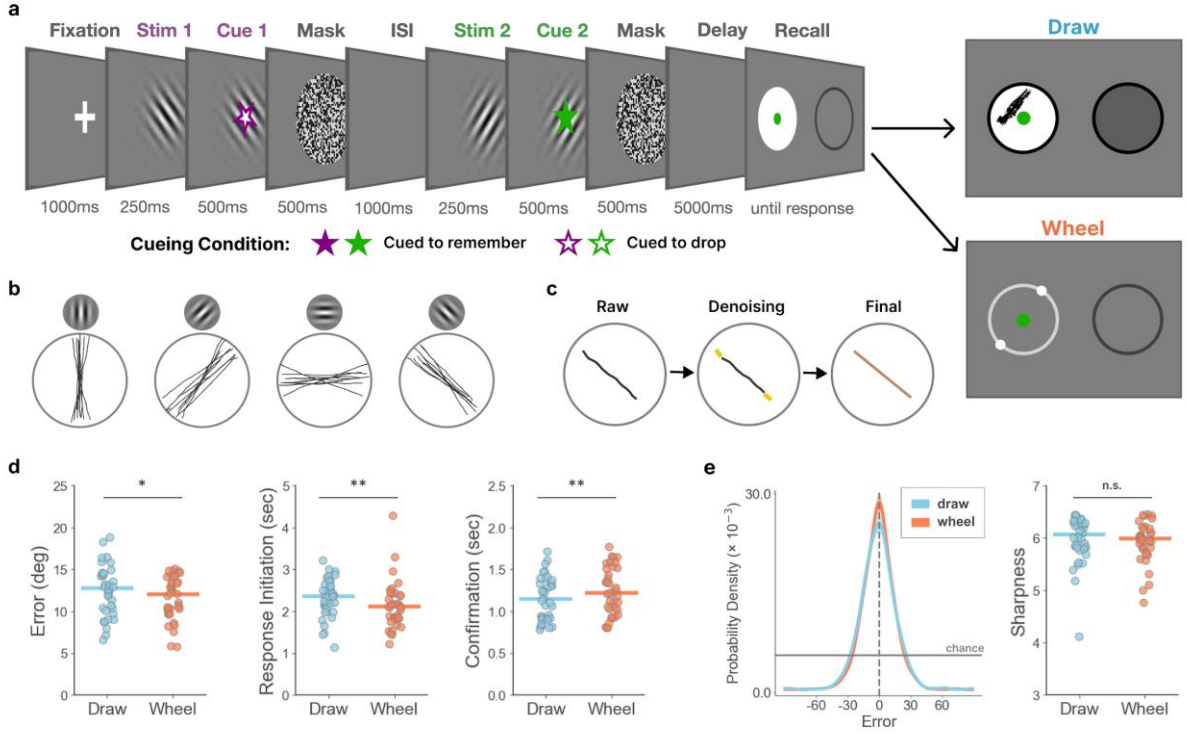

Figure 1: Experimental design and behavioral results. (a) *On each trial, the WM sample stimuli were two sequentially-presented Gabor patches, randomly varying in orientation. Shortly after each stimulus onset, a cue appeared indicating whether to remember (filled star)*

or drop (empty star) that stimulus. Either one or both stimuli could be cued to remember, but cues appeared after stimulus onset so that all stimuli would be initially encoded. In the test phase, two side-by-side circular response regions appeared, and a colored dot at the center of a region indicated which item should be recalled there. In addition to the color cue, the shading of the response region indicated if it was editable: only an illuminated wheel or input field could accept a response. In the 'draw' condition, participants saw a white input field where they could freely draw a line to report the remembered orientation. In the 'wheel' condition, they saw a light grey circle with two white dots opposite each other along the circumference, representing the endpoints of an oriented line for them to adjust. In both cases, they used a digital stylus and tablet to make the response, and on each trial they confirmed their response by pressing a 'continue' button to advance. The response format was instructed at the beginning of each block, and draw and wheel block conditions were given in quasi-random order (see Methods & Stats). ***(b)*** *Example drawings for a range of stimulus orientations (±1°), each from a sample of 9 participants.* ***(c)*** *The procedure for processing drawing responses. Each raw drawing was first denoised by removing jittering from the beginning and end of the response period (removed segments are shown in yellow in the 'denoising' panel). Then the response angle was calculated as the average angle of the remaining strokes (a cleaned drawing shown in brown in the 'final' panel).* ***(d)*** *Behavioral performance for both draw and wheel conditions, showing response error (i.e., angular difference between the response and the correct stimulus angle), response initiation time (i.e., time to first stylus contact in the response field), and response confirmation time (i.e., time from completing response to pressing the "continue" button). Point clouds show individual subject means and horizontal lines mark the group median. Asterisks mark significant Wilcoxon signed rank tests. (e) Smoothed distributions of response errors for the draw and wheel conditions across the group, and point clouds showing individual values for the sharpness of the fitted curve (reflecting response precision).*

After cleaning the drawing data (Figure 1b-c), we tested the differences between *draw* and *wheel* conditions with both *t*-tests and Wilcoxon signed ranks tests. Both tests showed similar results (see Methods & Stats) but we report the latter in the main text, because several conditions did not meet normality assumptions for *t*-tests.

We found that response errors were overall larger in the *draw* condition than in the *wheel* condition ($\boldsymbol{W = 171, n = 36, p = 0.010}$; Figure 1d l left). Responses were also initiated more slowly in the *draw* condition (vs. *wheel*; $\boldsymbol{W = 148, n = 36, p = 0.003}$; Figure 1d middle), but they were confirmed more quickly after completion ($\boldsymbol{W = 164, n = 36, p = 0.007}$; Figure 1d right). We also examined the distribution of response errors for each participant, and we used the sharpness of these distributions around zero error to further quantify performance (Figure 1e; see Methods & Stats). This measure of precision did not differ between the *draw* and *wheel* response format conditions ($\boldsymbol{W = 278, n = 36, p = 0.396}$). Response errors were overall greater when remembering two stimuli (vs. one), but this load effect did not interact with response format (see Supplement), so we collapsed across load conditions for most analyses.

In sum, *draw* responses showed slightly greater errors, but were overall highly precise and did not differ in sharpness from *wheel* responses. Drawing is therefore an effective tool to capture WM quality[19,20]. However, the results suggest that small differences in motoric response demands can influence WM performance. Namely, *draw* responses showed slower initiation times, but were finalized faster than *wheel* responses. This could indicate that subjects took longer to prepare the relatively elaborated motor plan for a drawing, but were then more confident to confirm it after completion. These behavioral differences could stem from differences in the motor control required to execute each response, or the memory demands inherent to the probe. For instance, drawing may index true free recall, because it is executed without an informative prompt[21]. Wheel responses may instead be subtly shaped by their dependence on adjusting a visible cue[22]. Growing evidence suggests that such task context factors might influence the neural WM content representation[13,14], and below we will examine whether they shape peripheral motor activity during the WM delay as well.

Gaze and hand motion patterns maintain WM features

WM is tightly intertwined with action, and neural representations of WM content can often reflect motor response preparation[8,9,23–26]. Simple features of remembered information can also now be detected in oculomotor signatures outside the brain. For instance, during a WM delay, gaze tends to veer in the direction where a relevant WM stimulus was encoded[11], and pupil size tracks its remembered brightness[27–30]. WM gaze biases can also reflect the remembered angle of objects in continuous orientation space[12], hinting that detailed WM content information may be expressed in peripheral sensors and effectors. This suggests that WM may recruit distributed structures across the nervous system that are relevant for current goals. If so, in addition to oculomotor activity, we would expect manual motor effectors to be modulated by corresponding WM demands.

Therefore, we tested whether memorized orientation features could be detected in both gaze and hand motion patterns during a WM delay, first collapsed across the *draw/wheel* task condition (Figure 2). From eye movement recordings, we determined each individual's baseline gaze location, and extracted distributions of both gaze offset angle and 2D fixation positions relative to that baseline (Figure 2a). From stylus recordings, we extracted the distribution of movements along each direction (Figure 2b). During the blank WM delay interval, we observed that small inflections in both gaze positions and hand trajectories appeared to vary systematically with the memorized orientations. For instance, gaze foci rotated around central fixation as the oriented stimuli rotated (Figure 2a), and there were more hand motions along the axis of the orientation held in memory (Figure 2b). These descriptive results illustrate that motor

activity appears to track WM content information, and we next formally quantified these patterns.

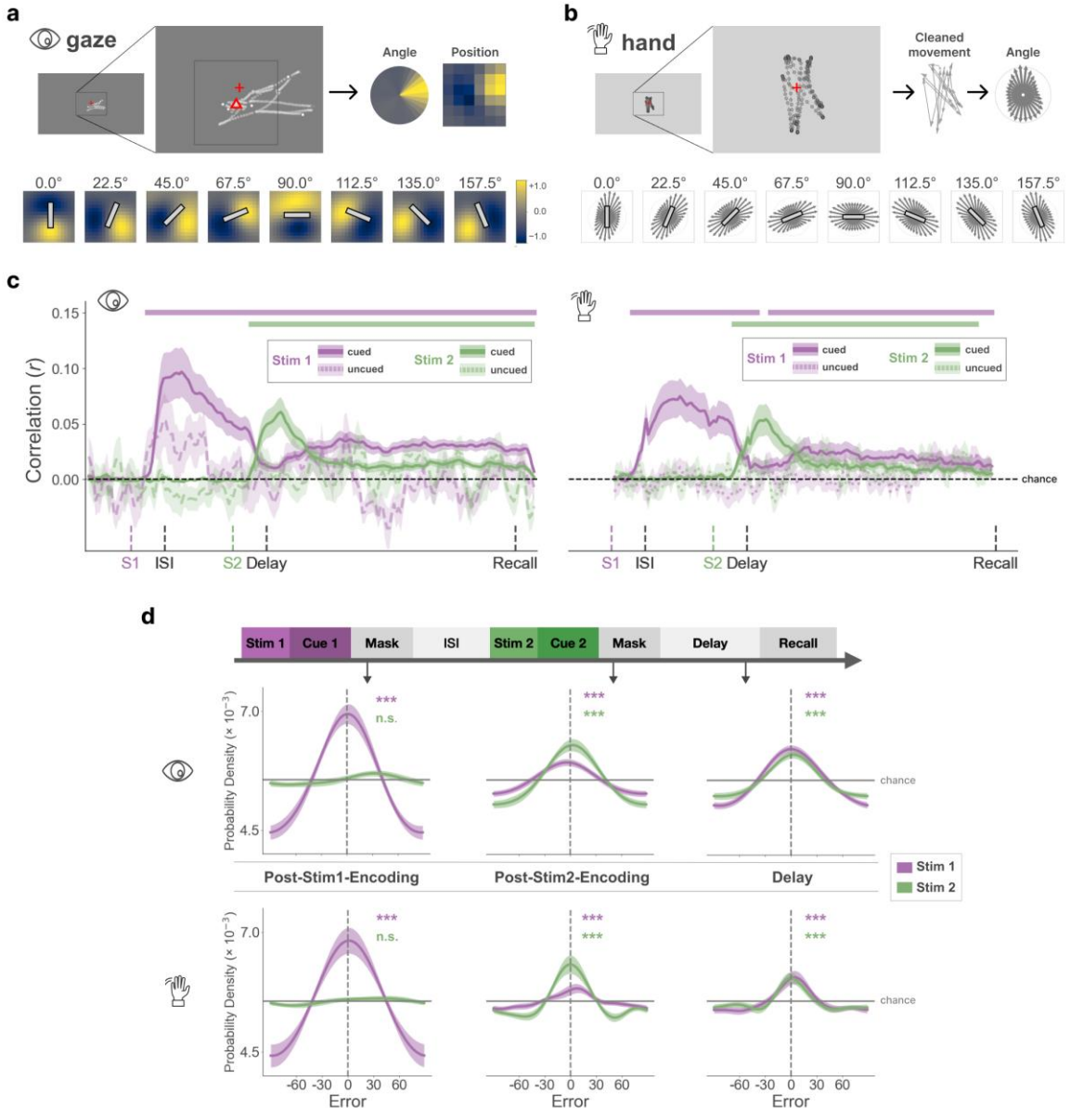

Figure 2: Tracking memorized orientations in gaze and hand motion patterns during WM. (a) *Top: features extracted from eye-tracking data during the delay period of an example trial. For each participant, we computed their median gaze location as a baseline (denoted by*

