UNIVERSITY OF CALIFORNIA
RIVERSIDE

Accuracy of Working Memory Representation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Hyung-Bum Park

September 2022

Dissertation Committee:
Dr. Weiwei Zhang, Chairperson
Dr. David Rosenbaum
Dr. Megan Peters

The Dissertation of Hyung-Bum Park is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to show my gratitude to everyone and every moment during my graduate program. The Ph.D. program to me was a journey with cycles of exploration and exploitation. Even when it is dark, in the solid belief in where our destination is, by taking comfort in the fact that we are (still) on the right path, I could enjoy every step I made with tremendous help from amazing people in the department. It could never be possible without unconditioned support from my advisor, Dr. Weiwei Zhang. I am pleased to be able to officially document a story here that my cohort used to call him as *Park's papa bear*. During my early coursework I was told several times by the cohort that whenever I made a speech in Dr. Zhang's seminars, he showed a gentle and proud smile as if saying, *yeah, that's my boy*. He gave me the greatest support with patience in my professional development and showed me what a desirable graduate advisor as well as an outstanding researcher would be like. His insightful criticisms and constructive advice taught me how to pursue perfection in research.

I am deeply grateful to my committee members, Dr. David Rosenbaum and Dr. Megan Peters. With their special expertise on important fractions of my dissertation study, they provided insightful suggestions with undoubted support on my work during the whole time in the program, including my qualifying exam. I would also like to thank my lab members, former and present times, Lilian Azer, Li Yang, Mack (Tianye) Ma, Dr. Bo Yeong Won, and Dr. Ilenia Salsano, as well as Dr. Zane (Weizhen) Xie and Dr. Marcus Cappiello. We made a great team and I owe a debt of gratitude for all the encouragement from each one of you.

Another special thanks to my fellow graduate students, Dr. Mahsa Alizadeh Shalchy and Dr. Esteban Sebastian Lelo de Larrea-Mancera. All of us international graduate students successfully ended up becoming proud survivors. I will always remember our good time back at the *International Cog House*. And my friend, Dr. Sang-Keun Yoo, for the entire time since we first met at the international student orientation at the start of our journey, I am indebted to you for all the support and the countless fun times we had together.

Lastly and most importantly, I would love to express great appreciation to my parents and family in South Korea for their continuous encouragement and prayers for me. And for Shinhae, it was my best luck to have you, and have been through all the hurdles and joys together in the last years. You fulfilled my world of love-and-science and taught this geek how to be a better person. I am looking forward to supporting you back on your academic journey and having celebrations together in the future.

ABSTRACT OF THE DISSERTATION

Accuracy of Working Memory Representation

by

Hyung-Bum Park

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, September 2022
Dr. Weiwei Zhang, Chairperson

The human mind represents the world by temporarily holding information in working memory that constructs an internal map of the external environment. This active information storage system does not simply produce veridical recordings of the moment-by-moment experience. Instead, it is rather a reconstructive process that possesses inherent, manifold limitations. A central question is thus how the erroneous nature of working memory can be best characterized. In this dissertation, I use behavioral and computational modeling methods to propose a *shift* in the central tendency of the internal representation as a psychologically valid aspect of working memory that is independent of other known limits such as capacity or resolution. In Chapter 1, I review general working memory literature and our current understanding of the representational limits, then introduce a shift component as another source of working memory quality. In

Chapter 2, I present computational modeling work with a set of data simulations and parameter recovery. I elucidate the necessity of the shift parameter in explaining working memory errors and how robust modeling of shift can be achieved using a hierarchical Bayesian approach. I then present a collection of empirical studies in Chapters 3 and 4 that experimentally show how this shift parameter can be utilized in theoretically important hypothesis testing. In Chapter 5, I extended the idea to mouse cursor trajectory data where the curved trajectory pattern can be operationalized as a moment-by-moment readout of representational shifts. In Chapters 6, I present an original empirical study that shows how the trajectory-based analyses of representational shift provide precise evidence crucial for constraining competing theoretical debates. Together, this dissertation provides a theoretical framework and a quantitative model to validate memory accuracy (i.e., representational appearance), manifested as the central tendency of the recall error distribution over sensory feature space, as another major source of working memory quality.

Table of Contents

PART I. 1

Chapter 1. General Introduction 1

1.1. Working Memory 1
1.2. Dissecting the Erroneous Nature of Working Memory 2
1.3. Accuracy, Bias, and Representational Shift 5
1.4. Dissertation Overview 11
References 14
Figures 18

Chapter 2. Modeling Systematic Bias of Working Memory Errors 22

2.1. Introduction 22
2.2. Hierarchical Bayesian Modeling of μ Shift 22
2.3. Simulation and Parameter Recovery 24
References 27
Figures 28

PART II. 31

Chapter 3. μ Shift Testing a Discrete Component in Visual Working Memory Encoding 31

3.1. Chapter Abstract 31
3.2. Introduction 31
3.3. General Method 37
3.3.1. Participants 37
3.3.2. Stimuli and procedure 37
3.3.3. Data analysis 40
3.4. Experiment 1 43
3.4.1. Results and Discussion 44
3.5. Experiment 2 48
3.6. General Discussion 51
References 57
Figures 63

Chapter 4. μ Shift Testing the Influence of Long-Term Memory on Working Memory 69
4.1. Chapter Abstract 69
4.2. Introduction 69
4.3. Method 72
4.3.1. Participants 72
4.3.2. Stimuli 73
4.3.3. Procedure 73
4.3.4. Data Analysis 75
4.4. Results 78
4.5. Discussion 79
References 83
Figures 87

PART III. 93

Chapter 5. μ Shift Testing Between-Item Interference Arising from Variance of Precision Across Multiple Working Memory Representations 95
5.1. Chapter Abstract 95
5.2. Introduction 95
5.3. Method 99
5.3.1. Data Reanalysis and Rationales 99
5.4. Results 104
5.4.1. Endpoint Bias and Its Impact on Between-Item Interference 104
5.4.2. Within-Trial Trajectory Predicts Relative Precision of VWM Items 106
5.5. Discussion 109
References 113
Figures 117

Chapter 6. μ Shift for Testing a Unit of Attentional Guidance by the Contents of Working Memory 124
6.1. Chapter Abstract 124
6.2. Introduction 125
6.3. Method 130
6.3.1. Participants 130
6.3.2. Stimuli and Procedure 131
6.3.3. Data Analysis 133

6.4. Results 138
6.5. Discussion 143
References 149
Figures 153
Chapter 7. General Discussion 162
7.1. Establishing a Psychological Meaning of μ Shift 162
7.2. Future Research Directions 164
References 168

List of Figures

Figure 1. *Change detection and continuous recall paradigms of visual working memory tasks.* 18

Figure 2. *A probabilistic mixture model of working memory recall error distribution.* 19

Figure 3. *Underestimation of precision by a model that does not account for bias.* 20

Figure 4. *Data simulation with μ shift and recovered without μ shift.* 21

Figure 5. *A graphical illustration of the hierarchical Bayesian mixture model for working memory recall errors.* 28

Figure 6. *Parameter recovery results under the maximum-likelihood fitting method* 29

Figure 7. *Posterior probability densities of the condition effect captured by the population-level μ shift parameter at each trial numbers.* 30

Figure 8. *Stimuli and procedure for the delayed estimation tasks in Experiment 1A, 1B, and 2.* 63

Figure 10. *Experiment 1A results of recall errors and the hierarchical Bayesian posteriors of the μ parameters.* 65

Figure 11. *Experiment 1B results in subjective judgments and recall errors.* 66

Figure 12. *Simulation of whole-report recall errors generated by the 2VM+U model and the hypothesis-driven predictions for the relative error difference.* 67

Figure 13. *Experiment 2 results of recall errors and the hierarchical Bayesian 2VM+U model posteriors of the μ parameters.* 68

Figure 14. *Model and the hypotheses.* 87

Figure 15. *Stimuli and procedure of the change detection task.* 88

Figure 16. *A summary of the face change detection task performance.* 89

Figure 17. *Overall celebrity familiarity ratings and the change detection performance.* 90

