



Trial-by-trial mouse trajectory predicts variance in precision across working memory representations: A critical reanalysis of Hao et al. (2021)

Hyung-Bum Park¹ · Weiwei Zhang¹

Accepted: 16 May 2022 / Published online: 6 June 2022
© The Psychonomic Society, Inc. 2022

Abstract

Multiple representations in visual working memory (VWM) can vary in mnemonic precision. This inhomogeneity of VWM precision has received some support from recent studies with the whole-report procedure, in which all memory items are recalled in free or forced orders. Recently, Hao et al. (2021, *Cognition*, 214, 104739) added a novel item-selection stage before each memory recall and found smaller between-trial variance in mouse trajectory during the selection stage in free-recall condition as compared with forced recall, which was taken as evidence for less between-item interference and the resulting precision benefit under free recall. Here, we reanalyzed the original dataset with a different analytic approach and attempted independent hypothesis testing focusing on within-trial trajectory deviations. We found that the direction of trial-by-trial trajectory bias for the first to-be-recalled item was predictive of the relative mnemonic precision of the remaining items. Critically, this relationship was only present for forced recall but not for free recall. Hierarchical Bayesian modeling of recall errors further identified that this relationship was selectively driven by VWM precision. Together, our reanalysis provides evidence for the source of between-item interference and its direct association with variable precision of VWM representations, and further highlights the novel methodological benefits of probing memory decisional processes using mouse trajectory data.

Keywords Working memory · Mouse trajectory · Precision · Whole report

Visual working memory (VWM) is an online cognitive system that represents task-relevant information over a short period at the service of ongoing mental activities. Models of VWM assume that it is a highly limited process and characterize such limit by estimating the variability of error responses from continuous estimation tasks, as a summary statistic of mnemonic precision (Bays & Husain, 2008; Zhang & Luck, 2008). This variability estimate has successfully captured individual differences in VWM precision in various populations (Xie et al., 2018) and experimental manipulations such as emotion induction (Xie & Zhang, 2017). However, the variability measure is based on data collapsing across items and trials. Consequently, it

is insensitive to capture inhomogeneity in VWM precision across items, even though it by no means implies that different memory items are of homogeneous quality (i.e., equal amount of attentional resources allocated for each item).

Growing evidence showed that the quality of retained representations considerably varies across items, some are represented more precisely than others due to intrinsic (Oberauer, 2002) or experimental factors (Zhang & Luck, 2008). This inhomogeneous VWM precision has received further support from studies using a whole-report procedure that allows investigation of within-trial variability of retained items (Adam et al., 2017; Fougny et al., 2012). The whole-report task requires sequential recalls of all memory items; thus, a key manipulation is the order of recall. Specifically, participants could either select to-be-reported item for each recall (free recall) or simply report items in a forced order determined by the experiment program (forced recall). Overall, memory performance tends to decrease from earlier to later recalls, presumably due to prolonged between-item interference and/or forgetting (Oberauer & Lin, 2017; Shin

✉ Hyung-Bum Park
hpark053@ucr.edu

¹ 3132 Psychology Building, Department of Psychology,
University of California, Riverside, 900 University Ave.,
Riverside, CA 92521, USA

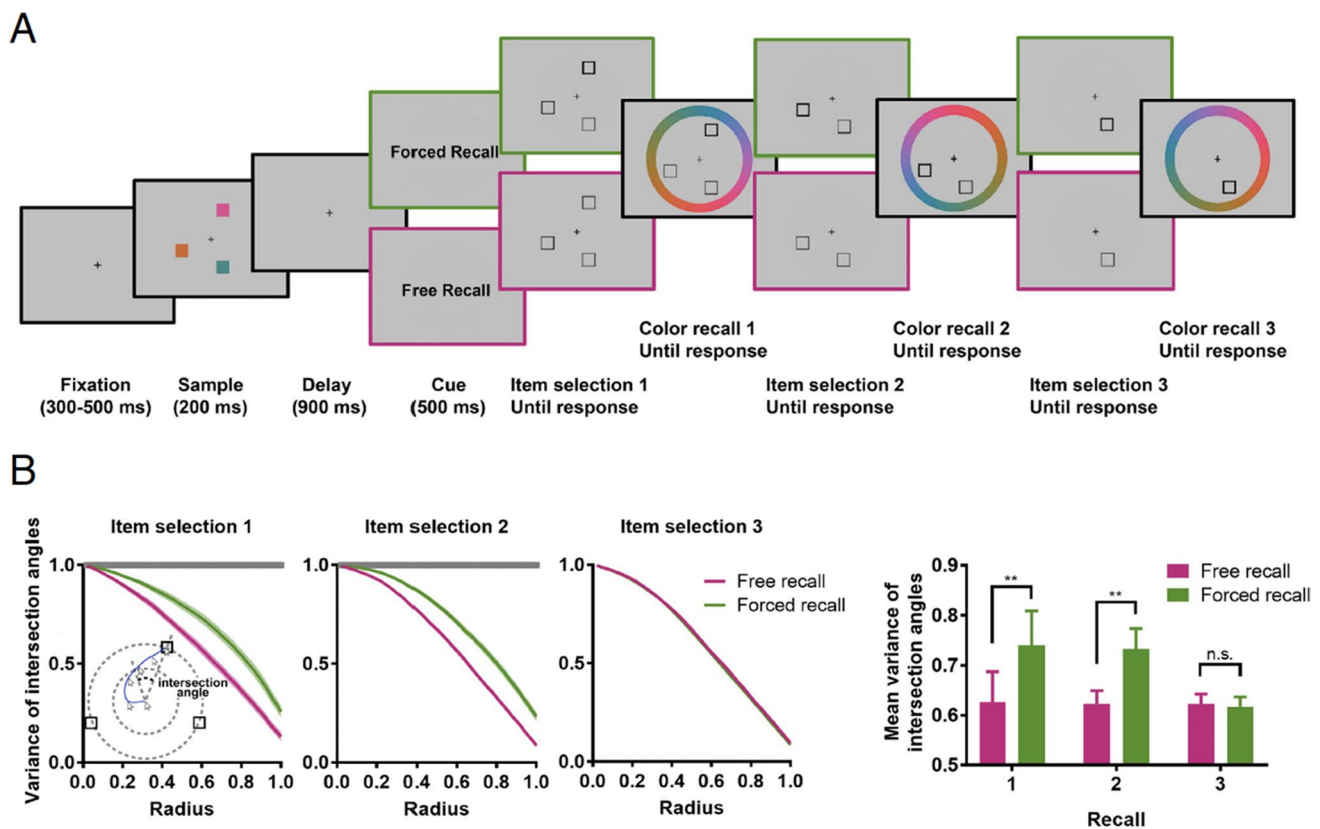


Fig. 1 Task procedure and primary results from Hao et al. (2021), reconfigured from the original Figs. 1 and 2. **a** Participants remembered three color squares briefly presented. After a short delay interval, a center cue appeared to indicate the recall type of the trial. Afterward, participants were asked to click the to-be-recalled item (item-selection; probe randomly selected under forced recall, whereas chosen by participants under free recall), then immediately followed by actual color recall. The item-selection and color recall were repeated for all three memory items. **b** Variance of mouse trajectory during the item-selection stage between free and forced