the red triangle in the figure). Then, after baseline subtraction, we computed the distributions of both gaze offset angle and 2D spatial fixation position. Bottom: Normalized 2D heatmaps of gaze positions during the delay period, aggregated across participants and task conditions. The orientation of the bars indicates the bin center for each subset of trials. Yellower color indicates more frequent gaze visitations at that location. ***(b)*** *Top: features extracted from stylus movement data during the delay period of an example trial. For each participant, we computed the angular distribution of cleaned hand movements. Bottom: median-centered angular distributions of hand movements during the delay period, aggregated across participants and task conditions. The length of each vector is proportional to the magnitude of hand movements in that direction. The orientation of the bars indicates the bin center for each subset of trials.* ***(c)*** *Time course of representational similarity analyses for gaze (left) and hand motion (right) patterns. The y-axis shows the strength of correlation between the motor activity and stimulus similarity matrices. Bold lines show correlations for stimuli that were cued to remember, dashed lines show correlations for stimuli that were perceived but cued to drop, and error bands reflect standard error of the mean. Shaded horizontal bars mark time periods when the correlation was significantly different from baseline (cluster-based permutation test, $p < .05$).* ***(d)*** *Inverted encoding modeling (IEM) of stimulus orientation, for cued items, during key trial epochs. After training, the model predicts the remembered orientation on each trial as a probability distribution from 0° to 180°. Distributions were recentered around 0° orientation (clockwise positive), and each plot depicts the distribution of model errors (i.e., the 'reconstruction'), averaged across trials and participants. Sharpness around 0° reflects the quality of stimulus reconstruction, which we take as the strength of evidence for a given orientation. Reconstructions from a leave-one-subject out model are shown for both gaze (top) and hand motion (bottom) data. Stars denote whether the sharpness is significantly greater than expected from a uniform distribution (i.e., zero).*

To gauge the quality of WM stimulus information in the eye and hand movement patterns, we applied two multivariate techniques that are commonly used for neuroimaging analysis. First, we conducted representational similarity analyses (RSA[31]) to evaluate whether the peripheral signals discriminated between different stimulus values (i.e., different orientations). This approach assesses whether the physiological activity patterns vary in a similar way to the remembered stimuli. If so, for instance, we would expect two orientations that are similar to each other to also evoke similar eye or hand movement patterns during WM[12]. To test this, we examined the correlation between the stimulus similarity structure (i.e., the matrix of pairwise similarity between stimulus orientations) and the similarity structure in the peripheral motor signatures (i.e., the matrix of pairwise cosine similarity between movement feature vectors when remembering different orientations). A strong correlation would indicate that the eye/hand movement patterns track the similarity structure in the stimulus space, and therefore that stimulus-specific information is contained in the peripheral motor signals.

We assessed the timecourse of this correlation across the trial (Figure 2c). We found that, shortly after encoding, stimulus-specific information emerged in both gaze and

hand movement patterns for stimuli that were cued to remember ('cued'), and it remained above chance throughout the entire WM delay interval. Critically, no above-chance information emerged for stimuli that were perceived but then cued to drop ('uncued'). Therefore, both manual and oculomotor signatures only reflect the task-relevant mnemonic content – they are explained by neither lingering sensory-evoked activation nor a reflexive motor response to visual experience for items that are seen but not remembered.

While RSA can tell us that activity patterns reflect the relational structure in the stimulus space, the analysis does not derive the specific item being remembered (in this case, the specific orientation). We therefore also conducted inverted encoding modeling (IEM[32,33]) to reconstruct memorized orientations (Figure 2d; see Methods & Stats for details). IEM learns an association between stimuli and neural activity patterns during training, and then generates predictions about what stimuli are reflected in untrained activity patterns. This approach has been used extensively to reconstruct population-level WM representations from neural activity[34–40], and here we apply it to peripheral motor activity patterns. The method is especially well-suited here to characterize the representation of a given WM item when multiple items are maintained simultaneously. We trained and tested models using both leave-one-subject-out cross-validation (*across-subject*) and 5-fold cross-validation (*within-subject*). Here, we report the across-subject results, and within-subject results are reported in the Supplement.

We evaluated stimulus reconstructions, for cued stimuli, from the error distributions of model predictions. We gauged the quality of reconstruction as the sharpness of this distribution above chance (centered to 0° in orientation space). We found that the relevant stimulus orientation could be reliably reconstructed during key trial epochs (Figure 2d). In the period immediately following the first sample stimulus, that orientation was reliably reconstructed ($p < 0.001$ for stim1, both gaze and hand). In the period immediately following the second stimulus, that stimulus was reliably reconstructed, whereas the first stimulus remained above chance but became relatively dampened ($p < 0.001$ for stim1 and stim2, both gaze and hand). During the blank memory maintenance period, both relevant stimulus orientations could be reliably reconstructed, from both gaze and hand movement patterns (all $p < 0.001$) (see Methods & Stats). Therefore, not only do motor activity patterns systematically vary when remembering different visual stimuli (as shown with RSA), but specific remembered items can be extracted from both gaze and hand signals.

To summarize these analyses, we used two multivariate approaches to quantify the WM information content in gaze and hand movement patterns. We found robust representation of task-relevant mnemonic content in both motor measurements, suggesting that visual WM content is distributed to multiple peripheral effectors during WM maintenance. Stimulus information within individual subjects could also be

reconstructed from models trained on other subjects' data, indicating that these motor WM signatures may be expressed similarly across people.

Motor signatures of WM shift with expected response format

Our theoretical framework predicts that the peripheral nervous system should be flexibly recruited during WM according to the task context. Indeed, there is evidence that cortical signatures of WM content can be flexible to how WM will be tested[3–6,14]. For instance, visuo-spatial WM content may be offloaded to a more motor-like neural representation when a specific response can be prepared[9]. This suggests that motor intentions might mold how or where visual WM content is maintained. In other words, the same visual WM content might be represented differently when it precedes different actions. To test this here, we next asked whether the WM signals in gaze and hand movements vary when a *draw* or *wheel* recall response is required.

Both *draw* and *wheel* task conditions tested memory for the same stimulus feature (angle), and used the same response implement (stylus). Nonetheless, we found that the relative strength of WM stimulus information in gaze and hand signals was adaptively modulated by the response format condition (Figure 3). Namely, the correlation in the gaze RSA was stronger in the *wheel* condition than in the *draw* condition. Conversely, the correlation in the hand motion RSA was stronger in the *draw* condition than in the *wheel* condition (Figure 3a). From this pattern, we might infer that the motor effectors are differently suited to represent WM content information in the two conditions – while the eyes appear to 'prefer' the *wheel* condition, the hands appear to prefer the *draw* condition.

This apparent preference was also evident in the IEM-based stimulus reconstructions from the delay period (Figure 3b). The gaze-based reconstructions showed stronger stimulus evidence (for both stim 1 and stim 2) in the *wheel* condition than the *draw* condition ($W = 202, n = 36, p = 0.039$). Conversely, hand-based reconstruction showed stronger stimulus evidence in the *draw* condition than the *wheel* condition ($W = 157, n = 36, p = 0.005$). The difference between *draw* and *wheel* response formats is therefore reliable across analysis methods. Note, however, that supplemental analyses suggest these differences reflect quantitative differences in representational strength rather than qualitative differences in coding format. Namely, the data from the *draw* condition could be used to successfully reconstruct WM content in the *wheel* condition (and vice versa), indicating that there is some shared representational code between the response formats (see Supplement).

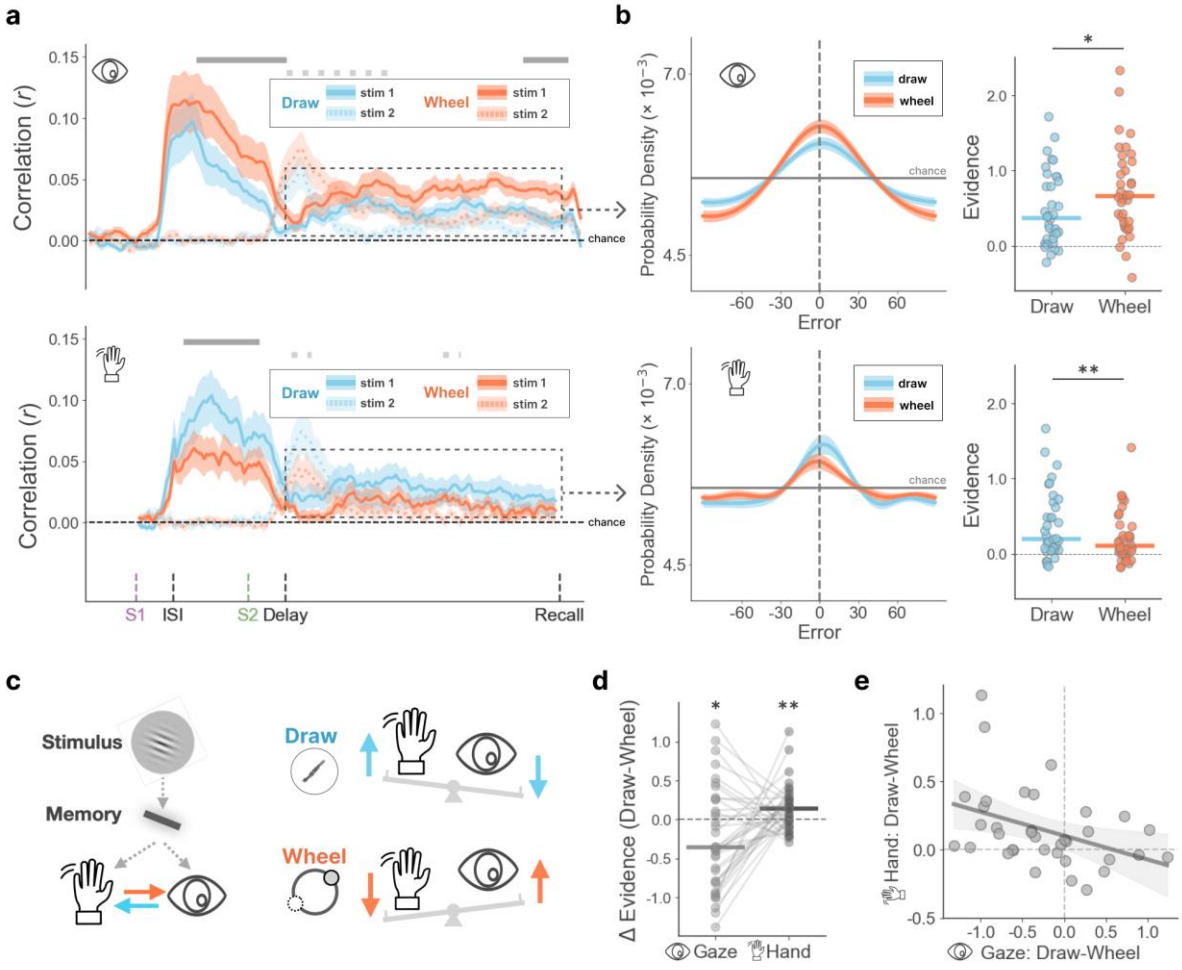

Figure 3: Gaze and hand movement WM representations under different response formats. (a) *Time course of representational similarity for gaze data (top) and hand motion data (bottom), comparing the draw (blue) and wheel (coral) conditions. The solid grey bars mark time points where permutation testing indicated that the correlation for stim 1 differed between conditions; dashed bars mark time points where the correlation for stim 2 differed between conditions.* ***(b)*** *Left: IEM reconstructions of memorized orientations, during the delay period, for draw and wheel conditions (combined across stim 1 and stim 2). Right: Points represent the IEM-based WM stimulus evidence (i.e., the sharpness of the error distributions) for each individual. Horizontal lines reflect the group medians for both the draw and wheel conditions.* ***(c)*** *Schematic depiction of the current framework: WM is adaptively balanced across peripheral effectors based on task demands. As the response format shifts from wheel to draw, the balance favors hand activity over gaze (and vice versa when it shifts from draw to wheel).* ***(d)*** *Difference in IEM evidence between the draw and wheel conditions (draw - wheel)*

for each subject, computed separately for gaze and hand data. The lighter (gaze) and darker (hand) points represent the differences between the blue and coral points in the upper and lower panels of (b), respectively. We refer to each corresponding difference score as the "contextual effect" for that effector. Here a positive contextual effect indicates stronger stimulus evidence in the draw condition, while a negative effect indicates stronger stimulus evidence in the wheel condition. ***(e)*** *Scatterplot shows the correlation between contextual effects in gaze and hand movement patterns, where each point represents an individual subject.*