Figure 18. *Population-level posteriors for bias (μ) and imprecision (σ) and the group difference effects.* 91

Figure 19. *Celebrity-by-celebrity results of the population-level posteriors for bias (μ) and imprecision (σ) measures.*.....92

Figure 20. *Task procedure and primary results from Hao and colleague (2021).* 117

Figure 21. *Differences in the analytic method for the mouse trajectory data from the original and the present study.* 118

Figure 22. *Overall mouse trajectories of the item-selection stage.* 119

Figure 23. *Comparison of the mouse trajectory results between endpoint correction methods.* 120

Figure 24. *Trial-categorization results*.....121

Figure 25. *The posteriors of the hierarchical Bayesian mixture model parameters.*.....122

Figure 26. *Predictions for mouse trajectory based on the single-item template (SIT) and the multiple-item template (MIT) hypotheses.* 153

Figure 27. *Stimuli and procedure of the working memory and perceptual matching dual task.*.....154

Figure 28. *Example of destination-vector transformation of mouse trajectory.* 155

Figure 29. *Imprecision of the reports for perceptual matching and working memory recall tasks.*.....156

Figure 30. *Mean reaction times and mouse onset latencies from perceptual matching task.* 157

Figure 31. *Perceptual matching mouse trajectory results.*.....158

Figure 32. *Analytic approach for testing the predictions for the circular trajectory bias distributions based on the single-item template hypothesis (H_{SIT}) and multiple-item hypothesis (H_{MIT}).* 159

Figure 33. *The hierarchical Bayesian posteriors of μ shift parameters.*.....160

Figure 34. *Trial-categorization results for each memory set size condition.* 161

PART I.

Chapter 1. General Introduction

1.1. Working Memory

In a complex environment, working memory (WM) serves as an online workplace for information temporally stored and manipulated. This short-term retention of information is central to cognition. One of the key characteristics of WM is that it is greatly limited in the capacity that only about 2-4 objects can be actively stored (Cowan, 2001; Luck & Vogel, 2013; Zhang & Luck, 2008) and also limited in processing speed that how much information can be encoded from perception to more durable form as memory representation at a moment (Chun & Potter, 1995; Huang et al., 2007; Vogel et al., 2006).

These intrinsic limits gave rise to the vast literature to investigate the nature of individual differences in WM capacity as well as specific mechanisms of how such limited capacity operates over various cognitive processes. Studies have shown that a large amount of variance in individual cognitive functions including general intelligence or emotional regulation, and cognitive deficits in clinical population can be attributed to individuals' WM capacity (Fukuda et al., 2010; Kane & Engle, 2002; Unsworth & Engle, 2007) and processing speed (Fuller et al., 2005; Kail & Salthouse, 1994).

While theories of WM agree on its capacity limit, they largely differ by whether the unit of these limits is discrete item-based or continuous resource-based. Specifically, the prominent "slot" models of WM postulate a fixed number of objects can be stored as discrete object-based representations (Luck & Vogel, 1997; Park et al., 2017; Zhang &

Luck, 2008), whereas the "resources" models argue that a fixed amount of resources is flexibly distributed across the items that need to be remembered concurrently (Bays et al., 2009; van den Berg et al., 2012). A longstanding debate on these contradictory accounts has brought great attention to the continuous nature of representation *quality* beyond the discrete item-based *quantity*.

1.2. Dissecting the Erroneous Nature of Working Memory

How to characterize and precisely measure the erroneous nature of WM has been of major interest in the field of cognitive psychology. Conventionally and even in recent studies, estimates of WM errors are obtained based on accuracy in the task where a set of discrete choices are allowed (e.g., Yes or No). For instance, in a *change detection* task (Luck & Vogel, 1997), one of the popular experimental paradigms to test visual WM (Figure 1A), two arrays of simple visual objects are sequentially presented as a memory and a test array, with a second of temporal separation in between. Participants in this task are required to detect whether the two arrays are the same or whether one of the remembered objects has a feature change in the test array, by responding "*Same*" or "*Different*". This discrete response method provided a simple calculation of the proportion correct and could be further converted to the number of successfully retained items among a total number of items presented to be remembered (i.e., the set size). For example, Cowan's K (Cowan, 2001) is a frequently used estimate of WM capacity, calculated by, $K = (\mathrm{HR} + \mathrm{CR} - 1) * \textit{set size}$, where HR and CR represent hit and correct rejection rates, respectively.

While these estimations of the aggregated-level task performance are useful in providing summary statistics of WM capacity, the aggregation of individual trials neglects important information in the variability across trials. For example, the binary responses can only tell us whether a test for a certain WM item was successful or not. That means, these measures have an inherent assumption that WM storage follows the all-or-none states, therefore the correct and incorrect responses are attributed to the success or failure of the WM process. However, successfully retained internal representations can continuously vary in signal strength (e.g., *d*-prime in signal detection model) that can be conceptualized as *mnemonic precision*. The variation in precision can vary from item to item as well as trial to trial, just as the aggregate-level estimates vary from individual to individual.

To overcome the limitations of the discrete responses, recent studies in visual WM utilize a continuous estimation method (Wilken & Ma, 2004; Zhang & Luck, 2008) in the memory recall paradigm instead of the change-detection, discrimination, or recognition task. A typical continuous recall task procedure is depicted in Figure 1B. The memory and maintenance periods are similar to other delayed match-to-sample tasks. At the test, on the contrary, probe items are presented with a surrounding color wheel. A target item among the probes can be indicated by a thicker outline of the placeholder. Participants are instructed to recall the color of the target item and choose the best match color on the color wheel by mouse clicking. WM recall errors can be calculated as the angular distance, in degree or radius, between the presented memory color and the reported color from the recall. The continuous recall errors over the repeated trials form a

bell-shaped distribution centered around zero representing the true color of the probe target. Because the stimuli and responses are bound to a circular space, the two ends of the error distribution can stretch toward negative π or positive π.

From the circular error distribution, the quantity (i.e., *capacity*) and the quality (i.e. *precision*) of WM can be independently estimated based on the state of mnemonic evidence that derived each recall response. For example, if the probed color is successfully retained in WM, the subject can make a memory-based response for the color that best matches the summary statistics of the internal representation. With an assumption that internal WM representation is noisy and probabilistic, the spread of the error distribution should represent the overall quality of the subject's WM. On the other hand, however, if the test probe is not stored at all, there is no internal evidence the recall response can map on. The subject should then eventually make a guess, and pick a random color among all possible color palettes as they are equally likely. These random guesses over trials will form a uniform distribution.

A measurement model by Zhang and Luck (2008) describes the observed error distribution and measures these two independent WM components, capacity and precision. The model is formulated as a mixture of the memory-based bell-shaped distribution and the guessing-based uniform distribution (Figure 2). The memory-based error distribution for the stimuli with a circular feature space can be formalized by the von Mises distribution, a Gaussian analog in circular space, described as:

$$p(x \mid \mu) = e^{\kappa \cos(x-\mu)} \Big/ 2\pi I_0(\kappa)$$

where μ and κ are the mean and the concentration parameters representing the shift in the retained feature relative to the external stimulus and the mnemonic precision. A mixture of the von Mises distribution and the uniform distribution can then be described as:

$$p(x|\mu) = (1 - g) \cdot VM(x; \mu, \kappa) + g(1/2\pi)$$

where g is the height of the uniform probability distribution representing the proportion of guessing-based responses.

The operationalization of mnemonic precision in the separation of capacity in the probabilistic mixture model has been extensively applied to study testing the experimental effect on the quality aspect of WM representation (Xie & Zhang, 2016; Zhang & Luck, 2009, 2011) and its individual differences (Gold et al., 2010; Xie et al., 2018).

1.3. Accuracy, Bias, and Representational Shift

Meanwhile, theories of WM have largely ignored another significant aspect of representational quality, *accuracy*, which is conceptually independent of precision. In a general sense of measurement, accuracy and precision reflect different types of noise. Accuracy reflects how close the measurements are to the true value thus describing a systematic error, whereas precision reflects how dispersed the measurements are from each other thus describing a random error (Walther & Moore, 2005). Statistically, inaccuracy and imprecision can be substituted with bias and variability that can be descriptively measured by the mean and the standard deviation, respectively, though the concept of accuracy sometimes embraces both bias and precision (i.e., accuracy can be identical to precision when bias is absent; Hellmann & Fowler, 1999).