recall, for each of the first, second, and third recalled items, respectively. The variance was estimated from intersection angles at each radius normalized from the starting point to the ending point. The free recall benefit was associated with less variance in item-selection mouse trajectory for earlier recalls, interpreted as less between-item interference. Note, the variance of intersection angle at the last radius (0.98; right before mouse click response) was well above zero and considerably different between free and forced recall in the first two item selections

et al., 2017; Zhang & Luck, 2009). In addition, a robust free-recall benefit was found in earlier recalls, indicating a tendency of reporting the *best*-remembered item first (Adam et al., 2017). This suggests that participants had fairly accurate meta-knowledge of the variable quality across concurrent VWM items, consistent with previous findings that subjective confidence rating of memory performance strongly predicts mnemonic precision (Adam & Vogel, 2017; Rade-maker et al., 2012).

Hao et al. (2021) recently revised the whole-report task by adding a separate item-selection stage immediately before each recall (Fig. 1). Specifically, following a cue indicating recall type (free or forced recall), participants were instructed to identify the next to-be-recalled item by clicking its placeholder. Mouse cursor trajectory during item selection was recorded to provide additional data to explore the nature of the free-recall benefit arise from variable WM

precision across items. Critically, they found that free-recall benefit was associated with less variance in mouse trajectories during the item selection across trials. As such, they reasoned that the free-recall benefit may largely arise from reduced between-item interference during retrieval, as well as minimized mnemonic cost due to switching of representations for the focus of attention (FoA; Oberauer, 2002). Under forced recall, however, when an item currently outside of FoA (or even out of representations successfully maintained) is prompted to be recalled, the current FoA item needs to be switched (or simply undergo prolonged delay), leading to mnemonic degradation.

Furthermore, they examined the location preference effect and its asymmetric consequences in free and forced recall. Under free recall, items in the upper-left visual field had highest probability to be chosen, but without precision benefit over the other items from the nonpreferred locations.

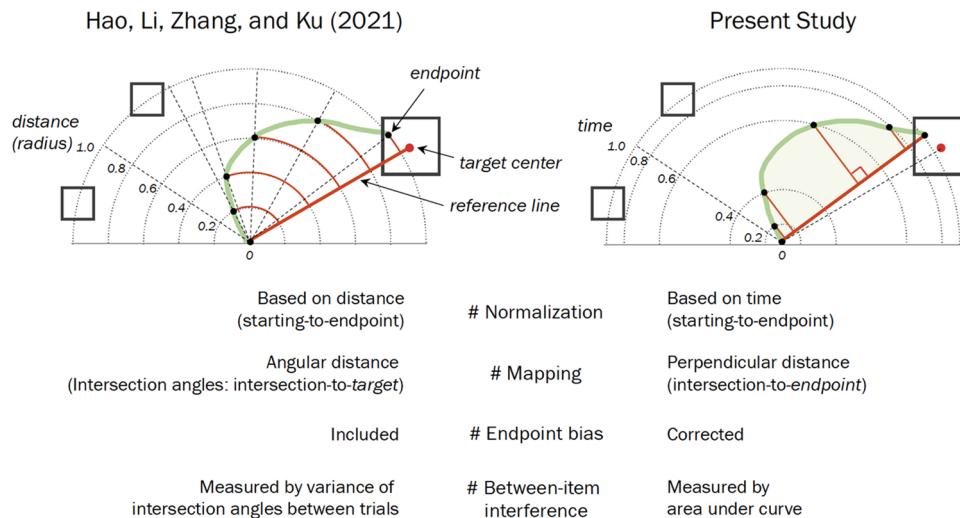


Fig. 2 Differences in the analytic method for the mouse trajectory data from the original and the present study. The green curve is an example trajectory from the mouse movement onset at the display center to the mouse click endpoint, depicted in a reconstructed setting of the item-selection stage. During this stage, three placeholders appeared at the locations of the three memory items and the partici-

pants selected one for the next memory retrieval. Four key differences are summarized: (1) data normalization, (2) mapping between interim mouse positions to the reference line, (3) inclusion of endpoint bias (i.e., the distance between the target center and the endpoint), and (4) between-item interference measure

Under forced recall, on the other hand, randomly chosen items from the preferred upper-left locations were associated with precision benefit, but this effect only held for the first recall. According to their interpretations inspired by Oberauer's interference model of WM (Oberauer & Lin, 2017), when an item in the preferred location (thus presumably in FoA) is not forcedly chosen for the first recall, the other chosen item must take place in FoA while the original FoA item is switched and underwent irreversible loss of precision.

Although we generally agree with their conclusions, their mouse trajectory dataset can be better utilized to provide a more direct examination of whether and how interitem dynamics can be explained by variable mnemonic quality across items. The rich mouse trajectory dataset from Hao et al. (2021) provides an additional opportunity to test the interference-based account of the free-recall benefit. For instance, between-item interference was inferred from between-trial estimates of circular variance in mouse cursor positions. Although it can be a straightforward index of between-item interference, the measure of between-trial variance fails to take advantage of within-trial dynamics of recall process in whole-report task. Alternatively, the source of between-trial variance can be reduced into directional biases in mouse position on each trial (e.g., clockwise or counterclockwise from target). As such, the patterns of trial-by-trial trajectory deviations could provide a more sensitive way to examine variable VWM precision given the substantial literature on memory-guided attention demonstrating that the contents of active WM can guide selective attention

toward an item location that matches WM (Olivers et al., 2011; Soto et al., 2008).