These findings illustrate that WM content information is distributed out to multiple peripheral systems. Critically, however, the information is adaptively allocated across those systems depending on which is most task-relevant. Namely, the *draw* condition can be considered as manual free recall, in that the response can be executed without visual input. But it may also require a relatively elaborated motor plan. The hands appear to represent WM information relatively more strongly in this *draw* condition (vs. *wheel*). The *wheel* condition, on the other hand, is prompted by a visible probe, and a successful response may require continuously adjusting the probe to match memory. The eyes appear to represent WM information relatively more strongly in this *wheel* condition (vs. *draw*).

If the different response demands do in fact modulate which peripheral effectors are most relevant for WM maintenance, we might expect to observe a hand-eye trade-off in stimulus evidence between the *draw* and *wheel* conditions (Figure 3c). For instance, within an individual, when the response format changes from *wheel* to *draw*, a greater increase in hand signal evidence should be accompanied by a corresponding decrease in gaze signal evidence. To test this, we calculated individual difference scores in IEM evidence, between the *draw* and *wheel* conditions, separately for gaze and hand data (Figure 3d). We refer to each of these difference scores as the "contextual effect" (i.e., the difference between *draw* and *wheel* task contexts), and we examined the correlation between the gaze and hand contextual effects. We found a negative correlation across participants ($\boldsymbol{r(34) = -0.39, p = 0.017}$; Figure 3e). This is consistent with the idea that the gaze and hand WM representations trade off with each other. For instance, participants with a relatively better hand-based reconstruction during *draw* (vs. *wheel*) also showed a relatively worse gaze-based reconstruction during *draw* (vs. *wheel*). This suggests that participants who showed a stronger preference for drawing (over *wheel*) in their hand-based representations, also disfavored drawing in their gaze-based representations to a corresponding degree. Thus the system may balance the distribution of WM maintenance activity across the relevant peripheral effectors.

Collectively, these analyses show that, during a blank WM delay period, the upcoming test demands may modify how information is represented. We found relatively stronger

hand-based WM representations during the *draw* condition (vs. *wheel*), which may reflect the conversion of stimulus information into a manual action plan. We found relatively stronger gaze-based WM representations during the *wheel* condition (vs. *draw*), which could reflect the greater need to rely on a visual representation for stimulus comparison. Thus, the expression of WM feature information in peripheral motor signatures may strategically adapt to expected task demands.

Serial dependence in peripheral motor signals

The previous sections addressed how the motor context of the current task (i.e. *wheel* versus *draw*) shapes the strength of peripheral representations for visual WM. However, our theoretical framework posits that WM representations that serve behavior should reflect an aggregate of contextual forces. For instance, events of the recent past – i.e., temporal context – can also influence WM performance and cortical representations[40–43]. If the observed peripheral WM signatures are more than fleeting motor traces, but instead adaptively anticipate task-relevant behavior, they should also incorporate temporal context. Therefore, lastly, we test whether gaze and hand movement WM signatures exhibit serial dependence.

Serial dependence describes a phenomenon whereby WM reports on a given task trial tend to be biased in the feature direction of a previous stimulus or response, as long as the two are fairly similar[44–46]. For instance, a red object may be seen as more orange if you just saw a similar yellow object (i.e., the red judgement becomes “attracted” toward yellow). This bias is thought to reflect an adaptive temporal smoothing, to strengthen perception and memory in noisy environments[47–49]. If that is the case, we would also expect serial dependence effects to be maximal in the more task-relevant peripheral motor signatures.

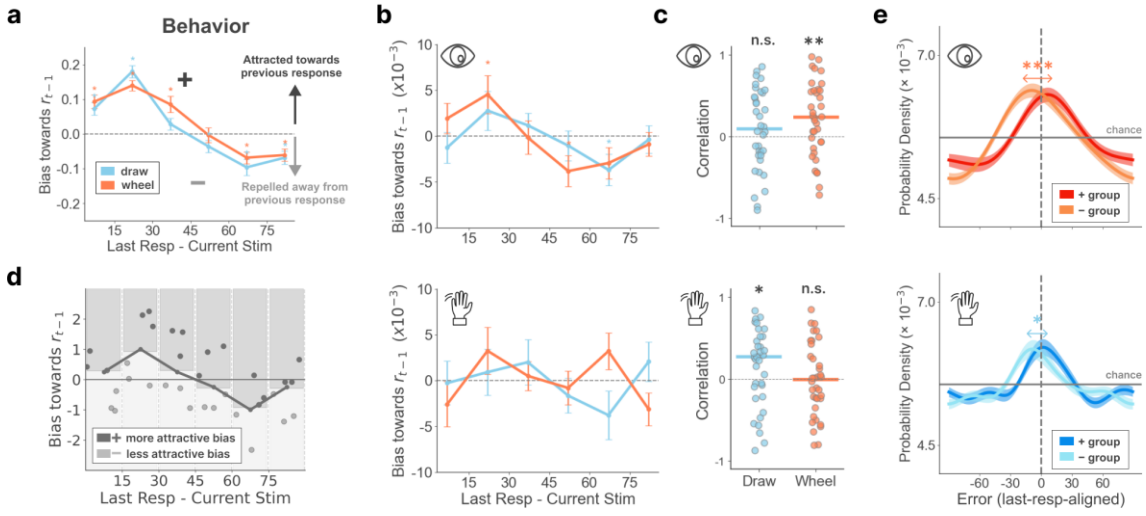

Figure 4: Serial dependence in behavior and peripheral motor WM signatures. (a-b) *Folded signed error as a function of the angular difference between the previous response and the current stimulus (grouped into six bins). Positive y-values reflect an attractive bias toward the previous trial, and negative y-values reflect a repulsive bias away from the previous trial.* ***(a)*** *Bias in behavioral response errors for the wheel (coral) and draw (blue) task conditions.* ***(b)*** *Bias in gaze and hand motion activity during the WM delay, assessed by shifts in the IEM error distributions.* ***(c)*** *Points represent the correlation between serial bias functions in behavior and peripheral activity for each individual. Blue dots represent the correlation between the blue lines in panels (a) & (b), while coral dots represent the correlation between coral lines in panels (a) & (b). Horizontal lines indicate the group median correlations for both the draw and wheel conditions.* ***(d)*** *Diagram illustrating how trials are further grouped based on behavior to examine aggregate stimulus reconstructions. For each subject, trials are first binned by the angular difference between the current target and the previous response (six bins). Within each bin, trials in the upper half of values for that bin are considered relatively more attractive, while trials in the lower half for that bin are considered relatively less attractive. Then stimulus reconstructions are examined for these two groups of more vs. less attractive behavioral responses.* ***(e)*** *IEM reconstructions (error distributions of model predictions) from gaze and hand motion data, grouped according to whether the behavioral responses on that trial were relatively more or less attractively biased (as described in panel c). Reconstructions are shown for the preferred task condition for each motor effector (i.e., wheel condition for gaze, draw condition for hand motions). The distributions are combined across trials and aligned such that positive values (rightward shift) reflect an attractive bias toward the previous trial, and negative values (leftward shift) reflect a repulsive bias away from the previous trial.*

Consistent with prior findings[50–52], we observed an attractive bias in behavioral error when the previous response was similar to the current stimulus, and a repulsive bias when the two were very different (Figure 4a) (see Methods & Stats). We then examined whether the WM activity in peripheral motor signatures tracked with this behavioral pattern.

We first assessed whether the stimulus representations in delay-period gaze and hand motion were shifted toward or away from the remembered stimulus value from the previous trial. Namely, we quantified whether the IEM error distributions were shifted off-center either toward or away from the previous trial's response[40]. As we did for behavior, we examined this IEM shift as a function of the angular difference between the current trial stimulus and previous trial response. We observed an attractive bias when the previous response was similar to the current stimulus, and a repulsive bias when it was dissimilar (Figure 4b), resembling the pattern in behavior. To quantify this resemblance, for each subject, we tested whether the serial bias function in motor effectors (Figure 4b) correlated with the serial bias function in behavior (Figure 4a). We found that the serial bias in gaze-based stimulus reconstructions only tracked the behavioral biases in the "preferred" *wheel* task condition (Figure 4c; *draw*: $W = 403, n = 36, p = 0.139$; *wheel*: $W = 498, n = 36, p = 0.004$). Similarly, the serial bias in hand-based stimulus reconstructions only tracked the behavioral serial bias in the "preferred" *draw* task condition (*draw*: $W = 446, n = 36, p = 0.039$; *wheel*: $W = 324, N = 36, p = 0.559$).

The prior analysis showed that both behavioral and motor serial biases vary with the feature distance between current and previous trials. Finally, we also tested whether motor biases tracked with behavioral biases when controlling for this distance function – for instance, whether relatively more attractive behavioral responses corresponded with more attractive stimulus reconstructions, even within the same range of similarity between current and previous trial orientations. We grouped trials according to whether the behavioral reports had been relatively more or less biased toward the previous trial (Figure 4d), and we separately assessed the IEM error distributions for those grouped trials. An error distribution centered around 0 would reflect an unbiased representation, one shifted to the right would reflect a bias toward the previous trial information, and one shifted to the left would reflect a bias away from the previous trial information. For both gaze and hand data, trials that showed a less attractive behavioral bias also showed stimulus reconstructions that were more strongly shifted away from previous trial reports (Figure 4e). Conversely, trials that showed a more attractive behavioral bias also showed stimulus reconstructions that were more strongly shifted toward previous trial reports. Again, this relationship was only evident in the *wheel* condition for gaze data (gaze (*wheel*): $\boldsymbol{W = 123, n = 36, p < 0.001}$) and in the *draw* condition for hand data (hand (*draw*): $\boldsymbol{W = 206, n = 36, p = 0.046}$). Therefore, temporal context biases in

peripheral signals do appear to track with behavioral outcomes, but only for the task conditions that are preferred by each peripheral motor effector (see Methods & Stats for non-preferred condition reconstructions).

In summary, during WM maintenance, stimulus information in eye and hand-movement signatures exhibited behavior-related serial dependence in 'preferred' task conditions. Therefore, signals in the peripheral motor system that are sensitive to prospective task demands may also reflect the integration of perceptual and decisional information over time. Both temporal and response context information may be adaptively integrated into distributed WM representations to best guide future responding.