These two concepts can be applied for further dissecting different types of WM errors. For instance, WM representations can be accurate-but-imprecise or inaccurate-but-precise (in Chapter 4, I present an empirical example of this dissociation as a result of experimental manipulation). In this dissertation, I argue that accuracy of WM reports should also be operationalized as a psychologically valid measure of representational quality. Although the accuracy and precision are mathematically independent, a measurement model that does not account for this bias type of error can lead to the underestimation of mnemonic precision. Figure 3 demonstrates this scenario. When a given dataset contains a systematic overall bias toward a certain direction, fixing μ shift parameter to zero constant results in a large overestimation of the width of the distribution (i.e., underestimation of precision). In fact, the probabilistic mixture model aforementioned already implemented the μ parameter reflecting the overall mean of the von Mises distribution. However, the μ shift is often excluded from the set of free parameters but instead fixed at zero as constant, in particular when there is no particular reason to predict any systematic shift of the underlying WM representation.

On the theoretical ground, the neglect of central tendency in characterizing WM quality originates from a particular assumption that WM errors are normally distributed to reflect underlying noisy neural representation centered around the true sensory value. It is indeed plausible considering how visual information is represented by population neural activity. In particular, WM representations are thought to be maintained in the form of sustained spiking activity in populations of neurons distributed across the brain regions including the prefrontal cortex, parietal cortex, and sensory cortex (Christophel et

al., 2017; Murray et al., 2017; Miller et al. 1996). Because these groups of neurons are selectively tuned to their preferred sensory feature values, their population neural activity should peak at the true sensory input and persist its activity over the delay period to service goal-directed maintenance of WM. This principle of input-matched representational central tendency is supported by decoding from the pattern of neural activity in the early visual cortex (Ester et al., 2013). This population coding scheme has successfully accounted for the precision of behavioral WM reports by gradual attenuation and/or diffusion (i.e., broadening) of population spiking activity (Barrouillet et al., 2007; Bays, 2014; Sprague et al., 2014). That is, variability of population neural activity and imprecision of the retained WM representation are closely linked each other.

On the other hand, however, the assumption of the correspondence between external stimulus and the central tendency of internal representation could be an oversimplification of the dynamic interaction of WM with other mental operations. Memory is rather a reconstructive process than passively carrying a veridical copy of sensory experience. A growing number of empirical studies support the reconstructive nature of WM by showing that the contents of WM can influence the perception of upcoming sensory inputs (Teng & Kravitz, 2019) or create adaptive distortions within the current representations based on the statistical regularities (Brady & Alvarez, 2011; Fischer & Whitney, 2014) or to avoid interference between memoranda (Chunharas et al., 2019). For example, when multiple visual items are remembered WM reports of the individual items are often biased toward the ensemble average of the group of the stimuli (i.e., attraction bias). Several studies attributed this type of attraction bias to Bayesian

inference (Brady & Alvarez, 2011; Huang & Sekuler, 2010), proposing that a prior assumption for the distribution of stimulus is utilized to interpret a new stimulus. Consequently, the resulting posteriors of internal representation can sometimes contain an attractive bias toward the prior mean. Importantly, note that in all these studies, distortions in WM representations are manifested as central tendency biases of the error distributions and thus could be conceptualized as *representational shift*.

Representational shift refers to a correspondence between the internal representation as measured and the original stimulus. In neuroscience literature, such representation shift is thought to originate from a *drift* of population neural activities (Panichello et al., 2019). The delay-period neural activity supporting WM maintenance is commonly modeled as recurrent synaptic networks that generate stable attractors along the continuum of sensory feature values. The continuous attractor network models of WM have shown that accumulated noise in prefrontal neurons over the delay period results in a systematic *drift*, but not *diffusion*, of population activity toward the nearest bump attractors or even collisions between attractors, which give rise to a shift in the represented feature value (Compte et al., 2000; Wei et al., 2012; Wimmer et al., 2014). In such case, even if the underlying neural representation of a given memory item is sharply maintained (i.e., inaccurate-but-precise), the error in memory report across trials can have a large variance, which could be solely captured by the *sd* parameter of the probabilistic mixture model and interpreted as mnemonic imprecision, instead of representational shift.

Together, these arguments cast a doubt on the estimate of WM precision measured by between-trial variability in errors. That is, when aggregated across trials

with different stimuli hence different representational profiles, a collection of representational bias that determines the direction and magnitude of error for a given stimulus will be interpreted as variability in WM recalls. Consistent with this notion, Wimmer and colleagues (2014) showed that a trial-wise drift of neural population activity in the monkey prefrontal cortex during the WM delay period could accurately predict the direction of behavioral error. This raises a theoretically important question: How much variability in WM errors is then reflective of overall mnemonic precision? To what extent or whether the representational bias irrespective of precision explains the overall quality of WM?

To quantify such artifact in estimating precision, I present a set of data simulation and parameter recovery results. Sets of 100,000 data points were simulated from the normal distribution, $SimData \sim Normal(\mu, SD_{base})$, with different combinations of varying means and standard deviations. Specifically, five different standard deviations were set as SD_{base} at 10, 15, 20, 25, and 30, and for each case, six different mean μ values were assigned proportional to its SD_{base} value, from 0% to 100% in steps of 20% increment. For example, for SD_{base} of 20, μ was set to 0, 4, 8, 12, 16, and 20, respectively. Note that the μ was implemented in an additive manner, thus representing a systematic mean shift.

A total of 5-by-6 matrix of 100,000 samplings from normal distribution were then again fitted by another normal distribution that does not have μ as a free parameter but instead fixed its mean at zero. Since we know the original variability in the simulated data, SD_{base}, we can evaluate the extent of the overestimated variability for each dataset by comparing the estimated standard deviation with SD_{base}. Figure 4A illustrates the

normal probability density functions of the simulated and recovered data, and Figure 4B shows a summary of the estimated *SD* as a function of SD_{base} and μ values used in data simulation. Systematic overestimation of variability was identified, and its magnitude was proportional to the systematic shift in μ.

To further establish the exact contribution of the mean shift to the overall variability of data, I next conduct another data simulation similar to the previous one but critically different in the implementation of a shift to the mean. The same number of data points were sampled from the normal distribution with the same set of SD_{base} for the variability term. Importantly, the mean of the sampling distribution was derived from another normal distribution having its mean fixed at zero but instead having variability of SD_{μ} for the same amount of μ values in the previous simulation. In other words, data was simulated by, $SimData \sim Norm(Norm(0, SD_{\mu}), SD_{base})$. Therefore, this represents the case of *random drift* of trial-wise representational shift, instead of a systematic shift in overall trials.

Interestingly, exactly the same results were found in the overestimation of standard deviation from the previous simulation of systematic shift (Figure 4B). Furthermore, this leads to a conclusion of the *root-sum-squared* relationship between mnemonic precision and representational shift, together accounts for the between-trial variability, described as:

$$SD_{between\text{-}trial} = \sqrt{SD_{precision}{}^2 + SD_{shift}{}^2}$$

This conclusion raises a question that the conventional operationalization of WM precision by the width of error distribution might be convoluted with non-systematic

shifts of the representation central tendency. This in part resonates with the discussion of probabilistic representation in perception literature (Block, 2018; Rahnev, 2017).

1.4. Dissertation Overview

In this dissertation, I propose an integrative account for the psychological validity of representational shift in modeling quality of WM representations. I present a collection of empirical studies and computational modeling methods that focus on the experimental effect of the representational appearance, manifested as the central tendency of the continuous reports of WM over sensory feature space. In the following Chapter 2, I introduce hierarchical Bayesian modeling of a systematic shift in WM error distribution while dissociating it from other components of errors such as capacity and resolution. I present data simulation and parameter recovery tests to highlight the reliability and the precision of parameter estimation under the hierarchical Bayesian method compared to the non-hierarchical model with the conventional maximum-likelihood estimation.