The present study has thus reanalyzed the dataset from Hao et al. (2021) with different analytic approaches. We first addressed the potential artifacts in mouse trajectory analyses in the original study. More importantly, we developed a trial-level mouse trajectory categorization method and tested the interference account of the free-recall benefit and its relationship with inhomogeneous WM precision based on within-trial trajectory deviations.

Method

Data reanalysis and rationales

The original data from Hao et al. (2021) were retrieved from the authors' deposit of the data at Open Science Framework (<https://osf.io/67upz/>). Figure 2 illustrates the differences in the analytic method for mouse trajectories from the original study and the present study. Specifically, the two methods are different in three aspects; normalization, mapping cartesian on polar coordinates, and inclusion of endpoint bias.

First, the original study normalized the mouse trajectories based on the *distance* from starting point at the center of the screen to its ending point to the circle ($x^2 + y^2 = 1$). This distance-based normalization can be useful in estimating variance at the same points along with starting-to-ending movements across trials. However, it ignores velocity which is likely variable across trials. We therefore normalized

trajectories based on two *time* points, from when a movement initiated (i.e., onset latency) to when the ending point is selected with a mouse click (i.e., reaction time). The onset latency is defined on each trial as the time the mouse cursor deviates from its original location by three pixels in a horizontal or vertical direction ($21.8 \pm 11.2\%$, $3.4 \pm 1.4\%$, and $3.3 \pm 1.2\%$ normalized time from display onset to click, for the first, second, and third item selection, respectively).

Second, to map the raw mouse cursor positions in Cartesian coordinates (x, y) to polar coordinate ($-\pi$ to π), Hao et al. (2021) computed *intersection angles* at different radii ranging from 0.02 to 0.98 in steps of 0.02. The intersection angle was calculated as the angular distance from each *intercept point* along the radii (origin-to-intersection vector) to the *target* (origin-to-target vector), not to the *endpoint*. The intersection angles rely on, maybe unnecessarily, an assumption that the interim mouse positions (and especially its angular distance from target) are psychologically meaningful such that they reflect the moment-by-moment movement goals. Hao et al. (2021) attempted to examine the source of trajectory deviation with a formal model comparison between two competing hypotheses, the Alternation and the Integration models. The two models attribute trajectory bias either to a probabilistic swap tendency (Bays et al., 2009) or to an attraction toward a vector sum over the target and distractor positions (e.g., response vector model; Tipper et al., 1997), respectively (see Hao and colleague's Supplemental Methods for details). Nonetheless, there has been an ongoing debate regarding underlying mechanisms of mid-flight deviations and what it truly means (Spivey et al., 2005; van der Wel et al., 2009). Moreover, intersection angles only assign polar distance from the target, regardless of the actual amount of deviation at different radii on the raw coordinate system. Consequently, this measure is likely to overestimate the circular variance at those intersection radii corresponding to early movements, since initial movements often tend to deviate drastically in *direction* due to motor noise but typically not in actual *distance*. For a concrete example, the two mouse cursor positions at 20% and 40% radii in Fig. 2 would be converted to similar intersection angles (i.e., angular deviation from the target) despite that fact that they are double-sized in perpendicular distance to the references line. For these reasons, the present study analyzes mouse trajectories in their raw coordinate system (e.g., pixels) which is more precise to characterize how trajectory tracks interdependency of multiple VWM representations.

Lastly, any biases in the endpoint (i.e., the distance from a mouse click location to the designated target probe location) would affect the measure of intersection angles in Hao et al. (2021), which could further contribute to the estimated circular variance postulated to measure between-item interference. There are several different methods quantifying curvature in trajectories (Ludwig & Gilchrist, 2002; Van

der Stigchel et al., 2006). Although all of them are with valid reasons to focus on different aspects of trajectories, a choice of reference straight line (either to a target or to the endpoint from the mouse onset position) may create considerable differences in the resulting measure. As one can infer from Fig. 2, the endpoint bias affects not only the variability at the very last radius but also every radius back to the starting point. Moreover, trajectory curvature is mathematically and conceptually independent from the endpoint bias. That is, a greater trajectory curvature is not necessarily accompanied by a greater amount of endpoint bias, and their directions may differ as well (e.g., curved clockwise from the target, but with a counterclockwise ending point, vice versa). Therefore, endpoint bias could be a serious artifact especially when the primary measure of interest is a variance of intersection angles. As shown in Fig. 1b, the variance estimates of the intersection angle at the last radius (0.98) were well above zero and considerably different between free and forced recall in the first two item selections. If such endpoint bias is controlled, the difference in the total variance of trajectory between free and forced recall might be diminished.

To overcome these artifacts of intersection angle measure, our reanalysis used the *area under trajectory curve* (AUC) to characterize the extent of trajectory deviation.¹ AUC is the geometric area between the observed trajectory and a reference straight-line from the onset to the endpoint. A higher AUC value indicates a greater deviation toward alternative locations. Also, the sign of AUC can represent the direction of deviation either toward clockwise (+) or counterclockwise (−) from the reference line.

Another critical hypothesis tested in the present study was that the between-item interference manifested as the between-trial variability of item-selection mouse trajectories, identified in Hao et al. (2021), could be directly accounted for by the within-trial dynamics of variable precision of multiple VWM representation (see Results for detail). For this novel hypothesis testing, we selected trials where the first recall item was placed within four displacements from the other two items on the opposite side among eight possible locations (i.e., first target placed somewhere in the middle of second and third items). For example, from Fig. 3a, when the first recall item was placed at the location marked as “1,” only those trials where the second and third items were placed at the opposite side between location “1” and “5” were included for data analysis. In other words, all other types of trials where second and third recall items were placed on the same side relative

¹ We also calculated another popular measure of trajectory deviation, the maximal deviation (MD) at the point where the curvature is maximum for sanity check. Participants' MDs were highly correlated with their AUCs, $r(14) = .95$, 95% CI [.87, .98], $p < .001$. The results were comparable between measures using AUC and MD.