CONCLUSIONS

Here, we observed a double dissociation between the WM stimulus information in gaze vs. hand movement patterns, during *draw* vs. *wheel* WM response conditions. Given identical visual stimulus content, the distribution of WM maintenance activity across manual and oculomotor effectors is tailored to subtle differences in impending action demand. These findings could stem from an emergent function where some of the WM maintenance onus is offloaded to the peripheral nervous system and balanced across relevant circuits in an adaptive way[53]. Beyond simple motor preparation or residual perceptual traces, these peripheral signals may aggregate contextual information across multiple scales, including both upcoming task demands and recent behavioral history. These findings echo work showing that cortical WM signals can flexibly reflect various motor and task factors[6,8,9,26,54]. Now, the current work shows that WM may similarly recruit the peripheral nervous system in a widely distributed scheme that is exquisitely sensitive to multiple levels of task context. These findings open up avenues of inquiry into individual variability in relying on this WM maintenance format, and whether these peripheral signatures play a functional role in behavior. For instance, future work should address whether this activity shoulders some of the burden for cortical storage mechanisms, reciprocally fortifies cortical signals (cf.[55]), or instead expresses the behavioral output of cortical representations.

STAR METHODS

Subjects

A total of 45 undergraduate students at the University of California, San Diego (UCSD) were recruited through the SONA participant pool. Participants received course credit

as compensation. The study was approved by the UCSD Institutional Review Board (IRB), and all participants provided informed consent electronically before participating.

Among them, 37 successfully completed the experiment without interruption (age: 18 to 26, M=20.60, SD=1.91; 29 female). One participant was left-handed. All participants had normal or corrected-to-normal vision (with 12 using contact lenses), and reported no history of neurological or psychiatric disorders. One additional participant was excluded from analyses due to failing the eye-data quality control criterion in more than 40% trials, resulting in a final sample of 36 participants.

Materials and Procedure

Apparatus

Participants were seated in a dimly lit room, approximately 60 cm away from a 24-inch monitor with a resolution of 1920 × 1080 pixels. Gaze data were recorded using a desk-mounted EyeLink (Portable Duo) eye tracker (SR Research Ltd), with a sampling rate of 1000 Hz. Data were exported using the EyeLink Data Viewer software package (v.4.3.1).

Responses were made on a Wacom tablet (Intuos CTH-690), which captures stylus movements with a resolution of 2540 lines per inch (LPI) and has an active area of size 8.5 x 5.3 in. The stylus functioned as a direct input device, with the tip position on the tablet mapped to the cursor position on the screen, effectively acting as a mouse. The cursor was invisible during the delay. The cursor's positions were recorded using PsychoPy. Participants were instructed to hold the stylus throughout the experiment, but they were otherwise unaware that stylus movements aside from explicit task responses were being recorded.

A chin rest was used to minimize head movements during the experiment and to maintain a consistent viewing distance. The tablet was placed in front of the monitor, on the side of the participant's dominant hand. Participants were allowed to adjust the tablet's position for comfort and ease of use during the calibration phase, but were instructed not to move it during the experiment.

The experiment was programmed and presented using Psychopy (v2024.2.4[56]) on a Windows 10 system.

Procedure and tasks

Each trial began with a 1000ms fixation period, followed by two sequentially-presented Gabor patches (referred to as 'stim1' and 'stim2' respectively) on a neutral grey

background. The orientation of each patch was independently sampled from a uniform distribution ranging from 0° to 180°. Each Gabor patch had a Gaussian envelope with a radius of 0.16 screen-heights (≈4.5° visual angle). The spatial frequency was set to 5 cycles per screen height (≈0.18 cycles/deg, yielding ~1.6 cycles across the patch). A cue appeared 250 ms after each Gabor patch onset, in the form of a star at the center of the screen, overlaid atop the Gabor, for 500ms. The color of the star tagged the stimulus so that it could be cued later for recall, and the filling of the star indicated whether that stimulus should be remembered (filled star) or could be dropped (hollow star). The color-cue mapping (i.e., which color corresponded to stim1 and stim2) was randomized across trials.

The cue and the Gabor patch offset simultaneously and were each followed by a noise mask. The mask consisted of random black (RGB: 0,0,0) and dark-gray (RGB: 187,187,187) squares, with a mean luminance of ~93 (on 0–255 scale) and a spatial frequency of ≈5.6 cycles/deg, presented at the same size as the Gabor patch. The first noise mask (after stim1) was followed by a 1s blank delay, during which only the neutral gray background was visible. We refer to this period as the inter-stimuli interval (ISI). The second noise mask (after stim2) was followed by a longer blank delay of 5s. This is the WM delay period on which IEM analyses are focused.

During the response phase, participants reported their memory of the cued orientation(s) using a stylus on a tablet. In the *draw* condition, participants were instructed to draw a single line to best match the angle of the remembered orientation. A white circular input region appeared, and participants could draw anywhere within the region. In the *wheel* condition, participants adjusted the positions of two opposing dots on a response wheel, so that the imaginary line formed by the dots matched the remembered orientation. A bare response wheel appeared initially, and once the participant clicked on the wheel, one dot appeared at the clicked location and the second dot appeared at the opposite position. The response phase was self-paced, and participants were able to erase and redo their drawings, or continuously adjust their wheel response, until they were satisfied and ready to press a ‘continue’ button to move on.

To prevent anticipatory gaze or hand motion biases associated with the expected locations of upcoming actions, the positions of the response regions were randomized across trials. The two response regions could be arranged either horizontally (to the left and right of the screen center) or vertically (above or below the screen center), with equal probability. The two stimuli were also assigned randomly to the two response regions (left vs. right, up vs. down), with equal probability. During the response phase, two response regions were always visible, but only an illuminated region was editable to receive a response. Such editable regions were also marked by a central colored dot (green or purple), indicating which stimulus should be reported within. On trials where

one WM item was probed, only one response region would be illuminated, whereas both would be illuminated on trials where both items were probed.

The task also included some conditions that are not analyzed here. Namely, the experiment included two types of blocks: *explicit* and *neutral*. In *explicit* blocks, when both items were cued to remember, they were both probed for recall. In *neutral* blocks, while both items were always cued, there was a 50% chance that only one of them would be probed. Therefore, across all conditions, participants may have been cued to remember either one or two items, and sometimes they only had to recall one item even if they were cued to remember two. However, the cues were always "valid" in that participants were never tested on an item that was uncued (see Table of trial/block conditions)**.**

Block	Trials
Explicit: During encoding, participants were certain that all and only the cued item(s) would be probed later.	• 50%: Remember one, recall one • 50%: Remember both, recall both
Neutral: Participants were cued to remember both items, but were uncertain whether one or both of the cued items would be probed later.	• 50%: Remember both, recall one • 50%: Remember both, recall both

The experiment consisted of 16 blocks: 8 each in the *draw* and *wheel* conditions. Each block contained 10 trials, yielding 160 trials in total (80 per response format condition). The experiment was conducted in two batches separated by a 10-minute break. Each batch consisted of 8 blocks, which were either all *explicit* or all *neutral* probe conditions, with the order randomized across participants. Within each batch, the first or second set of 4 blocks was randomly assigned to the *draw* or *wheel* condition. Thus, participants completed *draw* and *wheel* conditions in alternating sets of 4, which was aimed to minimize context-switching demands. Before each batch of the experiment, participants reviewed a diagram outlining the design of the upcoming blocks (see Figure S1), and completed 12 practice trials to familiarize themselves with the task.

Quantification and Data Analysis

Data recording and pre-processing

Behavioral response data

Although participants were allowed to redo their response in both response formats, our analyses included only the final response submitted. In the *wheel* condition, the

response was defined as the orientation of the line connecting the two dots. In the *draw* condition, the response was defined as the average orientation of the drawn line, computed after excluding the portions drawn during the first and last 10% of the drawing time.

Gaze data

Gaze data were recorded at a sampling rate of 1000 Hz. The eye tracker was calibrated before each batch of the experiment, with drift correction performed before each block. On each trial, gaze data were recorded from onset of the fixation period until 500 ms after the start of the response phase. All participants met the behavioral QA criteria that average response error was <30 deg, and at least 3/4 of trials were retained after outlier exclusion.

Preprocessing and quality control of gaze data closely followed published procedures29. Blinks were identified as time points where the rate of pupil-size change was an outlier (i.e., greater than 4 median absolute deviations). Gaze data from these time points and the immediately neighboring epochs was removed, and the resulting gaps were interpolated. Trials were excluded if more than 10% of data points had been removed, and participants were excluded from further analysis if they had more than 40% of trials excluded.

To characterize gaze patterns, we extracted features as follows. For each time window of 50 ms, we calculated the median x and y gaze coordinates across all trials for each participant. Gaze data were then re-centered based on this participant- and time-window-specific median locations. After re-centering, we included only gaze data falling between 15 and 150 pixels from the normalized center for further analysis, as very small displacements (<15 px) in our dataset were predominantly vertical and likely reflected noise, while large displacements (>150 px) were more likely due to distraction or off-task behavior. Next, for each time window within each trial, we extracted two types of statistics: (1) the distribution of gaze offset angles, and (2) the spatial distribution of gaze as a two-dimensional heatmap with a resolution of 50 pixels. These two vectors were concatenated to form a single feature vector describing gaze patterns. For each participant, a baseline feature vector was then computed as the average across all trials from phases with meaningful visual inputs or WM task (i.e., defined as the interval from the onset of the first stimulus to the start of the response phase). This baseline was subtracted from all feature vectors. Finally, we applied z-normalization to the resulting feature vectors.

Hand motion data

Hand motion data were collected using the stylus connected to the tablet. Position of the stylus on the tablet is captured by the cursor position on the screen (invisible during blank delay), recorded through Psychopy.

Participants were not aware of the recording of hand motion data, except during the response phase. On each trial, the hand motion data were recorded from the fixation period until the completion of responses.

Motions lasting less than 50 ms (i.e., when continuous position changes ceased within 50 ms) were considered sudden jumps and removed as artifacts. The resulting hand motion data were represented as a sequence of motion vectors, each defined by its direction, magnitude, and start and end times. To further smooth the data, consecutive motion segments with angular differences of less than 15° were merged into a single motion. The direction of the merged motion was computed as the weighted average of the constituent segment motions, and its magnitude was computed as the sum of their individual magnitudes.

To characterize hand motions, we extracted features as follows. For every 50ms time window within a trial, we computed the magnitude of motions within each of 30 directional bins spanning 360° (i.e., 12° per bin), resulting in a 30-dimensional vector. (For motions that spanned multiple time windows, the motion magnitude assigned to each window was proportional to the fraction of the motion's duration falling within that window). Finally, to address individual and temporal biases, we applied z-normalization to the feature vector across all trials for each participant and time window.

Data Analysis

Statistical analysis related to Figure 1d-f – behavioral performance

Average errors

For each condition, we computed the average deviation of each participant's responses from the true stimulus orientations. Trials with response errors greater than 45° were excluded, as they might reflect factors unrelated to memory or response precision.