In Part II, I present empirical studies that provide examples of how the shift parameter of WM errors can be utilized as a critical estimate for theoretically important hypothesis testing. Specifically, the Chapter 3 study introduces how the extent of representational shift can be coupled as a key experimental manipulation for testing the temporal dynamics of WM encoding. Thorough computational modeling for the profile of shift in recall error distributions for dynamically changing features of memory items revealed a temporarily discrete component in the WM encoding process for multiple items. Chapter 4 experimentally shows how the representational shift and precision independently and distinctively capture the influence of prior knowledge of celebrity

faces in LTM on how the celebrity look-alike faces are represented in WM. The results suggest that the stimulus familiarity effect can manifest as a more precise but biased representation toward the prototypical stimulus.

In the next Part III, I extended the empirical application of representational shift to mouse cursor trajectory data, specifically the pattern of trajectory bias during decision-making and motor response. Throughout two empirical studies, I show several advances in the mouse trajectory analyses combined with a specialized task paradigm for testing representational shifts in WM. As a part of the advances, I develop a novel method, destination vector transformation, mapping the raw cartesian coordinates of mouse cursor positions onto underlying feature space on a moment-by-moment basis. Specifically, the Chapter 5 study suggests how mouse trajectory data, specifically focusing on its biased curvature toward a certain item over the others, accounts for between-item interference in WM arising from variable precision across representations. In the following Chapter 6, I present the utilization of the circular trajectory shift measures for testing theoretical debates in WM research. The study tested the unit of attentional guidance by the contents of WM, whether multiple or only a single representation held in can guide attention toward the matching visual information in the subsequent, independent perceptual task. A closer examination of the mouse trajectory pattern during a perceptual matching task in the middle of the WM maintenance period revealed convincing evidence for the single-item template account. The results further support a hierarchical structure of WM representations via variable attentional accessibility.

Together, this dissertation presents a theoretical framework and a quantitative model to validate a shift in WM errors over sensory feature space as another major source of mnemonic quality. The collective evidence provided here further elucidates that the external sensory information is not processed in a veridical way but rather subject to interactions across various levels of mnemonic trace about one's internal representation of the world.

References

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570-585.

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), 7-7.

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience, 34*(10), 3632-3645.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic?. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1755). 1-10.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384-392.

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

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental psychology: Human Perception and Performance*, *21*(1), 109-127.

Chunharas, C., Rademaker, R. L., Brady, T., & Serences, J. (2019). Adaptive memory distortion in visual working memory. *PsyArXiv*.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral cortex*, *10*(9), 910-923.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–185.

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

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738-743.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity *Psychonomic Bulletin & Review*, *17*(5), 673-679.

Fuller, R. L., Luck, S. J., McMahon, R. P., & Gold, J. M. (2005). Working memory consolidation is abnormally slow in schizophrenia. *Journal of Abnormal Psychology*, *114*(2), 279-290.

Gold, J. M., Hahn, B., Zhang, W. W., Robinson, B. M., Kappenman, E. S., Beck, V. M., & Luck, S. J. (2010). Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. *Archives of General Psychiatry*, *67*(6), 570-577.

Hellmann, J. J., & Fowler, G. W. (1999). Bias, precision, and accuracy of four measures of species richness. *Ecological Applications*, *9*(3), 824-834

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*(5839), 823-825.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision*, *10*(2), 24.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199-225.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637-671.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391–400.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154-5167.

Murray, J. D., Jaramillo, J., & Wang, X. J. (2017). Working memory and decision-making in a frontoparietal circuit model. *Journal of Neuroscience*, *37*(50), 12167-12186.

Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Communications*, *10*(1), 1-11.

Park, H. B., Zhang, W., & Hyun, J. S. (2017). Dissociating models of visual working memory by reaction-time distribution analysis. *Acta Psychologica*, *173*, 21-31.

Rahnev, D. (2017). The case against full probability distributions in perceptual decision making. *BioRxiv*, 108944.

Sprague, T. C., Ester, E. F., & Serences, J. T. (2014). Reconstructions of information in visual spatial working memory degrade with memory load. *Current Biology, 24*(18), 2174-2180.

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827-836.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104-132.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780-8785.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436-1451.

Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, *28*(6), 815-829.

Wei, Z., Wang, X. J., & Wang, D. H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *Journal of Neuroscience*, *32*(33), 11228-11240.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120-1135.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), 431-439.

Xie, W., & Zhang, W. (2016). Negative emotion boosts quality of visual working memory representation. *Emotion, 16*(5), 760-774.

Xie, W., Cappiello, M., Park, H.-B., Deldin, P., Chan, R. C. K., & 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.

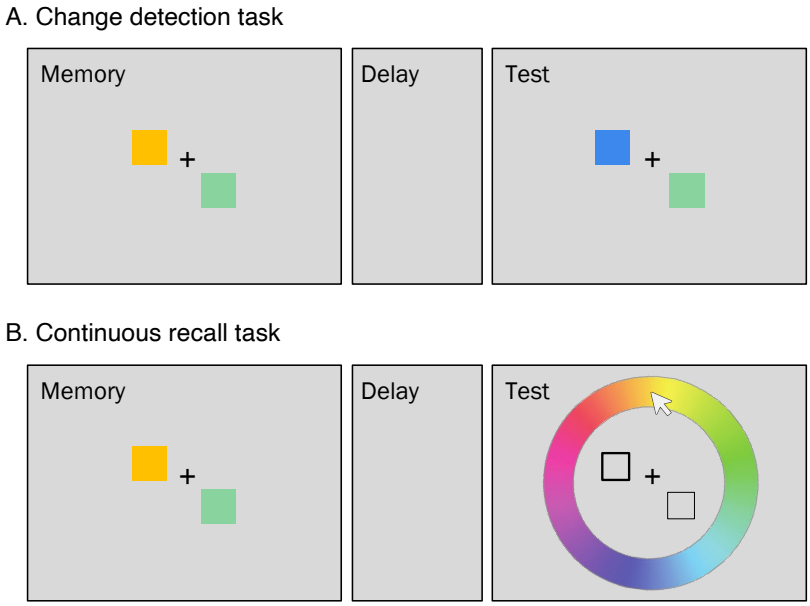
Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*(4), 423-428.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*(11), 1434-1441.

Figures

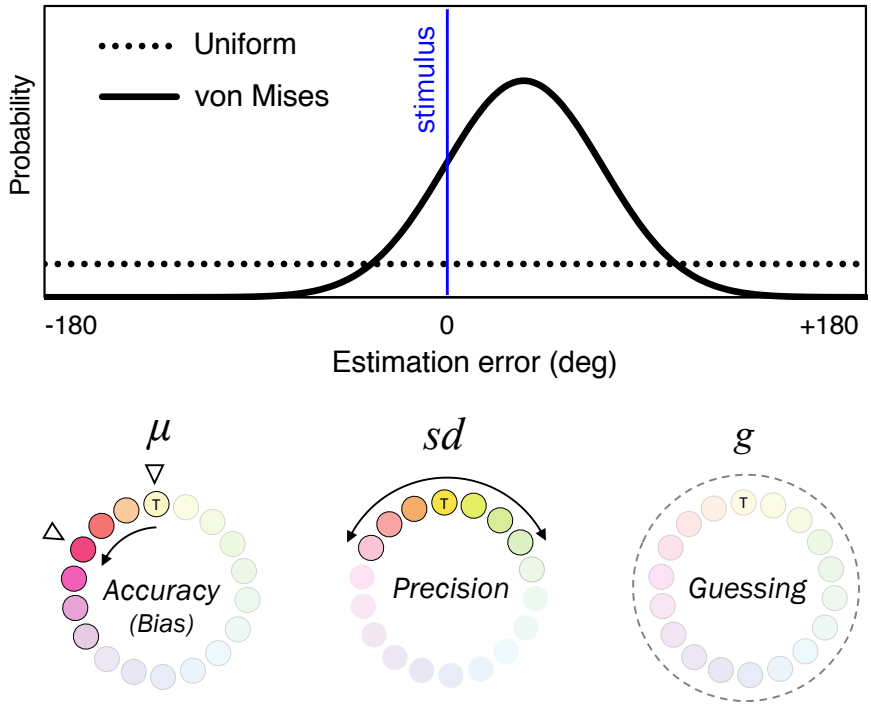
Figure 1.