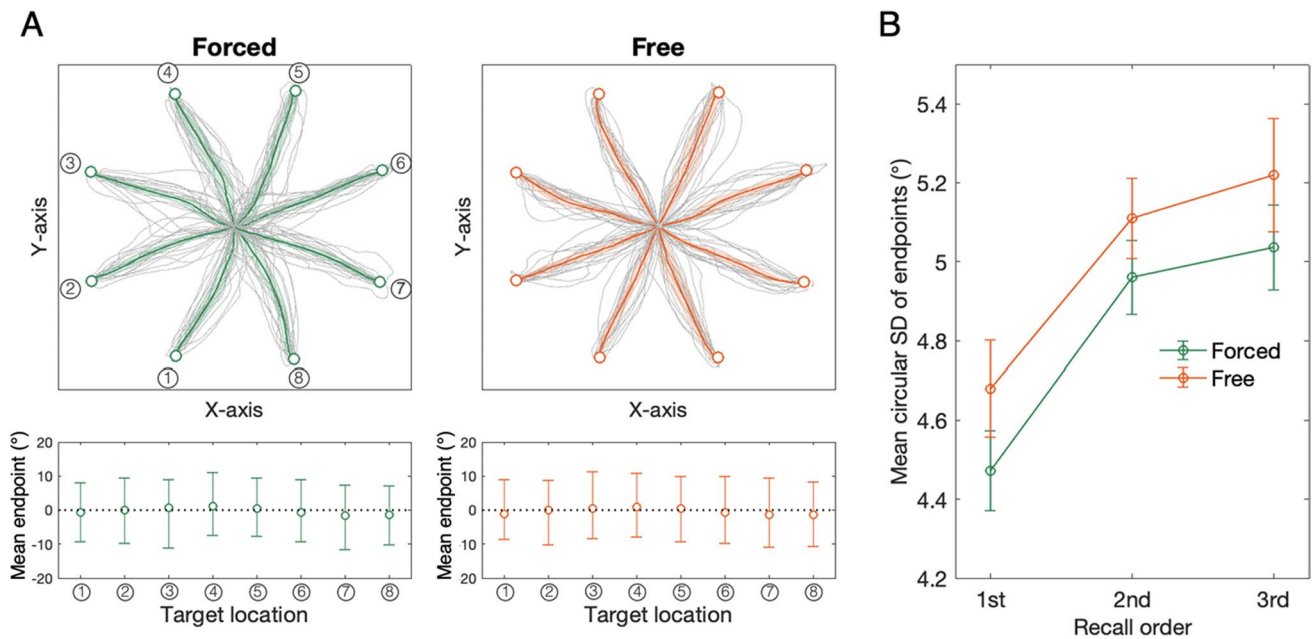


Fig. 3 a Reconstructed item-selection mouse trajectories over eight possible target locations for forced and free recall, collapsed across participants and selections for the three items. Colored curves and shades represent group mean trajectory and 95% CI, whereas grey curves are individual participant mean trajectories (top panel). The bottom panel illustrates the extent of variability at the mouse-click endpoints across trials (indicated by error bars representing 95% CI

of total trials), which was similar across eight target locations (M and SD of $CI_{95\%}$ upper – lower bound range = $18.36 \pm 0.99^\circ$). **b** The mean and standard error of the endpoint variability measured by circular standard deviation (SD) as a function of recall order (first, second, and third item selection) and recall type (forced vs. free recall). The endpoint variability was systemically larger under free recall than forced recall

to the first recall item were excluded. We allowed uneven clockwise and counterclockwise displacements from the first item to maximize the trial number (e.g., second and third item placed three steps clockwise and two steps counterclockwise to first item, respectively) This resulted in 1,917 trials remained for data analysis out of the total of 3,840 trials (49.9%).

For the analyses of recall errors, we applied the hierarchical Bayesian approach for the extended mixture model (Zhang & Luck, 2008). The three free parameters, μ , SD , and *guessing*, of the model represent the location (μ , mnemonic appearance) and the width (SD ; inversely related to mnemonic precision) of the central peak distribution of the recall errors (i.e., noisy mnemonic representation), and the probability of random *guessing* that is not driven by mnemonic evidence, respectively. The hierarchical Bayesian estimation samples plausible posterior parameter values at the population level, while simultaneously accounting for different sources of variabilities from individuals, conditions, and trials using Markov Chain Monte Carlo simulations (16,000 MCMC samplings after 16,000 warm-ups). The main effects of each population-level parameter were estimated in a general linear model, sampling from the normal distribution where the mean is a sum of the fixed (condition) and random effect (individual), and the variability

term describes the individual-by-condition interaction effect (Rouder et al., 2014).

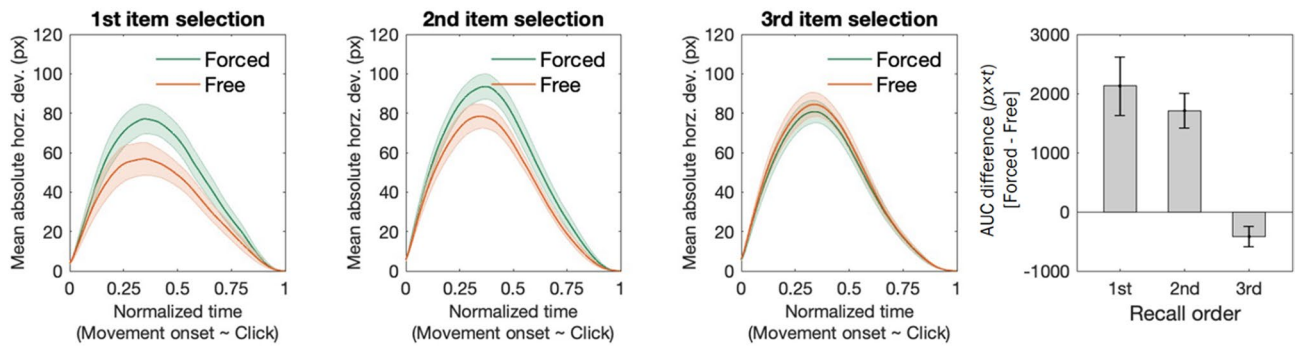
We chose reasonable to non-informative priors for all parameters to minimize biases due to the choice of priors. The mean and the 95% credible interval (highest density interval, HDI) of the posterior distribution were treated as a point estimate and an analogue of a frequentist confidence interval (CI), respectively. Statistical inferences were made based on the range of HDIs (e.g., whether the positive or negative side of 95% HDIs for the condition effect crosses over zero), as the strength of evidence (Kruschke, 2014).