Behavior: error

Condition	M ± SEM	Mdn (Q1, Q3)
Draw	12.51 ± 0.50	12.77 (10.33, 14.46)
Wheel	11.59 ± 0.43	12.02 (10.10, 13.65)

Both paired samples t-test and Wilcoxon signed-rank test suggest that errors are significantly greater in the *draw* condition ($t(34) = 2.95, p = 0.006$; $W = 171, n = 36, p = 0.010$)

Reaction time

For each trial, we computed two timing indices for behavioral responses: initiation time (the interval between the onset of the recall phase and the start of the first response) and confirmation time (the interval between completing the final response and pressing the "continue" button). We then computed the mean initiation and confirmation time for each participant across their trials. The table below summarizes the results (in seconds):

Behavior: response time

	Condition	M ± SEM	Mdn (Q1, Q3)
Initiation	Draw	2.32 ± 0.08	2.37 (2.06, 2.61)
	Wheel	2.14 ± 0.09	2.11 (1.83, 2.40)
Confirmation	Draw	1.15 ± 0.04	1.15 (0.92, 1.35)
	Wheel	1.25 ± 0.05	1.22 (1.03, 1.51)

Both the paired samples t-test and Wilcoxon signed-rank test suggest that compared to the *wheel* condition, in the *draw* condition, people start responding more slowly ($t(34) = 2.39, p = 0.022$; $W = 148, n = 36, p = 0.003$) but press the "Continue" button more quickly ($t(34) = 2.98, p = 0.005$; $W = 164, n = 36, p = 0.007$).

Precision (sharpness) metrics

To quantify memory precision, in addition to calculating the average error, we computed the error distribution for each participant, centered around $\text{error} = 0$. To address data sparsity, the distribution was smoothed using a weight defined as:

$$K(x_1, x_2) \propto \cos\big(\angle(x_1, x_2)\big) \cdot k$$

where k was set to 150. Performance was then defined by the sharpness of the smoothed distribution around zero, quantified as:

$$\text{Performance} = \cos(x) \cdot \hat{P}(\text{err} = x)$$

A greater value indicates that errors are more concentrated around zero, suggesting better performance.

Behavior: precision (sharpness)

Condition	M ± SEM	Mdn (Q1, Q3)
Draw	5.97 ± 0.08	6.06 (5.81, 6.30)
Wheel	5.92 ± 0.07	5.99 (5.79, 6.18)

Paired samples t-test suggests no significant difference in error between conditions ($t(34) = 0.89, p = 0.378$). A Wilcoxon signed-rank test also suggests that the error is not significantly greater in the *draw* condition, with $W = 278, n = 36, p = 0.396$.

Statistical analysis related to Figure 2c and Figure 3a – representational similarity analysis (RSA)

RSA details

We follow a similar approach to that described[12]. To quantify similarity between the peripheral motor activity patterns, we use cosine similarity between pattern feature vectors. To quantify similarity between stimuli, accounting for perceptual inhomogeneities, we use the cosine similarity between their corresponding response distributions from an independent behavioral sample.

We calculated gaze or hand motion pattern similarity across trials, stimulus similarity across trials, and the correlation between the two. Specifically, for each participant and each 50 ms time window, we considered all pairs of trials i and j ($i \neq j$), and we constructed two matrices. For each pair, we computed (1) the cosine similarity between the participants' motor activity feature vectors within that time window (producing the motor activity matrix), and (2) the cosine similarity between the response distributions from the independent behavioral sample (producing the stimulus similarity matrix). We then calculated the Pearson correlation between the motor activity matrix and the stimulus similarity matrix. Significance was assessed using a two-tailed permutation test (iterations=1000), evaluating whether the observed correlation differed significantly from zero (i.e., no correlation).

Statistical analysis related to Figure 2d-e, Figure 3b-d – inverted encoding modeling (IEM)

Reconstruction of memorized features through IEM and quantification of evidence

Inverted encoding model (IEM) has often been used to reconstruct perceived or remembered stimuli from brain activity patterns[32,33]. Here, we adapted this approach to reconstruct the memorized features from gaze and hand motion patterns.

We used a model with 18 channels, each assumed to be best tuned to a different orientation: 0°, 10°, …, 170°. Let the peripheral signal features be denoted by X, the channel activations by C, and the channels' basis function as W. Assuming

$$X = WC$$

We estimate W using the peripheral signal patterns X_1 and the estimated channel activations C_1 from the training data:

$$\widehat{W} := X_1 C_1^T (C_1^T C_1)^{-1}$$

Then, for a separate set of test trials with signal patterns X_2, we can estimate the corresponding channel activations as:

$$\widehat{C_2} := \left(\widehat{W}^T \widehat{W}\right)^{-1} \widehat{W}^T X_2$$

Let C_z denote the activation of the channel tuned to orientation z. We first define the similarity between two angles (i.e., y_1 and y_2, whose angle between is $\angle(y_1, y_2)$) as:

$$s(y_1, y_2) \propto \cos\left(\angle(y_1, y_2)\right)$$

For a given stimulus y, the activation of each channel can be defined as:

We denote the projection from stimulus space Y to channel activation space C as $\Pi_{Y \to C}$. Thus, based on the predicted channel activations $\hat{C}$, the predicted stimuli $\hat{Y}$ could be obtained by:

The only free parameters in this model are the sharpness parameter k, which determines the tuning width of channel activations. To find the optimal k, we compute the reconstruction loss for all k as:

$$\mathcal{L} := \| \widehat{W}\hat{C} - X \|_2$$

With the optimal sharpness k:

$$k^* := \underset{k}{\operatorname{argmin}} \mathcal{L}$$

For each trial, the model predicts a distribution, $\hat{P}(s = x)$ representing the predicted probability that the target stimulus s has value x. To evaluate model performance, we first compute the distribution of reconstruction errors across all trials for each participant. We then quantify the quality of the reconstruction as follows:

$$\text{Evidence} = cos(x) \cdot \hat{P}(err = x)$$

A smaller reconstruction error should correspond to a sharper reconstruction error distribution curve around zero, thereby leading to higher evidence. A uniform distribution of reconstruction errors yields an evidence value of zero. In contrast, a model that systematically tracks the orientation orthogonal to the target is expected to yield a negative evidence value.

For across-subject analyses, we used a leave-one-subject-out cross-validation scheme: the training set included all trials from all participants except one, and the test set comprised all trials from the held-out participant. For within-subject analyses, we applied a 5-fold cross-validation scheme, with each fold containing 20% of trials from each participant. When training and testing across different experimental phases or trial conditions, we first split the test data at the subject- (across-subject) or trial- (within-subject) level, and then for each test-set selected the training data to ensure no overlap between train and test sets in terms of participants (across-subject) or trials (within-subject).

Statistical analysis for Figure 2d-e: IEM evidence throughout trial

The following are the IEM-based evidence traces of the first and second stimuli in gaze and hand motion patterns throughout the trial. Evidence is calculated per subject. For the post-stim1-encoding phase and ISI, trials where stim1 is cued to drop are excluded. For the post-stim2-encoding and delay phases, to calculate the evidence for stim1 and stim2, trials are only included when the corresponding item is cued to remember . Additionally, we calculated 'combined' evidence, which is extracted from the average error distribution from decoding stim1 and stim2.

The following table summarizes the IEM evidence statistics. We report both parametric tests (t-test) and non-parametric tests (Wilcoxon signed rank test) results, because in some conditions the data did not pass the normality check (i.e., p(normal)<0.05)

IEM evidence in gaze data

	Target	M $\pm$ SEM	$t(35)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n=36)$ (p)
Post-stim1	stim1	1.43 $\pm$ 0.19	7.43 (<.001)	0.245	1.22 (0.63, 2.10)	650 (<.001)
	stim2	0.07 $\pm$ 0.07	0.90 (0.374)	0.934	0.06 (-0.22, 0.38)	399 (0.154)
Post-stim2	stim1	0.39 $\pm$ 0.08	5.15 (<.001)	0.021	0.30 (0.11, 0.60)	609 (<.001)
	stim2	0.73 $\pm$ 0.12	6.36 (<.001)	<.001	0.50 (0.24, 0.91)	662 (<.001)
	combined	0.56 $\pm$ 0.08	7.15 (<.001)	0.012	0.43 (0.23, 0.83)	661 (<.001)
Delay	stim1	0.70 $\pm$ 0.08	8.35 (<.001)	0.804	0.75 (0.40, 1.07)	646 (<.001)

	stim2	0.50 ± 0.08	6.19 (<.001)	0.188	0.42 (0.16, 0.78)	631 (<.001)
	combined	0.61 ± 0.07	8.50 (<.001)	0.454	0.61 (0.27, 0.82)	659 (<.001)

IEM evidence in hand motion data

	Target	M ± SEM	$t(35)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n = 36)$ (p)
Post-stim1	stim1	1.33 ± 0.27	4.90 (<.001)	0.003	0.87 (0.02, 2.16)	590 (<.001)
	stim2	0.05 ± 0.05	1.00 (0.323)	0.611	0.03 (-0.14, 0.25)	385 (0.212)
Post-stim2	stim1	0.20 ± 0.06	3.21 (0.003)	<.001	0.18 (-0.04, 0.28)	546 (<.001)
	stim2	0.49 ± 0.10	4.84 (<.001)	<.001	0.26 (0.12, 0.75)	630 (<.001)
	combined	0.35 ± 0.06	5.99 (<.001)	0.004	0.26 (0.52, 0.65)	645 (<.001)
Delay	stim1	0.33 ± 0.09	3.78 (0.001)	<.001	0.14 (-0.01, 0.62)	555 (<.001)
	stim2	0.26 ± 0.06	4.29 (<.001)	0.052	0.21 (-0.01, 0.49)	557 (<.001)
	combined	0.30 ± 0.07	4.48 (<.001)	0.003	0.15 (-0.01, 0.55)	562 (<.001)

Statistical analysis for Figure 3b-d: IEM evidence compared across modality

The following tables show the evidence of memorized features in gaze and hand motion, comparing the *draw* and *wheel* response formats.

IEM evidence – comparing draw and wheel

	Condition	M ± SEM	Mdn (Q1, Q3)
Gaze	Draw	0.48 ± 0.08	0.37 (0.08, 0.88)
	Wheel	0.73 ± 0.10	0.66 (0.31, 1.10)
Hand	Draw	0.38 ± 0.08	0.20 (0.06, 0.64)
	Wheel	0.23 ± 0.06	0.11 (0.01, 0.38)

Wilcoxon signed-rank tests revealed significant differences in the strength of evidence between the two response format conditions, in both gaze-based and hand-motion-

based decoding. The table below summarizes the comparison between response formats (i.e., *draw – wheel*) for gaze and hand motion respectively.

IEM evidence – Draw - Wheel

	M ± SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n = 36)$ (p)
Gaze	-0.25 ± 0.11	-2.22 (0.033)	0.394	-0.36 (-0.80, 0.26)	202 (0.039)
Hand	+0.15 ± 0.05	+3.06 (0.004)	0.008	+0.14 (-0.04, 0.28)	157 (0.005)

Due to noise, some conditions produced negative evidence values for certain participants. These negative values are potentially problematic for interpretation; for example, if a participant shows evidence of -1 in condition A and -3 in condition B, both are below chance, and it is not meaningful to claim better performance in A than in B. Therefore, we repeated the analysis after converting all negative evidence values to 0, but this did not change the findings.

IEM evidence – comparing draw and wheel, negative values corrected

	Condition	M ± SEM	Mdn (Q1, Q3)
Gaze	Draw	0.49 ± 0.08	0.37 (0.08, 0.88)
	Wheel	0.74 ± 0.09	0.66 (0.31, 1.10)
Hand	Draw	0.40 ± 0.07	0.20 (0.06, 0.64)
	Wheel	0.25 ± 0.05	0.11 (0.01, 0.38)

IEM evidence – Draw - Wheel, negative values corrected

	M ± SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n = 36)$ (p)
Gaze	-0.25 ± 0.11	-2.31 (0.027)	0.397	-0.35 (-0.79, 0.13)	182 (0.029)
Hand	+0.15 ± 0.05	+3.04 (0.004)	0.001	+0.09 (-0.02, 0.25)	132 (0.005)

Statistical analysis related to Figure 4a-c, d-e: evaluate serial bias in IEM-based decoding

To quantify the impact of preceding trials on the pattern of current trial peripheral signals, we extract the following metric from the IEM-based decoding results using a method modified from[40]. For each trial, the IEM decoding error distribution is aligned to

the direction pointing from the current target to the previous response: if the last response in the previous trial is counter clockwise relative to the current target stimulus, the error distribution is flipped at zero. For a given trial condition, we then aggregate the error distributions across trials for each participant. The bias of the aggregated distribution is defined as follows**:**

$$\text{Bias} = sin(x) \cdot \hat{P}(err = x)$$

A positive bias indicates that the pattern is shifted toward the previous response, suggesting an attractive serial effect, while a negative bias indicates a shift away from the previous response, suggesting a repulsive serial effect. The bias is computed per subject using leave-one-subject-out cross-validation approach.