Change detection and continuous recall paradigms of visual working memory tasks.



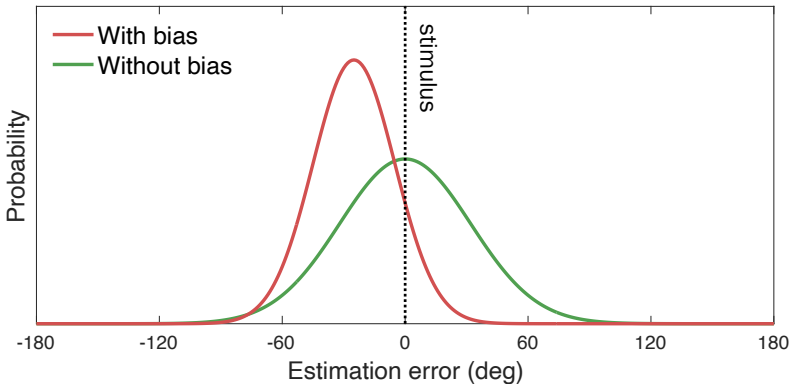
Note. (A) Change detection task. Participants are presented with two arrays consisting of simple feature objects. A short blank display is presented as delay interval between the memory and test array. At the test array, participants need to report whether the test probes are the same or whether the color of one of the objects is different from the remembered objects. (B) In the continuous recall task, the test display contains a color wheel and placeholders at the corresponding locations of the memory objects. A target probe is indicated by a thicker outline of the placeholder. Participants need to recall the color of the target probe and choose the best match color on the color wheel by a mouse clicking.

Figure 2.
A probabilistic mixture model of working memory recall error distribution.



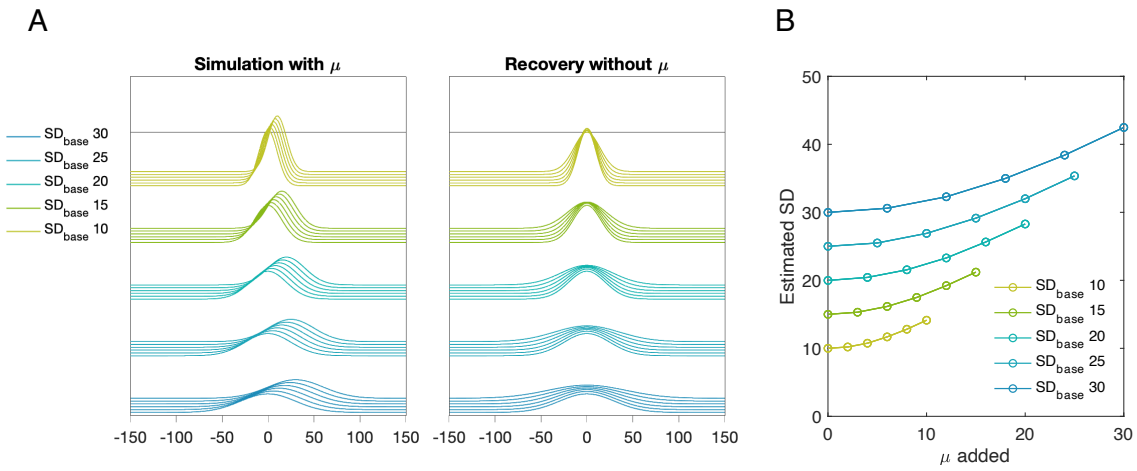
Note. The model is formulated as a mixture of the uniform and von Mises distributions, each captures guessing and memory-based responses, respectively. The von Mises distribution has μ and *sd* parameters as the mean and standard deviation of the distribution (the directionality in each type of error from the true value, *T*, is illustrated with an arrow), reflecting accuracy and precision aspects of memory quality. The height of the uniform distribution is parameterized by guess rate, *g*.

Figure 3.
Underestimation of precision by a model that does not account for bias.



Note. When the overall errors exhibit a systematic left-ward bias (red curve), a measurement model that does not account for bias can result in underestimation of precision to describe the data given (i.e., overestimation of the width of the distribution; green curve).

Figure 4.
Data simulation with μ shift and recovered without μ shift.



Note. (A) The normal probability density functions of the simulated and recovered data (B) A summary of the estimated *SD* as a function of SD_{base} and μ values added to simulate data.

Chapter 2. Modeling Systematic Bias of Working Memory Errors

2.1. Introduction

In previous Chapter 1, I present a brief overview of how the errors from continuous estimation WM task can be modeled to capture distinct constructs of representations, the qualitative and quantitative aspects. A vast of recent studies have utilized this measurement model to characterize the mnemonic quality and to evaluate individual differences in WM function. However, overall memory quality could also manifest as the accuracy of representation (i.e., the correspondence between internal representations and stimuli). This chapter will elucidate more details of the practices modeling this representational shift from error distribution.

The representational shift can manifest as the systematic changes in the central peak location of the error distribution, which can be represented as the third parameter μ in Zhang and Luck (2008) probabilistic mixture model. The direction and magnitude of the μ shift parameter could be central for testing theoretically important hypotheses. However, its numerical effect often tends to be very subtle, thus vulnerable to multiple sources of variability such that the central tendency may vary from participant to participant, trial to trial, or even item to item. This could fail to reach statistical significance even when the effect is true.

2.2. Hierarchical Bayesian Modeling of μ Shift

For the numerically and statistically weak effect in μ shift, the hierarchical Bayesian approach could be a useful solution to estimate the true effect over the noisy surface of a dataset. The hierarchical Bayesian modeling combining hierarchical models

and Bayesian parameter estimation techniques offers substantial advantages over the non-hierarchical models under the maximum-likelihood or least-square estimation (Oberauer et al., 2017). Specifically for the advantage in parameter estimation, multiple levels of variability (e.g., population-level, subject-level, and condition-level) are simultaneously accounted for by pooling the data from all participants and conditions (Lee & Wagenmakers, 2014). Throughout informing and updating the parameter estimates across hierarchical structures in data, the posteriors of parameters can be shrunk toward the population mean (Shiffrin et al., 2008). This results in a probabilistic density of posteriors over all plausible parameter values, instead of a point estimate of the best-fitting parameter value under the maximum-likelihood method, giving the uncertainty of parameter estimation.

A graphical illustration of the hierarchical Bayesian mixture model is shown in Figure 5. For the population parameters of the mixture model, *shift*, *precision*, and *proportion of guessing*, the main effects are estimated in a general linear model, sampled from the normal distribution where the posterior mean is defined by the sum of fixed effect (i.e., condition effect) and random effect (i.e., individual effect), and the variability term describes the interaction between the individual and condition effects (Rouder et al., 2014). For example, the population μ shift parameter is drawn from,

$$\mu_{ij} \sim Normal(\alpha_i + \beta_j, \delta^2)$$

$$\alpha_i \sim Normal(0, \varepsilon)$$

$$\varepsilon \sim Gamma(1, 1); \quad *in\ radius$$

$$\beta_j \sim Normal(0, pi/18)$$

where α_i and β_j are the i_{th} subject effect and the j_{th} condition effect, respectively, and δ^2 describes their interaction effect.

2.3. Simulation and Parameter Recovery

To test the robustness of parameter estimation for the μ shift under hierarchical Bayesian mixture modeling, here I present a set of data simulations and parameter recovery tests. The 3-parameter Zhang and Luck (2008) mixture model was used to simulate and fit the simulated data. Specifically, assuming that a researcher conducted a study where the difference between two experimental conditions is predicted only by a 3° shift in the reported errors but not in precision and guessing rate. The preset parameters for these two conditions are as follow: Condition 1 (μ: 0°, sd: 20°, p_{guess}: 0.2) and Condition 2 (μ: 3°, sd: 20°, p_{guess}: 0.2).