Results

Endpoint bias and its impact on between-item interference

In Hao et al. (2021), three memory item locations were randomly selected from eight possible placeholders. We sorted every trial based on those eight recall target locations across participants and recalls and reconstructed the original mouse trajectories. We then estimated the amount of endpoint bias at each location, separately for forced and free recall (Fig. 3a). The extent of endpoint variability was

A. Endpoint-corrected



B. Endpoint-included

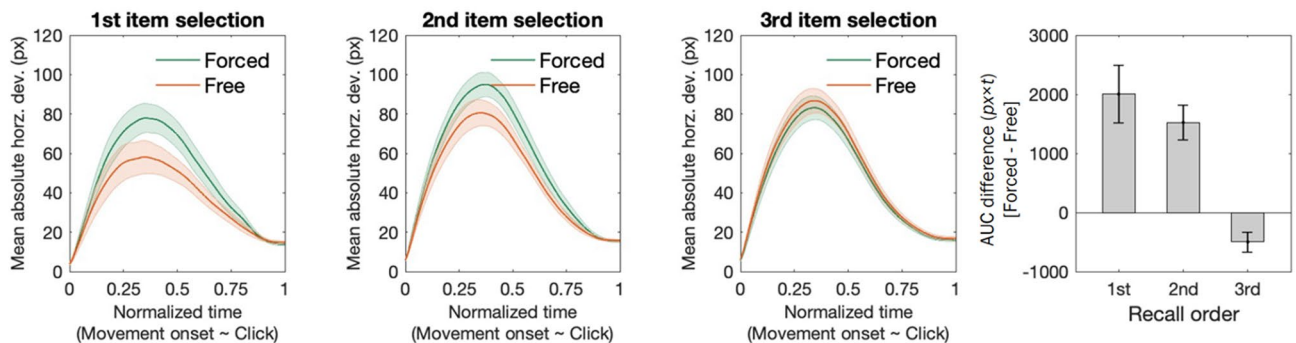


Fig. 4 Mean absolute horizontal deviations as a function of time normalized from the mouse movement onset to click endpoint, separately for recall type (forced vs. free recall) and recall order (first, second, and third item selection), respectively. Either trial-specific endpoint (a) or target probe center (b) was set to be placed at the top of the mouse onset at the display center as a reference point. The rightmost

panels illustrate the differences in trajectory area under curve (AUC) between forced and free item-selections, as a measure for the relative amount of between-item interference between recall type. Positive values indicate greater between-item interference. All error bars including the shaded curves represent the standard error of mean

estimated for individual participants by calculating circular SD as a function of recall order (first, second, vs. third) and recall type (forced vs. free), separately (Fig. 3b). A two-way repeated-measures analysis of variances (ANOVA) revealed significant main effects of recall order, $F(2, 30) = 15.36, p < .001, \eta_p^2 = .51$, and recall type, $F(1, 15) = 10.24, p = .006, \eta_p^2 = .41$, without significant interaction effect, $F(2, 30) = 0.06, p = .943, \eta_p^2 = .00$. Surprisingly, endpoint bias was greater in free recall than in forced recall, which is opposite to what can be inferred from Hao et al. (2021). This discrepancy is likely due to different analytic approaches for mouse trajectories.

Our next aim is to replicate the primary finding of greater between-item interference under forced recall while evaluating the impact of endpoint bias by comparing two sets of results *with* and *without* endpoint bias. Figure 4 illustrates the mean absolute horizontal deviations along time normalized to movement onset-to-endpoint. All trajectories were rotated such that a reference point of endpoint (endpoint corrected) or target (endpoint included) to be placed

perpendicular from the starting point. Consequently, any interim deviations can be represented as deviation on the horizontal dimension. We calculated AUCs then took [forced – free] difference as a measure for the relative amount of between-item interference.

For the endpoint-corrected set, planned one-sample t tests comparing AUC differences from zero at each recall revealed significant *positive* differences in first selection ($+2,129.1_{px \times t}$ [CI_{95%} hereafter: $+1164.5, +3093.6$]), $t(15) = 4.33, p = .001$, Cohen's $d = 1.12$, and second selection ($+1,709.7_{px \times t}$ [$+1138.5, +2280.9$]), $t(15) = 5.87, p < .001, d = 1.51$, whereas a significant *negative* difference in third selection ($-419.2_{px \times t}$ [$-758.3, -80.2$]), $t(15) = -2.42, p = .028, d = -0.63$). These results replicate the original finding of greater between-item interference under forced recall in selection of first two items. We observed similar patterns for the endpoint-included set (Fig. 4b), with AUC differences [forced – free] in the first, second, and third item selections were $+2004.3$ [$+1047.6, +2961.0$], $+1523.5$ [$+954.0, +2093.1$], and -497.7 [$-831.6, -163.8$], respectively, all ps

$< .011$, $d_s > 0.75$. Additional two-way repeated-measures ANOVA for AUC differences as a function of recall order and endpoint (corrected vs. included) revealed significant main effects of recall order, $F(2, 30) = 15.71$, $p < .001$, $\eta_p^2 = .51$, and endpoint, $F(1, 15) = 26.69$, $p < .001$, $\eta_p^2 = .64$. However, there was no significant interaction between them, $F(2, 30) = 1.04$, $p = .361$, $\eta_p^2 = .07$.

To summarize, we identified substantial amount of endpoint bias varied between forced and free recall. The endpoint bias nonetheless played a minimal role in between-item interference. These findings reaffirm the original conclusion that under forced recall, participants suffered greater between-item interference during the item-selection stage.

Within-trial trajectory predicts relative precision of VWM items

Hao et al. (2021) discussed potential connection between *between-item interference* and *VWM inhomogeneity*. Specifically, the greater between-item interference under forced recall was attributed to increased swap tendency toward the alternatives (i.e., misdirecting mouse cursor toward the other non-probe items), whereas VWM inhomogeneity was supported by asymmetric consequences of the location preference effect in free and forced recall. These interpretations, however, are drawn from less concrete evidence by combining two separate empirical results supporting each concept. Further, it relies on particular assumptions for the nature of FoA such as irreversible mnemonic degradation due to switch of items for a single, fixed capacity of FoA state (see for other accounts of FoA capacity; Beck & Hollingworth, 2017; Beck et al., 2012; Cowan, 2001; Williams et al., 2022).