Previous studies on serial dependence in behavior have shown that the direction of the bias can differ depending on the similarity between the current and previous stimuli—specifically, attraction when successive stimuli are similar and repulsion when they are dissimilar. We tested whether such a pattern would emerge for WM representations in peripheral motor activity as well. To do this, we grouped trials based on the angular difference between the current stimulus and the previous response (referred to as 'serial difference') into six bins (15° per bin). For gaze and hand data, shifts were calculated from IEM-based reconstructions, whereas for behavioral data, shifts were calculated from the distribution of errors in responses.

Thus, for each participant, we obtained three functions describing bias as a function of serial difference: one for behavior, one for gaze, and one for hand motion. First, the behavioral bias shows a pattern consistent with previous studies: the bias is significantly positive (i.e. attractive toward the previous response) when the current stimulus and previous response are similar (0–45° difference), but negative (i.e., repulsive) when they are dissimilar (45–90° difference). The following table summarizes the statistical test of whether the behavioral bias significantly differs from 0 (i.e., unbiased).

Behavioral serial bias (Draw)

serial-diff	M ± SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n = 36)$ (p)
$[0 - 15)$	0.07 ± 0.02	3.92 (<.001)	0.993	0.08 (0.00, 0.14)	121 (0.001)
$[15 - 30)$	0.18 ± 0.02	10.18 (<0.001)	0.244	0.18 (0.08, 0.28)	2 (<.001)
$[30 - 45)$	0.03 ± 0.02	1.70 (0.098)	0.628	0.05 (-0.04, 0.09)	228 (0.101)
$[45 - 60)$	-0.04 ± 0.02	-2.08 (0.045)	0.657	-0.04 (-0.09, 0.02)	199 (0.035)
$[60 - 75)$	-0.10 ± 0.02	-4.08 (<.001)	0.657	-0.10 (-0.17, -0.02)	112 (<.001)

$[75-90)$	-0.07 $\pm$ 0.02	-3.68 (<.001)	0.951	-0.08 (-0.14, 0.00)	129 (0.001)

Behavioral serial bias (Wheel)

serial-diff	M $\pm$ SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n=36)$ (p)
$[0-15)$	0.09 $\pm$ 0.02	4.84 (<.001)	0.902	0.09 (0.01, 0.18)	85 (<.001)
$[15-30)$	0.14 $\pm$ 0.02	9.23 (<0.001)	0.634	0.14 (0.09, 0.20)	13 (<.001)
$[30-45)$	0.09 $\pm$ 0.02	3.61 (0.001)	0.589	0.09 (0.01, 0.17)	127 (0.001)
$[45-60)$	0.00 $\pm$ 0.02	-0.14 (0.887)	0.344	0.02 (-0.09, 0.10)	333 (1.000)
$[60-75)$	-0.07 $\pm$ 0.02	-4.03 (<.001)	0.001	-0.04 (-0.12, 0.01)	118 (<.001)
$[75-90)$	-0.06 $\pm$ 0.02	-3.43 (0.002)	0.147	-0.05 (-0.10, -0.02)	123 (0.001)

We then computed, for each participant, the correlation between their behavioral bias function and their gaze and hand motion bias functions, respectively.

The following table summarizes the results of this analysis (with $n = 36$):

Correlation between peripheral signals and behavior serial bias patterns

	Condition	M $\pm$ SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n=36)$ (p)
Gaze	draw	0.09 $\pm$ 0.08	1.06 (0.148)	0.220	0.09 (-0.23, 0.50)	403 (0.139)
	wheel	0.22 $\pm$ 0.08	2.85 (0.004)	0.566	0.24 (-0.08, 0.57)	498 (0.004)
Hand	draw	0.14 $\pm$ 0.08	1.83 (0.038)	0.065	0.27 (-0.12, 0.48)	446 (0.039)
	wheel	-0.01 $\pm$ 0.07	-0.16 (0.562)	0.475	0.00 (-0.26, 0.32)	324 (0.559)

The results indicate that serial bias in gaze significantly aligns with behavioral bias only in the *wheel* condition, while hand motions only significantly aligns with behavior in the *draw* condition.

We further tested whether the bias in peripheral signals differed significantly between trials where the behavioral bias was more versus less attracted toward the previous trial. We compared the IEM reconstructions between trials of more positive and less positive behavioral bias (i.e., group of more attractive behavioral bias - group of less attractive behavioral bias). The following table summarizes the results:

Behavior-relevant serial bias in peripheral signals

	Condition	M $\pm$ SEM	$t(34)$ (p)	p(normal)	Mdn (Q1, Q3)	$W(n = 36)$ (p)
Gaze	draw	2.04 $\pm$ 1.32	1.55 (0.130)	0.541	3.01 (-3.47, 5.86)	240 (0.148)
	wheel	5.97 $\pm$ 1.52	3.92 (<.001)	0.659	5.55 (0.13, 11.57)	123 (0.001)
Hand	draw	5.32 $\pm$ 2.99	1.78 (0.083)	0.171	4.95 (3.19, 13.84)	206 (0.046)
	wheel	3.40 $\pm$ 2.31	1.47 (0.149)	0.295	3.12 (-2.99, 13.72)	215 (0.064)

ACKNOWLEDGEMENT

We thank Connie Xie, Yun-Chen Hung and Lana Gaspariani for their help with data collection. We also thank Pria Daniel, Ana Chkhaidze and Zhuojun Ying for helpful discussions and feedback throughout this study. This work was funded in part by NIH award 1R01EY036843-01 to Anastasia Kiyonaga.

REFERENCES

1. D'Esposito, M., and Postle, B.R. (2015). The cognitive neuroscience of working memory. Annual review of psychology *66*, 115–142.

2. Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.-D. (2017). The distributed nature of working memory. Trends in cognitive sciences *21*, 111–124.

3. Myers, N.E., Stokes, M.G., and Nobre, A.C. (2017). Prioritizing information during working memory: Beyond sustained internal attention. Trends in cognitive sciences *21*, 449–461.

4. Nobre, A.C., and Stokes, M.G. (2019). Premembering experience: A hierarchy of time-scales for proactive attention. Neuron *104*, 132–146.

5. Serences, J.T., Ester, E.F., Vogel, E.K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. Psychological science *20*, 207–214.

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

7. Pratte, M.S., and Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. Journal of vision *14*, 22–22.

8. Boettcher, S.E., Gresch, D., Nobre, A.C., and Ede, F. van (2021). Output planning at the input stage in visual working memory. Science Advances *7*, eabe8212.

9. Henderson, M.M., Rademaker, R.L., and Serences, J.T. (2022). Flexible utilization of spatial-and motor-based codes for the storage of visuo-spatial information. Elife *11*, e75688.

10. Shao, Z., Zhang, M., and Yu, Q. (2025). Stimulus representation in human frontal cortex supports flexible control in working memory. eLife *13*, RP100287.

11. Van Ede, F., Chekroud, S.R., and Nobre, A.C. (2019). Human gaze tracks attentional focusing in memorized visual space. Nature human behaviour *3*, 462–470.

12. Linde-Domingo, J., and Spitzer, B. (2024). Geometry of visuospatial working memory information in miniature gaze patterns. Nature Human Behaviour *8*, 336–348.

13. Sheehan, T., Carfano, B., Richmond, D., and Serences, J. (2024). Serial dependence persists longer for imaginary versus real stimuli.

14. Trentin, C., Falanga, L., Jeske, J., Olivers, C.N., and Slagter, H.A. (2024). Action similarity warps visual feature space in working memory. Proceedings of the National Academy of Sciences *121*, e2413433121.

15. Prinzmetal, W., Amiri, H., Allen, K., and Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. Journal of Experimental Psychology: Human Perception and Performance *24*, 261.

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

17. Van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W.J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. Proceedings of the National Academy of Sciences *109*, 8780–8785.

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

19. Bainbridge, W.A., Pounder, Z., Eardley, A.F., and Baker, C.I. (2021). Quantifying aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. Cortex *135*, 159–172.

20. Fan, J.E., Bainbridge, W.A., Chamberlain, R., and Wammes, J.D. (2023). Drawing as a versatile cognitive tool. Nature Reviews Psychology *2*, 556–568.

21. Wixted, J.T., and Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. Psychonomic Bulletin & Review *1*, 89–106.

22. Tulving, E., and Osler, S. (1968). Effectiveness of retrieval cues in memory for words. Journal of experimental psychology *77*, 593.

23. Olivers, C.N., and Roelfsema, P.R. (2020). Attention for action in visual working memory. Cortex *131*, 179–194.

24. Ede, F. van (2020). Visual working memory and action: Functional links and bi-directional influences. Visual Cognition *28*, 401–413.

25. Jonikaitis, D., Noudoost, B., and Moore, T. (2023). Dissociating the contributions of frontal eye field activity to spatial working memory and motor preparation. Journal of Neuroscience *43*, 8681–8689.

26. McDougle, S.D., and Hillman, H. (2025). Motor working memory. Trends in Cognitive Sciences. https://doi.org/https://doi.org/10.1016/j.tics.2025.08.011.

27. Hustá, C., Dalmaijer, E., Belopolsky, A., and Mathôt, S. (2019). The pupillary light response reflects visual working memory content. Journal of Experimental Psychology: Human Perception and Performance *45*, 1522.

28. Zokaei, N., Board, A.G., Manohar, S.G., and Nobre, A.C. (2019). Modulation of the pupillary response by the content of visual working memory. Proceedings of the National Academy of Sciences *116*, 22802–22810.

29. Dong, Y., and Kiyonaga, A. (2024). Ocular working memory signals are flexible to behavioral priority and subjective imagery strength. Journal of Neurophysiology *132*, 162–176.

30. Dong, Y., Hung, Y.-C., Xie, C., and Kiyonaga, A. (2025). Pupillary signatures of working memory content are flexible to visual vs. Semantic task demands. Journal of Vision *25*, 2871–2871.

31. Kriegeskorte, N., Mur, M., and Bandettini, P.A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience *2*, 249.

32. Brouwer, G.J., and Heeger, D.J. (2009). Decoding and reconstructing color from responses in human visual cortex. Journal of Neuroscience *29*, 13992–14003.

33. Sprague, T.C., and Serences, J.T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nature neuroscience *16*, 1879–1887.

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

35. Foster, J.J., Sutterer, D.W., Serences, J.T., Vogel, E.K., and Awh, E. (2017). Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. Psychological science *28*, 929–941.

36. Lorenc, E.S., Sreenivasan, K.K., Nee, D.E., Vandenbroucke, A.R., and D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. Journal of Neuroscience *38*, 5267–5276.

37. Sprague, T.C., Adam, K.C., Foster, J.J., Rahmati, M., Sutterer, D.W., and Vo, V.A. (2018). Inverted encoding models assay population-level stimulus representations, not single-unit neural tuning. Eneuro *5*.

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

39. Barbosa, J., Stein, H., Martinez, R.L., Galan-Gadea, A., Li, S., Dalmau, J., Adam, K.C., Valls-Solé, J., Constantinidis, C., and Compte, A. (2020). Interplay between persistent activity and activity-silent dynamics in the prefrontal cortex underlies serial biases in working memory. Nature neuroscience *23*, 1016–1024.