For the maximum-likelihood fitting method, a large sample size of 5,000 data points (i.e., subjects) was simulated from the model to minimize sampling noise, separately for each condition and each six different trial numbers: 30, 50, 100, 200, 500, and 1,000 trials. These simulated data are then again fitted by the same model to estimate the best-fitting parameter values, separately for each data cell. To evaluate the parameter recovery reliability, I focus on the variability of the recovered parameter values across 5,000 subjects. Figure 6A shows a decreasing pattern in the standard deviations of the best-fit parameters over increasing trial numbers. More importantly, Figure 6B shows the condition effect (i.e., condition 2 – condition 1) on each parameter by fitting curves to the normal probability density functions, as a function of trial numbers. Among the six populations of the condition effects on the μ shift centered around 3°, a simulation set of

500 trials per condition revealed a 95% confidence interval between -0.05° and 5.98°. This indicates that about 500 trials per condition are required to reliably capture the predicted effect of 3° on μ shift.

For the hierarchical Bayesian method, I simulated only 20 data points from the model separately for each condition and three numbers of trials: 50, 100, and 200 trials. This huge reduction in the number of samples was to show the power of hierarchical Bayesian parameter estimation pooling information from all data points from different samples (i.e., subjects) for estimating population-level parameters while accounting for individual differences. For Bayesian parameter estimation, I took a total of 10,000 samples after 10,000 warm-ups from four Markov Chain Monte Carlo chains. For the Bayesian posteriors of model parameters, statistical inference can be made based on the 95% highest density interval ($HDI_{95\%}$). The $HDI_{95\%}$ tells the parameter values covered by 95% of the posterior density, thus a conceptual analog of the frequentist confidence interval ($CI_{95\%}$).

Figure 7 plots the posterior probabilities of the population-level μ shift parameters for the condition effect at the corresponding trial numbers, 50, 100, and 200 trials. The results show a drastic reduction in the number of trials required to recover the known difference of 3° μ shift between conditions, with a reasonable number of 20 subjects. Specifically, even at 50 trials per condition per subject, the lower boundary of the $HDI_{95\%}$ [0.7°, 4.9°] did not cross over zero, indicating a credible difference between the two conditions. The $HDI_{95\%}$ range becomes narrower as the trial number increases, $HDI_{95\%}$

[1.4°, 4.3°] and $HDI_{95\%}$ [1.8°, 3.9°] for 100 trials and 200 trials per condition per subject, respectively.

The present data simulation and parameter recovery confirms the primary advantage of hierarchical modeling in estimating population-level parameter estimates. It is particularly useful in the context of studies using continuous WM recall tasks and when more than two experimental conditions need to be tested or only a limited number of subjects are available (e.g., special populations such as experts or clinical patients in a specific domain). In the following chapters, I introduce empirical examples of how this hierarchical Bayesian mixture modeling for the representational shift component in WM errors.

References

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H. Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5), 11.

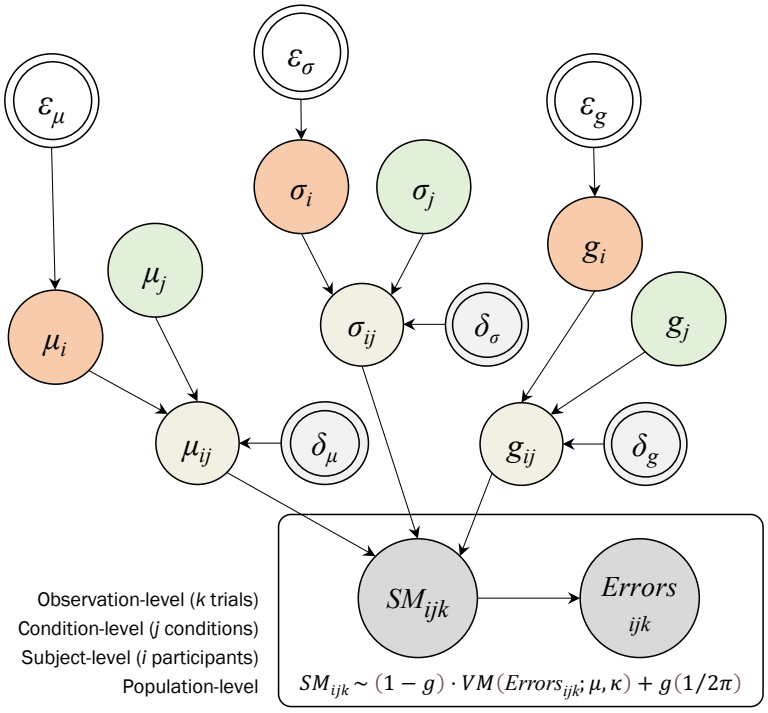
Rouder, J. N., Morey, R. D., & Pratte, M. S. (2014). Bayesian hierarchical models. In *New Handbook of Mathematical Psychology, volume. 1: Measurement and Methodology*. Cambridge University Press.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248-1284.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.

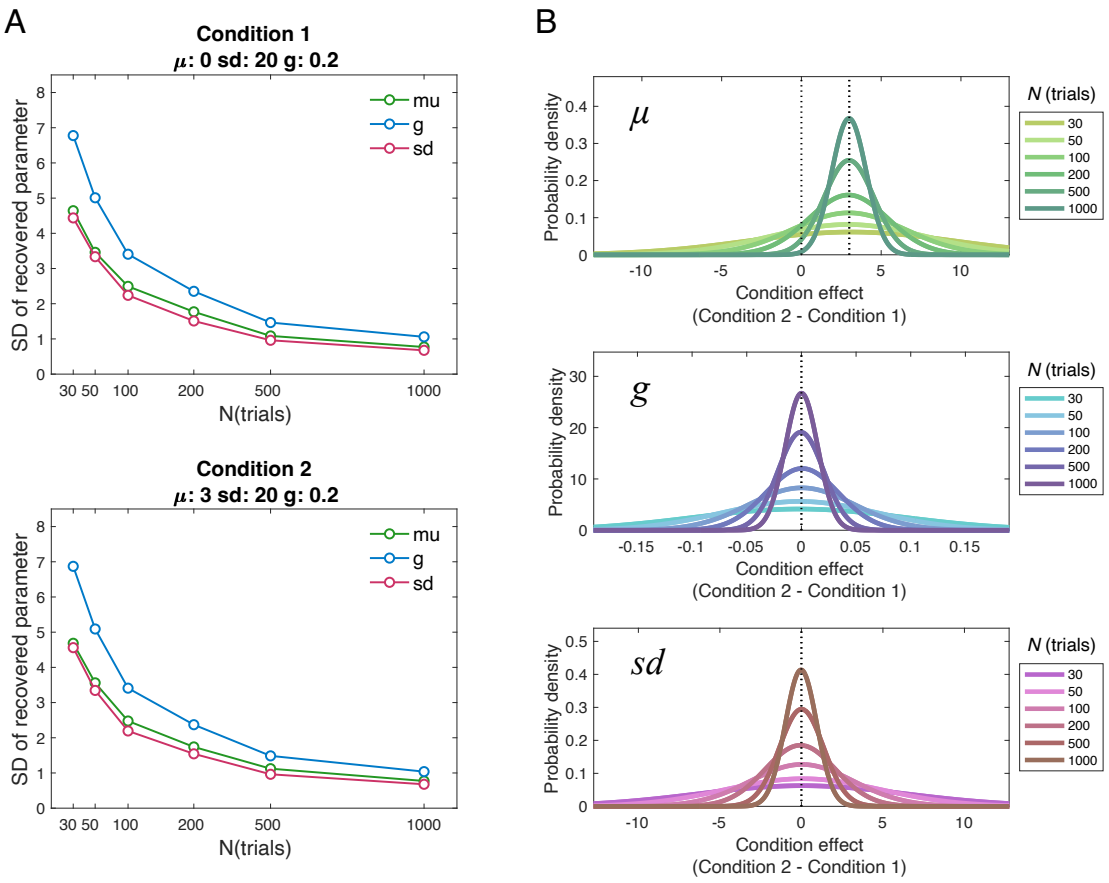
Figures

Figure 5.
A graphical illustration of the hierarchical Bayesian mixture model for working memory recall errors.