Alternatively, here, we hypothesize that inhomogeneous VWM precision may directly give a rise to inhomogeneous between-item interference between forced and free recall. It is expected that trial-by-trial dynamics of mouse trajectory during the first item selection are related to subsequent item recall performance. We predict that mouse trajectory to the first forced target would be attracted toward one of the two nontargets, where its direction may depend on the relative precision of those items. Specifically, on two-thirds probability, the first forcedly chosen target would not be of the *best* precision and mouse movement to this nonbest item would be deviated toward the other *best* item, reflecting the relative attractive force driven by variable precision among three memory items.

To test this idea, we created a two-by-two trial-categorization matrix, where one dimension categorizes whether the first item-selection trajectory deviated toward second or third recall item (by the sign of trajectory AUC), whereas the other dimension categorizes the relative recall performance (second – third absolute recall errors; a negative value indicates better recall for the second item than the

third item). Note, for the first dimension of the sign of trajectory AUC, AUCs from trials with the third item presented counterclockwise (–) to the first item were reversed. Hence, negative and positive AUCs represent first item-selection trajectory deviated toward the second item (–) and the third item (+), respectively. For the second dimension of the absolute (second – third) recall error differences, the overall main effect of recall order (i.e., second recall more precise than third recall in general) was corrected by weighting a half-size of the main effect (i.e., the difference between the bottom row sum and the top row cell sum) to the top row cells (i.e., when third recall more precise) while subtracting it from the bottom row cells (i.e., when second recall more precise). This was to equate the null probability of top and bottom cells at 50% each. Trials with the absolute relative (second – third) error smaller than 2° were excluded to prevent ambiguity in categorization.

Each trial was categorized into one of quadrant categories based on the sign of the measures. Critically, according to our hypothesis, trial-categorization accuracy defined by the summed probability of the first (+/+) and the third quadrants (–/–) should be greater than the chance-level at 50%. In other words, there will be a greater proportion of trials in which the direction of the first item-selection trajectory bias was predictive of the relative recall precision of the second and third items. Moreover, this asymmetric diagonal pattern in the two-by-two matrix should only be present for forced recall, but not for free recall where the first recalled item had the best precision.

The results were consistent with these predictions (Fig. 5). We found trial-categorization accuracies to be significantly above chance for forced recall (54.7% [51.0%, 58.3%], one-tailed one-sample $t(15) = 2.69$, $p = .008$, $d = 0.70$, $BF_{10} = 3.60$, but stayed around chance for free recall (50.6% [46.1%, 55.0%]), $t(15) = 0.27$, $p = .396$, $d = 0.07$, $BF_{10} = 0.26$. The same one-tailed paired-samples t test between forced and free recall yielded only marginally significant difference but with reliable effect size measures, $t(15) = 1.74$, $p = .051$, $d = 0.43$, $BF_{10} = 1.63$. This suggests that the direction of trajectory bias during the first item selection was to some extent predictive of the relative precision of the two subsequently recalled items. This provides supporting evidence for our hypothesis that between-item interference may be directly related to variable precision among VWM representations.

To further identify how the first item-selection trajectory predicts representational quality of the subsequently recalled items, we fitted second and third recall errors with the extended mixture model (Zhang & Luck, 2008) using hierarchical Bayesian method, separately for different first item-selection trajectory bias direction (toward second or third item; left and right cells in Fig. 5). Figure 6 summarizes the resulting population-level posteriors of μ , SD , and *guessing* parameters, showing their difference between two

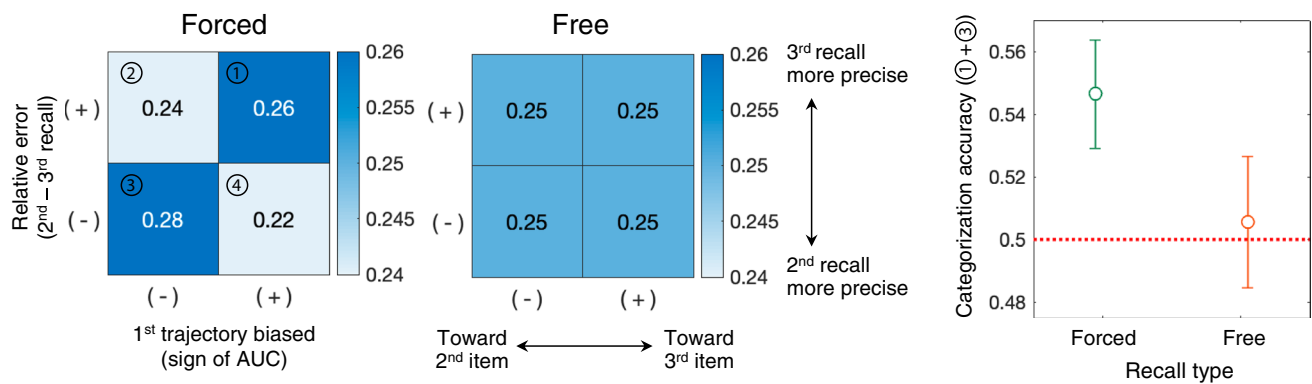


Fig. 5 Proportion of trials as a result of trial-categorization consisting of a 2-by-2 matrix, where the horizontal dimension categorizes whether the first item-selection mouse trajectory deviated toward the second or third recalled item (by the sign of area under curve; AUC), whereas the vertical dimension categorizes by the sign of the relative absolute recall errors (second – third; negative sign indicates the second recall being more precise than the third recall). The first

(+, +) and the third quadrants (–, –) are consistent with the prediction that the direction of first item-selection mouse trajectory under forced recall will be predictive of the relative precision of the second and third recall items. (Right-panel): Consistent with this prediction, trial-categorization accuracy (proportion of trials in the first and third quadrants) was significantly above chance under forced recall but stayed at around the chance level under free recall

recalls (second – third). Further analyses based on posterior mean with 95% HDI for the difference between two cases when the first item-selection trajectory deviated toward the second or third recalled item revealed only credible difference from the *SD* parameter for forced recall (+6.84, HDI_{95%} [+1.09, +13.42]), whereas no other parameters from both recall type were credibly different. These results indicate that the variability in precision, but not the representational shift (μ) or *guessing*, among VWM representations on a given trial manifested in the trial-specific directional mouse trajectory bias during the first item selection under forced recall.