40. Fischer, C., Kaiser, J., and Bledowski, C. (2025). A direct neural signature of serial dependence in working memory. Elife *13*, RP99478.

41. Fischer, J., and Whitney, D. (2014). Serial dependence in visual perception. Nature neuroscience *17*, 738–743.

42. Sheehan, T.C., and Serences, J.T. (2022). Attractive serial dependence overcomes repulsive neuronal adaptation. PLoS biology *20*, e3001711.

43. Luo, M., Zhang, H., Fang, F., and Luo, H. (2025). Reactivation of previous decisions repulsively biases sensory encoding but attractively biases decision-making. PLoS biology *23*, e3003150.

44. Manassi, M., Murai, Y., and Whitney, D. (2023). Serial dependence in visual perception: A meta-analysis and review. Journal of vision *23*, 18–18.

45. Pascucci, D., Tanrikulu, Ö.D., Ozkirli, A., Houborg, C., Ceylan, G., Zerr, P., Rafiei, M., and Kristjánsson, Á. (2023). Serial dependence in visual perception: A review. Journal of Vision *23*, 9–9.

46. Cicchini, G.M., Mikellidou, K., and Burr, D.C. (2024). Serial dependence in perception. Annual Review of Psychology *75*, 129–154.

47. Kiyonaga, A., Scimeca, J.M., Bliss, D.P., and Whitney, D. (2017). Serial dependence across perception, attention, and memory. Trends in Cognitive Sciences *21*, 493–497.

48. Cicchini, G.M., Mikellidou, K., and Burr, D.C. (2018). The functional role of serial dependence. Proceedings of the Royal Society B *285*, 20181722.

49. Manassi, M., and Whitney, D. (2022). Illusion of visual stability through active perceptual serial dependence. Science advances *8*, eabk2480.

50. Bliss, D.P., Sun, J.J., and D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. Scientific reports *7*, 14739.

51. Fritsche, M., Mostert, P., and Lange, F.P. de (2017). Opposite effects of recent history on perception and decision. Current Biology *27*, 590–595.

52. Samaha, J., Switzky, M., and Postle, B.R. (2019). Confidence boosts serial dependence in orientation estimation. Journal of vision *19*, 25–25.

53. Postle, B.R. (2006). Working memory as an emergent property of the mind and brain. Neuroscience *139*, 23–38.

54. Curtis, C.E., Rao, V.Y., and D'Esposito, M. (2004). Maintenance of spatial and motor codes during oculomotor delayed response tasks. Journal of Neuroscience *24*, 3944–3952.

55. Goldin-Meadow, S. (2005). Hearing gesture: How our hands help us think (Harvard University Press).

56. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behavior research methods *51*, 195–203.

SUPPLEMENTARY

Behavioral data: how precise is drawing

We also report performance for the *draw* and *wheel* conditions under different WM load conditions. We separate trials into those where only one vs. both items were cued to remember (Figure S1).

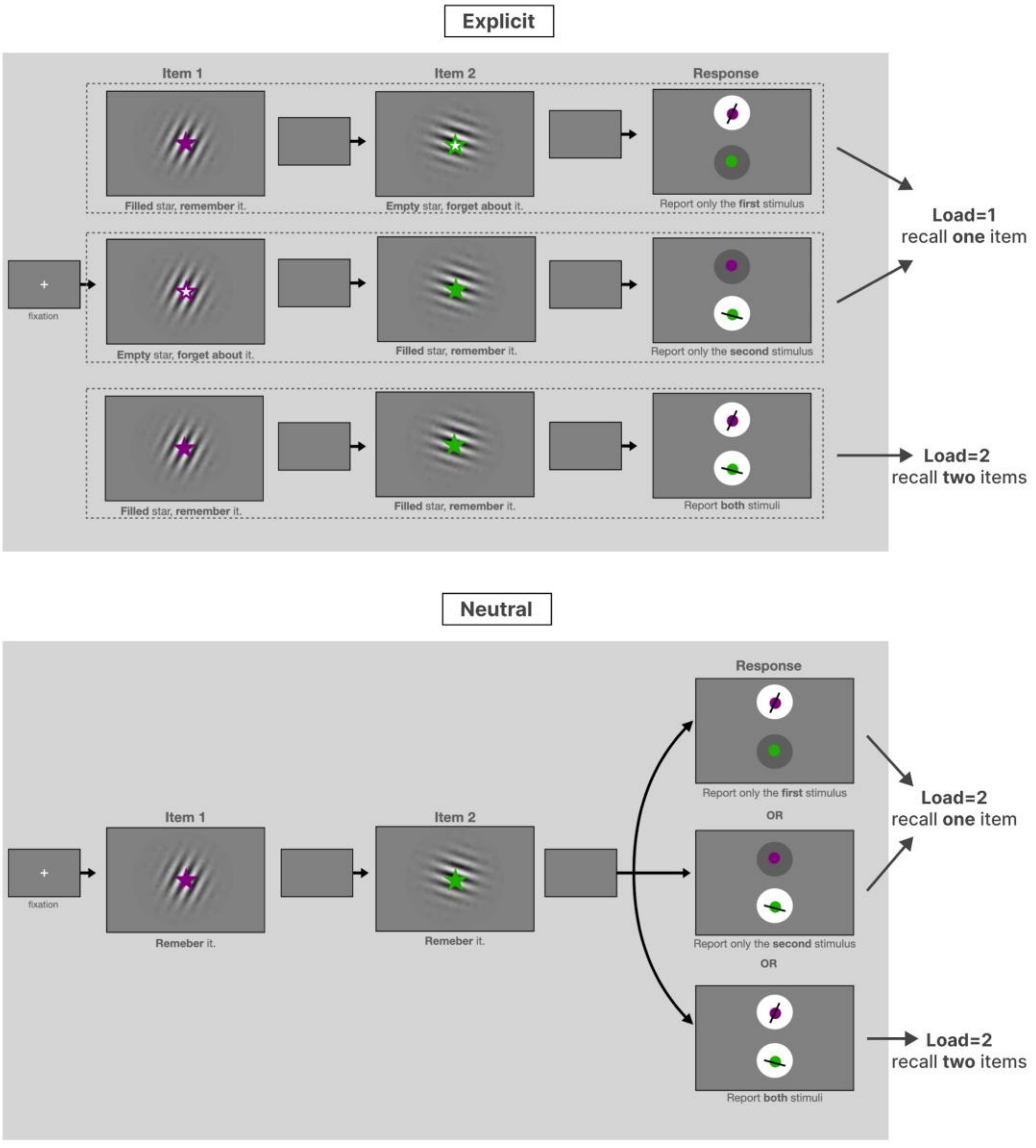

Figure S1: Experimental design diagram for the draw condition, as shown to participants (the wheel condition used the same design except that the response diagram consisted of a wheel and dots). Top: *Explicit condition, where participants always recalled the cued item(s). When one item was cued, memory load was 1 and one item was reported; when two items were cued, memory load was 2 and both items were reported.* ***Bottom:*** *Neutral condition, where both items were encoded but participants did not know which would be probed until the response phase. Memory load was always 2; in half the trials one item was reported, and in the other half both items were reported.*

We compare behavioral performance metrics between the two response formats as a function of WM load (Figure S2). When only one item is remembered, *draw* and *wheel* conditions do not differ significantly on either metric (error: $W = 248, n = 36, p = 0.187$; sharpness: $W = 316, n = 36, p = 0.798$). When both items were remembered the *draw* condition shows greater average error ($W = 181, n = 36, p = 0.016$), while sharpness remains comparable ($W = 298, n = 36, p = 0.592$).

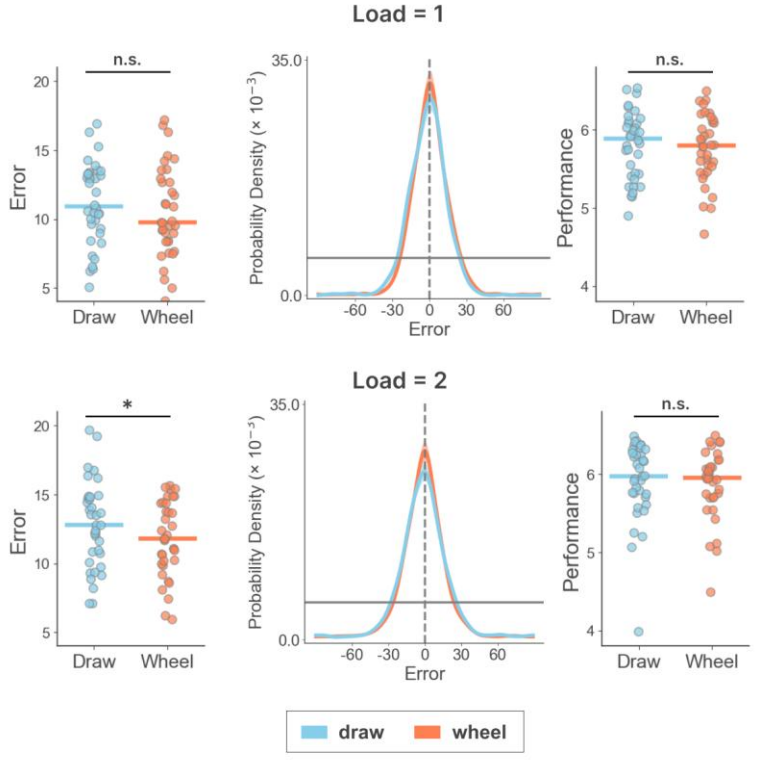

Figure S2: Behavioral performance is comparable between the draw and wheel conditions. Top: *Load = 1;* ***Bottom:*** *Load = 2.* ***Left:*** *Participants' average response error.* ***Middle:*** *Distribution of response errors.* ***Right:*** *Sharpness of the error distribution curves.*

Since average error effects differed between one- and two-item memory, we conducted a mixed-effects linear regression to further examine whether there is an interaction between memory load and response format. The results are summarized in the table below (both with n=36), showing that neither the response format nor its interaction with load has a significant effect on behavioral performance. Thus, results in the main text collapse across load conditions ($n = 36$):

Error: mixture linear model

	Estimate (β)	SE	z	p	95% CI
Intercept	11.12	0.51	21.98	<.001	[10.13, 12.11]
Load (2 vs 1)	1.66	0.44	3.77	<.001	[0.80, 2.53]
Format (draw vs wheel)	0.65	0.44	-1.47	0.140	[–1.51, 0.21]
Load × Format	–0.30	0.62	–0.49	0.627	[–1.52, 0.92]

Which suggest that neither the response format nor its interaction with load has a significant effect on behavioral performance.

Generalizability of encoding across conditions

Generalizing across response formats

In the main text, we reported successful reconstruction of the memorized orientation from gaze and hand motion patterns using a decoder trained on data pooled from both *draw* and *wheel* conditions. This was based on our assumption that the encoding of WM features in peripheral motor signals (i.e., the mapping between memorized orientation and gaze/hand movement patterns) is consistent across response formats, even if the strength of that encoding could differ (as shown in main text). In other words, we assume that if people tend to look more rightward when thinking of 45° in the *draw* condition, a similar rightward bias should also be present for 45° in the wheel condition – even if the precise mean position, variance, or strength of the bias differs. Thus, we assumed that our response format manipulation should not qualitatively change how WM features are encoded in peripheral signals, but may induce a qualitative change in the strength of stimulus evidence.

Here, we provide supplementary analyses to support this assumption.