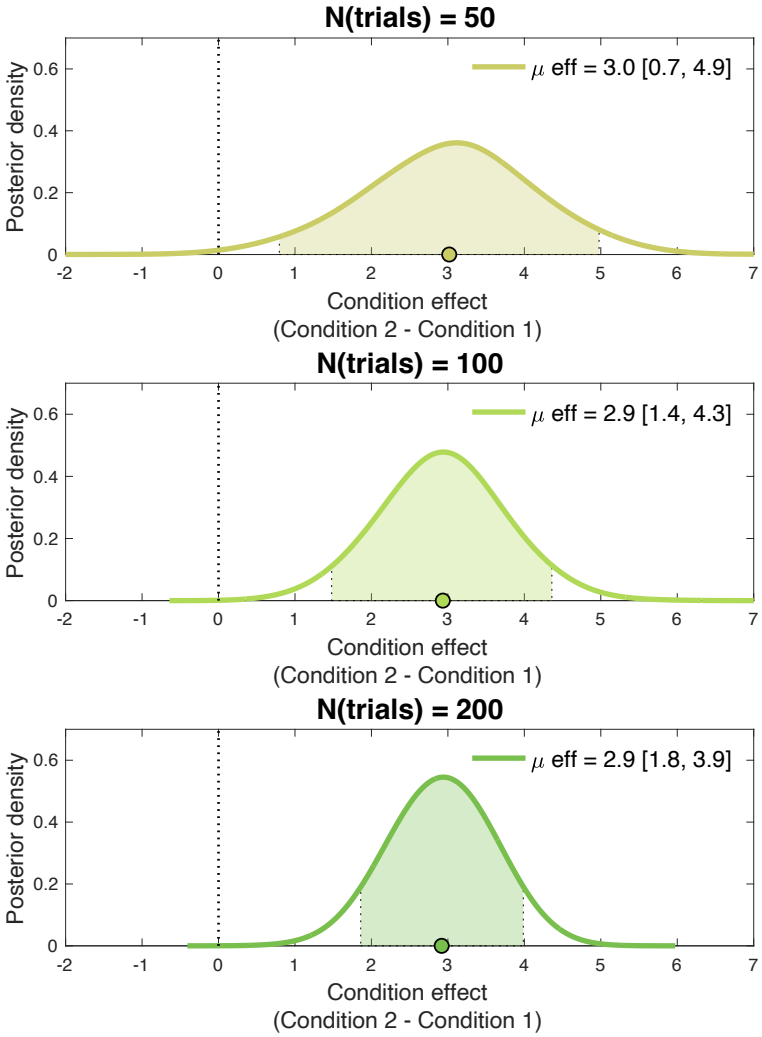
Note. The model contains different parameters for hierarchies following given data structure, observation-level (gray-filled circles), subject-level (orange), condition-level (green), and population-level (ivory). Parameters in double-border circles describe variabilities of the target parameters. The population-level parameters μ_{ij}, κ_{ij}, and g_{ij} are fitted to the error from each participant, condition, and trial.

Figure 6.
Parameter recovery results under the maximum-likelihood fitting method



Note. (A) The standard deviations of the best-fit parameters of the probabilistic mixture models for condition 1 and condition 2, as a function of trial numbers. The larger values indicate greater variability in the recovered parameter values. (B) The population condition effect (condition 2 – condition 1) illustrated by the normal probability density function fit curves based on the mean and standard deviation of the recovered parameter values.

Figure 7.
Posterior probability densities of the condition effect captured by the population-level μ shift parameter at each trial numbers.



Note. The curves are the kernel density fits of the parameter posteriors, and the shaded areas are the range of the 95% highest density interval ($HDI_{95\%}$). The legends summarize the mean and $HDI_{95\%}$ of the condition effect on the μ shift parameter.

PART II.

Chapter 3. μ Shift Testing a Discrete Component in Visual Working Memory Encoding

3.1. Chapter Abstract

Working memory (WM) as a central bottleneck in cognition has primarily been attributed to its well-known storage limit. However, relatively little is known about a processing limit in the initial memory encoding stage, whose temporal characteristics may constrain various cognitive processes. The present study has developed a novel method of dynamic stimulus presentation with hierarchical Bayesian modeling to quantitatively estimate WM encoding speed. Specifically, two memory items for the delayed-estimation task continuously changed color hues in perceptually unnoticeable steps. Across three experiments, the recall errors systematically shifted toward the direction of color change, providing a proxy measure of encoding speed. These shifts can be accounted for by a mixture of two distributions of encoding time with credible separations in-between, providing strong evidence for a discrete component in WM encoding of multiple items.

3.2. Introduction

In natural vision, objects' appearance often changes subtly due to naturally occurring events in the environment (e.g., changes in luminance from shadow of other objects) and eye movements or body movements of the observers. Consequently, adaptive behaviors often rely on the observers' ability to rapidly encode and update relevant information in a consciously accessible form that can last more than few hundred

milliseconds in the absence of sensory inputs (e.g., during saccadic eye movements or blinks), as well as to resist interference and memory decay (Irwin, 1991; Sperling, 1960). Longstanding research on this active memory representational system, visual working memory (VWM) has shown that it is severely limited in storage capacity (Cowan, 2001; Luck & Vogel, 1997; Park et al., 2017; Zhang & Luck, 2008, 2009a, 2011), mnemonic precision (Bays & Husain, 2008), or strength (Schurgin et al., 2020). Despite the effort on characterizing these representational limitations of VWM and its relation to cognition (Fukuda et al., 2010; Kane & Engle, 2002; Oberauer et al., 2005; Unsworth & Engle, 2007), only a handful of studies assessed the broader contributions of the encoding process transferring fragile perceptual representations into durable mnemonic representations in VWM (Vogel et al., 2006; Cappiello & Zhang, 2016; Cotton & Ricker, 2022).

VWM encoding process is functionally important for cognition in several ways (Xie & Zhang, 2022). First, VWM encoding is highly limited in its processing speed (Jolicœur & Dell'Acqua, 1998; Vogel et al., 2006). That is, the amount of information that can be encoded into VWM at a given time is limited, and this may in turn limit the processing speed of other cognitive processes, such as the speed of conscious perception tested in the attentional blink literature (Chun & Potter, 1995; Vogel & Luck, 2002). VWM encoding speed as a limiting factor for conscious perception may even lead to continuous sensory inputs being perceptually encoded into discrete chunks (Vanrullen et al., 2007). Beyond individual processes, VWM encoding speed may be associated with the overall mental processing speed (Kail & Salthouse, 1994), and consequently mediate

the relationship between working memory capacity and higher cognition such as fluid intelligence (Fry & Hale, 1996; Salthouse, 1991). Relatedly, impairments in VWM encoding are often observed in various clinical populations. For instance, schizophrenia patients often exhibit larger attentional blink effects (Cheung et al., 2002) and reduced VWM capacity (Gold et al., 2006), potentially due to slower VWM encoding (Fuller et al., 2005).

Second, VWM encoding speed may underlie some behavioral effects on VWM capacity. For instance, behavioral (Jannati et al., 2015) and electrophysiological (Linke et al., 2011) measures of VWM encoding speed seem to be closely related to VWM storage capacity. In addition, some recent reports suggest that VWM capacity boost by stimulus familiarity (Sørensen & Kyllingsbæk, 2012; Shen et al., 2018; Xie & Zhang, 2017a) can be accounted for by the familiarity effect on VWM encoding speed (Blalock, 2015). Further studies have suggested that familiarity-induced capacity effect may only manifest when VWM encoding is interrupted by short encoding time or use of masking stimuli (Xie & Zhang, 2017b; Xie & Zhang, 2018).

Given the functional significance of VWM encoding and its impact on cognition, it is pivotal to establish reliable experimental measures of VWM encoding speed. The encoding speed/rate limit can manifest, and thus could be behaviorally assessed, in two different ways. First, if the amount of encoding time is short, then VWM task performance will deteriorate as the memory set size increases (Becker et al., 2013; Miller et al., 2014). Reversely, at a given memory set size, VWM task performance will increase continuously with encoding time until it reaches a plateau (Vogel et al., 2006).