Discussion

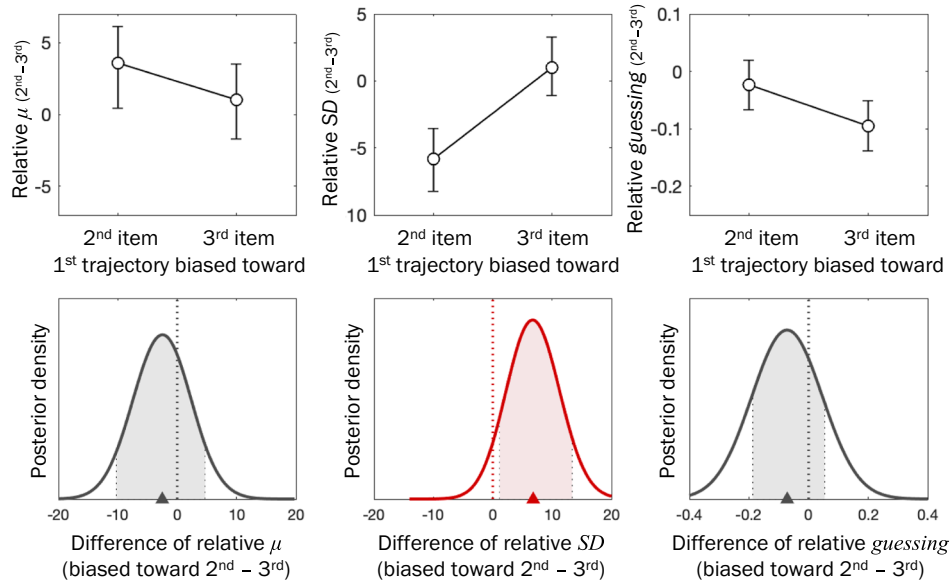
The present study reanalyzed the behavioral and mouse trajectory data from Hao et al. (2021), with a different analytic approach that controls for substantial and asymmetric variance in mouse endpoints between free and forced recall. We first replicated the original findings of the greater between-item interference under forced recall. More importantly, we examined the trial-level mouse trajectories to investigate how they are inherently associated with inhomogeneous VWM precision. The results from our novel trial-categorization method revealed that the direction of trajectory bias during the first item selection was predictive of the relative recall performance of the remaining items. Specifically, a memory item producing stronger attraction of the item-selection mouse movement was recalled more accurately than the other item. This was only valid for forced recall but not for free recall. Furthermore, using hierarchical Bayesian modeling of the recall errors, the categorization pattern was solely driven

by WM precision, not by other factors of recall such as bias in feature appearance or failure of remembering. These findings provide strong support for our hypothesis that variable precision of concurrent VWM representations directly gives a rise to the asymmetric trajectory bias resulting from nontarget items at the moment, a novel manifestation of between-item interference. Taken together, the present reanalysis provided direct mouse trajectory evidence for the link between between-item interference and variable precision.

These results are in general consistent with the operations of active-state WM contents. Accumulating evidence suggests that new sensory inputs that match active WM contents can capture attention, even when such guidance is irrelevant to the current task goal (Pan et al., 2016; van Moorselaar et al., 2014). WM guidance not only occurs to the internal shift of attention but also manifests in preparation and execution of motor actions such as eye movements or reaching behavior (Hollingworth et al., 2013; Theeuwes et al., 2009). Our findings thus provide an important extension of these previously observed interactions between WM and attention.

Although the present results are in support of the inhomogeneity in VWM precision, it does not directly speak to the ongoing debate on the capacity of FoA or the number of active templates that can guide attention and action (Cowan, 2001; Oberauer, 2002; Olivers et al., 2011; McElree, 2001; Zhou et al., 2020). It is also possible that not necessarily one or more discrete items have to enter the qualitatively privileged state to guide behavior. Instead, all the representations in mind at the moment may compete for selection and result in a gradient interaction with attention, depending on their relative precision. Mouse movement trajectory could be an effective measure to address this question in future research.

Forced-recall



Free-recall

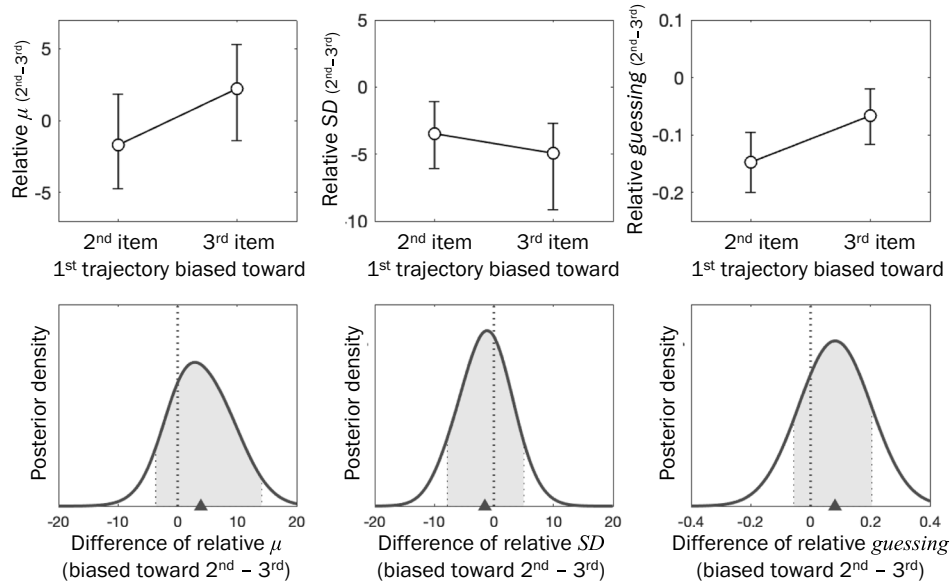


Fig. 6 Hierarchical Bayesian posterior distributions of the three parameters (μ , SD , and *guessing*) from the extended Zhang and Luck (2008) mixture model, representing the difference between second and third recall errors. For each recall type (forced and free recall), the top row shows the mean and the 95% highest density interval (HDI) of the posterior parameter values (second – third recall) as a function of the first item-selection trajectory bias direction (toward second or third item). The bottom row illustrates the difference between these two conditions (trajectory biased toward second – third recall item), depicted in the nonparametric kernel density fits

of the resulting posteriors (solid black curves) with shaded grey area representing the lower and upper bound of $HDI_{0.95}$. These resulting posteriors, in other words, represent the interaction effect between recall order (second and third) and trajectory bias (toward second and third). Only the precision parameter (SD) under forced recall reveals its $HDI_{0.95}$ not crossing over zero, thus indicating a credible interaction effect. In other words, it is the variability in precision, but not representational shift (μ) or *guessing*, across VWM representations on a given trial that gave rise to the trial-specific directional mouse trajectory bias during the first item selection under forced recall

There are some caveats in the current findings. First, we found that the item selections were more *erroneous* (i.e., greater endpoint variance) in free recall than forced

recall, which is opposite to Hao and colleagues’ results. This discrepancy may originate from different trajectory analysis methods. In addition, participants were not

instructed to click at the center of the item placeholder, thus the amount of endpoint bias could arise from motor execution. That is, under free recall, the participants might try to proceed to the recall quicker with faster but less careful mouse clicking. Consistent with this speculation, item-selection time was faster for free recall than forced recall (Hao et al., 2021).