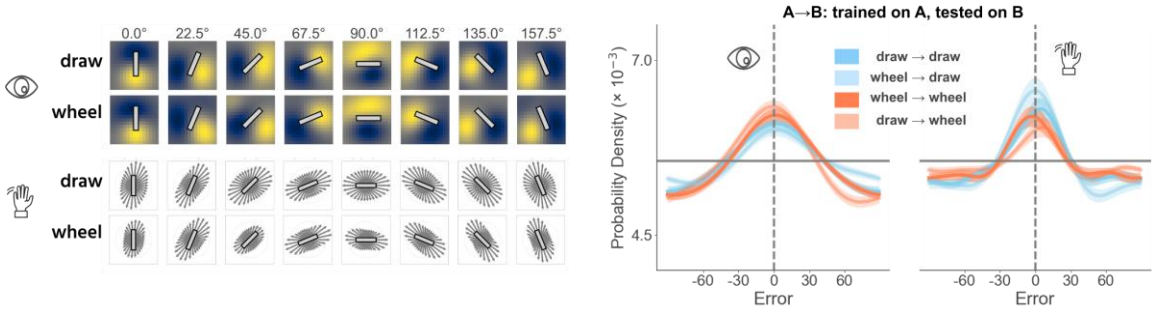

Figure S3: Encoding patterns are generalizable across two response format conditions. Left: *Gaze patterns (i.e. normalized 2D heatmaps of gaze distribution) and hand motion patterns (i.e. normalized angular distribution) as shown in Figure 2a, are plotted separately for the draw and wheel condition.* ***Right:*** *Distributions of reconstruction errors from the IEM analysis using a leave-one-subject-out cross validation paradigm. Each line corresponds to training the model on gaze or hand motion data from either the draw or wheel condition and tested on one of the two conditions. Error distribution curve for the model trained on condition A and tested on condition B is denoted as 'A → B'.*

First, for trials with a memory load of 1, we grouped trials by their memorized orientation (as in Figure 2a) and plotted the normalized patterns of gaze and hand motions during the delay separately for each response format (Figure S3left). Visually, the gaze and hand motion patterns appear similar across the two conditions. Specifically, we observe greater hand movements in the direction of the memorized orientation, a downward gaze bias for vertical orientations, and a counter-clockwise rotation of the 'hotspot' (i.e. regions with above-baseline gaze density, shown as yellow blobs) as the memorized orientation rotated clockwise.

Next, we quantified this apparent similarity. We trained and tested the IEM model on data from each response format condition separately, and compared the reconstruction errors across within- and cross-format cases (Figure S3right). The reconstruction results of the model trained on format A and tested on format B is denoted as 'A → B'. If the encoding patterns differ qualitatively between response format conditions, we would expect cross-format decoding (i.e. 'A → B') significantly worse than within-format ones (i.e. 'A → A' and 'B → B'). The results are summarized in the table below.

IEM evidence cross- vs. within-formats in gaze data

cross-formats	within-formats	$t(34)$ (p)	p(normal)	$W(n = 36)$ (p)
draw → wheel	draw → draw	1.70 (0.098)	0.010	284 (0.451)

	wheel → wheel	1.11 (0.276)	0.445	284 (0.451)
wheel → draw	draw → draw	-1.85 (0.073)	0.816	218 (0.072)
	wheel → wheel	-2.44 (0.020)	0.530	190 (0.024)

IEM evidence cross- vs. within-formats in hand motion data

cross-formats	within-formats	$t(34)$ (p)	p(normal)	$W(n = 36)$ (p)
draw → wheel	draw → draw	-4.32 (<.001)	0.288	98 (<.001)
	wheel → wheel	-2.30 (0.027)	0.932	202 (0.039)
wheel → draw	draw → draw	2.94 (0.006)	0.988	159 (0.005)
	wheel → wheel	3.60 (0.001)	0.055	142 (0.002)

We use ‘X > Y’ to denote that decoding performance is significantly better in X than in Y (here if ‘X’ is ‘A → B’, ‘Y’ could be any of ‘A → A’ or ‘B → B’). With the significance threshold adjusted using Bonferroni correction (0.0125), we did not find significant difference between cross-formats and within-formats in the gaze data. In the hand data, we found mixed results: (1) draw → draw > draw → wheel, (2) wheel → draw > wheel → wheel, and (3) wheel → draw > draw → draw. The evidence regarding whether within-format decoding is superior to across-format decoding is mixed: (1) supports it, whereas (2) and (3) argue against it.

With the significance threshold adjusted using Bonferroni correction (0.0125), we did not find significant differences supporting the hypothesis in gaze data, and found mixed results in hand data. We use ‘X > Y’ to denote that decoding performance is significantly better in X than in Y (here ‘X’ and ‘Y’ could be any two of ‘A->A’, ‘A->B’, ‘B->A’, ‘B->B’). In the hand data, we found (1) ‘draw → draw > draw → wheel’, (2) ‘wheel → draw > wheel → wheel’, and (3) ‘wheel → draw > draw → draw’.

As we did not find evidence that within-format decoding is superior, we can conclude that response format does not qualitatively alter the peripheral encoding of working memory. The stronger decoding evidence under certain response format conditions (as discussed in the main text) likely reflects greater consistency between peripheral signals and memorized orientations in those conditions.

Generalizing across individuals

In the main text we showed that models trained on the data across participants were used to successfully reconstruct the memory stimulus from a held-out participant's data. This implies that the patterns encoding WM features in the peripheral nervous system may generalize across people. In addition to training and testing using a leave-one-subject-out approach (denoted as the 'across-subject' model), we also trained and tested individual models for each subject (denoted as the 'within-subject' model), and examined the relationships between them.

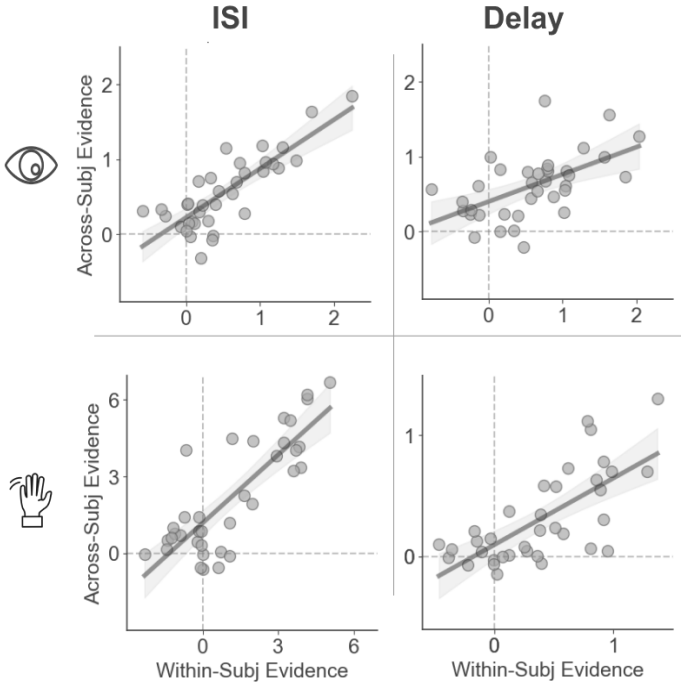

Figure S4: *Within- and across-subjects decoding evidence highly correlates. Correlations between within-subjects vs. across-subjects IEM decoding evidence, based on both gaze and hand movement recordings, during both the ISI and delay periods. Within-subjects scores reflect each individual's stimulus reconstruction strength when the model was trained and tested on their own data. Across-subjects scores reflect each individual's stimulus reconstruction strength when the model was trained on the data from all the other subjects and tested on the individual.*

The strength of an individual's stimulus evidence was strongly correlated between models (Figure S4; ISI gaze: $r = 0.65, p < 0.001$; ISI hand: $r = 0.89, p < 0.001$; delay gaze: $r = 0.36, p < 0.001$; delay hand: $r = 0.55, p < 0.001$). That is, participants whose

gaze or hand movement patterns enabled more precise within-subject reconstructions (using models trained on other trials) also showed more precise reconstructions when the model was trained on data from other participants. In other words, individual WM-related peripheral signals that are more internally consistent are also better decodable by models trained on aggregated data from others. These findings provide additional evidence for the generalizability of WM-related peripheral encoding patterns across individuals.

Time course of the contextual effect of response format

In the main text we have reported contextual effects of the task response-format,in both eye and hand effectors during the WM delay. In this section, we further explore the source of this effect, particularly at what stage of processing it starts to emerge.

Figure S5: The impact of response format emerges faster in gaze than in hand motions. ***(a)*** *Decoding evidence from a model trained on a fixed time window and tested on that window and all subsequent ones. Eleven critical 500 ms time windows were selected: the late encoding period (from cue onset to the end of the first stimulus display) (enc), the inter-stimulus interval (ISI), and nine windows during the delay period (d[i] for the i-th window in the delay). To reduce noise and maintain consistent window sizes, the first and last 250 ms of the ISI and delay were excluded. For each response format condition and effector, we trained a decoder on one specific time window and tested it on the same window and all those following. Sky blue and coral dots indicate test time points where decoding evidence significantly differs from the decoding evidence at the training window, for the draw and wheel conditions, respectively. Gray dots indicate significant differences in test decoding evidence between the two response format conditions at that time point.* ***(b)*** *Decoding evidence from a*

decoder trained and tested within the same time window. Sky blue and coral dots indicate that decoding evidence in the current window differs significantly from the preceding window for the draw and wheel conditions, respectively. Gray dots indicate significant differences in decoding evidence between the two response format conditions at that time point. Note that in both panels, training and testing include only items that had been presented and cued to remember up to that point.

First, we examined when during the trial the representation of memorized content begins to transform in preparation for the final motor response. To do this, we selected three critical 500 ms time windows—late encoding, the inter-stimulus interval (ISI), and the start of the delay period—and trained a decoder on each. We then tested each decoder on its own training window and all subsequent windows (Figure S5a). If decoding evidence from a later time window t_{test} significantly differs from that of the training window t_{train}, this suggests that the underlying representation has changed between those time points, either qualitatively or quantitatively.

For gaze data, we observed that when the decoder was trained on the encoding window, decoding evidence significantly decreased when tested on later delay windows in the *wheel* condition, compared to testing on the training window itself. However, when the decoder was trained on the ISI or the start of delay, decoding evidence at later windows remained comparable to that of the training window. This suggests that in the *wheel* condition, substantial transformation in the gaze-based representation occurs immediately after encoding, and stabilizes thereafter.

For hand data, regardless of which of the three time windows was used for training, decoding evidence changed significantly at later time windows. Notably, decoding accuracy increased sharply immediately after encoding, suggesting that hand-based rehearsal takes longer to build up. In subsequent windows, decoding evidence gradually declined, implying that hand-based WM representations evolve continuously over time, gradually forming action-ready states. Interestingly, the divergence between encoding and later windows was greater in the *draw* condition, whereas differences emerged later (from the delay phase onward) in the *wheel* condition, suggesting an even slower transformation in the *wheel* condition.

Given these differences in temporal dynamics, we expected the response-format–related contextual effects to emerge at different times: earlier in gaze, and more gradually in hand. This expectation was confirmed in a parallel analysis in which the decoder was trained and tested within the same time window (Figure S5b). In gaze data, the contextual effect (i.e., a significant difference in decoding between *draw* and *wheel* conditions) emerged as early as the encoding phase. In contrast, hand motion data showed no such difference until the later part of the delay period.

In summary, the transformation from input to action-ready WM representations follows distinct temporal profiles across effectors and response formats. For gaze, transformation occurs early and is most pronounced in the *wheel* condition. For hand, transformation begins during the ISI and continues throughout the trial in both formats, and progresses more slowly in the *wheel* condition. As a result, contextual effects of response format also emerge at corresponding stages: early in gaze, and later or more gradually in hand. Eyes and hands may be involved in different stages of processing of WM representations.