Second, although we obtained the above-chance trial-categorization accuracy for forced recall, the effect was weak (54.7%) though exhibited reliable effect sizes (Cohen's $d = 0.70$, $BF_{10} = 3.60$). This could be largely due to the absence of the trajectory effect on one-third of forced-recall trials where the first forced target would be by chance of the best precision, yielding no apparent trajectory bias to the other items. Consequently, this one-third of forced-recall trials was nothing different from those free-recall trials, attenuating the predictive power of trajectory bias direction for the relative precision of the other items. Also, the relative distance of the second and third item locations to the first item was not controlled to be symmetrical in the original experiment, adding additional noise to the measure of the AUC sign. Nonetheless, our hierarchical Bayesian modeling of the second and third recall errors identified which of the mixture model parameters was the source of asymmetric trajectory bias, suggesting its robustness to small experimental effects (Park et al., 2021).

The present study further extended the literature on investigating the continuous nature of internal cognitive processing using response trajectory data across the various domains such as eye movement (Kowler et al., 1995; Van der Stigchel et al., 2006), hand movement (Abrams & Balota, 1991; Song & Nakayama, 2008; Welsh & Elliott, 2004), and mouse trajectory (Spivey et al., 2005; van der Wel et al., 2009). However, most previous studies relied on trial-average effects and draw conclusions from the comparison between experimental conditions. The present study assesses the within-trial trajectory in response to ongoing WM-based decisions, which provides a more effective investigation of the dynamic interactions among multiple memory representations. Taken together, the present findings have demonstrated that mouse trajectory during the selection of recall item not only tracks the total amount of between-item interference but rather directly reflects the variation in representational precision of VWM items. Our findings further highlight the methodological importance of mouse trajectory analysis for hypothesis testing.

Authors' contributions Not applicable.

Funding This study was made possible by funding support from the National Institutes of Health (1R01MH117132).

Data availability The datasets from the original study (Hao et al., 2021) are available in the Open Science Framework (<https://osf.io/67upz/>).

Code availability Not applicable.

Declarations

Conflicts of interest/Competing interests The authors have no relevant conflicts of interest or competing financial/non-financial interests.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science*, 2(3), 153–157.
- Adam, K. C., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, & Psychophysics*, 79(5), 1506–1523.
- Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, 97, 79–97.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890), 851–854.
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 1–11.
- Beck, V. M., & Hollingworth, A. (2017). Competition in saccade target selection reveals attentional guidance by simultaneously active working memory representations. *Journal of Experimental Psychology: Human Perception and Performance*, 43(2), 225–230.
- Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science*, 23(8), 887–898.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3(1), 1–8.
- Hao, Y., Li, X., Zhang, H., & Ku, Y. (2021). Free-recall benefit, inhomogeneity and between-item interference in working memory. *Cognition*, 214, 104739.
- Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*, 24(5), 790–796.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35(13), 1897–1916.
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Ludwig, C. J., & Gilchrist, I. D. (2002). Measuring saccade curvature: A curve-fitting approach. *Behavior Research Methods, Instruments, & Computers*, 34(4), 618–624.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 817–835.

- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421.
- Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1), 21–59.
- Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7), 327–334.
- Pan, Y., Luo, Q., & Cheng, M. (2016). Working memory-driven attention improves spatial resolution: Support for perceptual enhancement. *Attention, Perception, & Psychophysics*, 78(6), 1625–1632.
- Park, H. B., Ahn, S., & Zhang, W. (2021). Visual search under physical effort is faster but more vulnerable to distractor interference. *Cognitive Research: Principles and Implications*, 6(1), 1–14.
- Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, 12(13), 1–13.
- Rouder, J. N., Morey, R. D., & Pratte, M. S. (2014). Bayesian hierarchical models. In H. Batchelder, H. Colonius, E. Dzharaforov, & J. I. Myung (Eds.), *New handbook of mathematical psychology. Volume I: Measurement and methodology*. Cambridge University Press.
- Shin, H., Zou, Q., & Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, 17(14), 1–24.
- Song, J. H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research*, 48(7), 853–861.
- Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, 12(9), 342–348.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102(29), 10393–10398.
- Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, 132(2), 106–114.
- Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, 4(1), 1–38.
- Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews*, 30(5), 666–679.
- van Der Wel, R. P., Eder, J. R., Mitchel, A. D., Walsh, M. M., & Rosenbaum, D. A. (2009). Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: A commentary on Spivey, Grosjean, and Knoblich (2005). *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 588–594.
- van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1450–1464.
- Welsh, T. N., & Elliott, D. (2004). Movement trajectories in the presence of a distracting stimulus: Evidence for a response activation model of selective reaching. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 57(6), 1031–1057.
- Williams, J. R., Brady, T. F., & Störmer, V. S. (2022). Guidance of attention by working memory is a matter of representational fidelity. *Journal of Experimental Psychology: Human Perception and Performance*, 48(3), 202–231.
- Xie, W., & Zhang, W. (2017). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition*, 166, 73–83.
- Xie, W., Cappelletto, M., Park, H. B., Deldin, P., Chan, R. C., & Zhang, W. (2018). Schizotypy is associated with reduced mnemonic precision in visual working memory. *Schizophrenia Research*, 193, 91–97.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235.
- Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, 12(1), 24–25.
- Zhou, C., Lorist, M. M., & Mathôt, S. (2020). Concurrent guidance of attention by multiple working memory items: Behavioral and computational evidence. *Attention, Perception, & Psychophysics*, 82(6), 2950–2962.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.