To measure the encoding rate from these aspects, Vogel and colleagues (2006) designed a masked encoding procedure in a change-detection task. In the task, a brief presentation of memory items was followed by an additional array of pattern masks at various time intervals (i.e., the stimulus onset asynchrony (SOA) between the onset of the memory display and the mask array). These backward masking stimuli largely interrupt the VWM encoding process such that any perceptual representations that have yet been encoded will be masked out, whereas the already encoded memory representations can survive. By systematically manipulating the memory encoding time with the memory-mask SOAs, they found a continuous increase in change-detection accuracy with the encoding time, but only until it reached an asymptote level. In addition, the SOA at the asymptote performance and the asymptote performance itself, measured as Cowan's *K* (Cowan, 2001), increased with memory set sizes, suggesting more representations are encoded into VWM with longer encoding time. A linear function adequately captured this pattern at set size four, yielding a slope of 50 ms with an intercept of 60 ms, indicating that VWM encoding rate is approximately 50 ms per stimulus and the initial formation of perceptual representation is about 60 ms.

These methods have provided simple empirical measures of VWM encoding speed that can capture various experimental effects (Xie et al., 2022) and individual differences in healthy and clinical populations (Habekost & Starrfelt, 2009). Nonetheless, these behavioral measures may have some caveats originated from particular assumptions and methodological constraints (also see Xie & Zhang, 2018). For example, these methods assume that the backward pattern masking effectively terminates VWM

encoding with comparable effectiveness across memory set sizes and memory-mask SOAs. An additional assumption is the homogeneity in VWM encoding rate across memory items (i.e., linear or exponential function; also see Bundesen, 1990). However, these assumptions have not been systematically tested. Also, these methods require artificial interruption of VWM encoding with backward masking at various SOAs, rendering it less ecologically valid given that artificial backward masking is uncommon in natural vision. Lastly, the masking stimuli could introduce an additional task demand for active inhibition to avoid replacing memory representations with the masking stimuli, similar to object substitution masking (Enns, 2004). Given that inhibition of distractors may be more crucial than VWM encoding in limiting the processing speed of rapidly presented stimuli (Dux & Harris, 2007), it is necessary to ensure that the estimates of VWM encoding rate are not confounded by the speed of active inhibition.

These potential issues with the current behavioral methods for assessing VWM encoding call for a new method that does not rely on the artificial backward masking procedure and those unnecessary assumptions. In the present study, we have thus developed a novel task procedure of presenting two dynamically changing memory colors to assess VWM encoding time based on the recalled color values using the delayed estimation paradigm (Zhang & Luck, 2008). Although this dynamic memory feature change is perceptually subtle and unnoticeable (see Experiment 1B results), the encoded memory color will vary based on the encoding speed. Consequently, the temporal dynamics of VWM encoding process could manifest as and thus be behaviorally measured by the magnitude of systematic shifts in the remembered feature values from

the memory onset, which can in turn be converted to the encoding speed measure. In other words, a slower VWM encoding will lead to larger shift in the retained color representation toward the direction of the memory color rotation.

Furthermore, we utilized this novel measure of VWM encoding speed to test the temporal dynamics in the VWM encoding of multiple items. Despite our subjective experience of representing the external world in a continuous manner, growing evidence point out a discrete sampling mechanism in attention, perception, and internal attention directed to working memory representations (Barrouillet & Camos, 2012; Landau, 2018; Pomper & Ansorge, 2021; VanRullen, 2016). We sought out a discrete component within the VWM encoding of two colors by comparing two descriptive models (see Method for detail) generated from different theoretical perspectives, using a hierarchical Bayesian method.

The *discrete encoding* hypothesis predicts dissociable memory encoding time (starting and/or completion) for the two memory items that manifests as two separable central peaks of the recall error distributions under our experimental manipulation of dynamically changing memory colors. Specifically, the trial-by-trial variability in the circular shift measure could be better accounted for by the model with two central peaks (i.e., the to-be-recalled memory item was either encoded earlier or later than the other item). In contrast, the *concurrent encoding* hypothesis predicts overlapping encoding time for two changing colors that subsequently manifests as a single central peak of the recall error distribution. According to this hypothesis, the seriality predicted by the discrete encoding hypothesis could be nothing but a consequence of trial-by-trial

variability of parallel encoding completion time (Townsend, 1990). In this case, the variable magnitudes of shift on the continuous feature space should originate from *random variation between trials*, while *comparable between the encoded items*. Consequently, the error distribution would be better described by the model with a single central peak.

Together, the present study tested whether or how VWM encoding of two dynamically changing colors manifest as the shifts in the recall error distributions for one of the two remembered colors (Experiment 1A) or both colors (Experiment 2). To rule out an alternative account that these effects are potentially driven by strategical adjustment of the recalled color, Experiment 1B tested whether the dynamic color rotation was perceptually noticeable and subsequently associated with the recall performance.

3.3. General Method

3.3.1. Participants

Twenty-seven (19.0 ± 1.2 years old, 3 males), 33 (19.7 ± 1.1 years old, 17 males), and 25 (20.0 ± 1.2 years old, 9 males) college students at University of California, Riverside participated in Experiment 1A, Experiment 1B, and Experiment 2, respectively, for course credits. All participants provided informed-consent and reported to have normal or corrected-to-normal visual acuity and normal color vision.

3.3.2. Stimuli and Procedure

Stimuli were presented on an LCD monitor with a refresh rate of 60 Hz and a dark grey background (6.1 cd/m^2) at a viewing distance of 57 cm. The memory colors were

randomly selected from a 360-degree color-wheel with a minimum circular distance of 24°. The color-wheel, with a radius of 8.2° and a thickness of 2.2°, consisted of 180 colors evenly distributed on a circle in CIELAB space (for details, see Zhang & Luck, 2008). All colors had equal luminance but varied in hue and slightly in saturation. As shown in Figure 8, for the delayed estimation task in Experiment 1, 1B, and 2, each trial started with an 800 ms fixation, followed by a memory array consisting of two colored squares (2.0° × 2.0° in visual angle). The colored squares were presented at locations randomly selected from a set of four equally spaced locations on an imaginary circle with a radius of 5.3° from the center.

Importantly, the memory colors continuously rotated by 2° at every screen refresh (16.7 ms) throughout the 167 ms of presentation duration, resulting in a total of 18° color rotation over nine steps (velocity of 0.12°/ms). With this color rotation rate, 1° color rotation in the color space corresponds to 8.33 ms in time. The directions of the color rotation were counterbalanced between clockwise (CW) and counterclockwise (CCW) across trials but stayed the same for the two memory colors on each trial.

In all experiments, participants were instructed to remember the two memory colors as accurate as possible over a 1,000 ms blank interval. A test array followed, containing two outline squares at the corresponding memory item locations and a color-wheel. One square was thicker than the other item, indicating the probed item that needs to be recalled. Participants reproduced the color of the probed item by clicking the mouse cursor on the best-matching color on the color-wheel. Accuracy was stressed over speed. Immediately after the recall response, a 500 ms online feedback was presented with a

cross and an arrow marking the color the participants picked and the actual color presented in the memory array, respectively. Each trial sequence was repeated 400 trials with a second of inter-trial interval.

In Experiment 1B, an additional perceptual task consisted of two questions was presented after the memory recall response in every trial, to assess whether the participants noticed the color rotation of the memory colors or not. The same color-wheel from the test array of the delayed estimation task stayed on the screen, as a reference for the participants. The first question of the perceptual task, presented at the display center, asked the participants whether they noticed any color changes during the presentation of the memory colors. Regardless of the Yes/No answer and its accuracy, the second question followed and asked whether the memory colors changed direction clockwise or counterclockwise on the color-wheel. Participants responded to each question by clicking on a Yes/No scale or CW/CCW color change scale with the computer mouse. Critically, unlike Experiment 1, the memory colors only rotated in the circular space on half of the trials, whereas remained static during memory presentation for the other half of the trials. The directions of the color change were again counterbalanced between CW and CCW. A total of 64 trials were collected for Experiment 1B.

Experiment 2 procedure was identical to the one in Experiment 1A except that the first recall test was followed by a 1-s blank display and then another recall test for the other item (i.e., the whole-report procedure). In correspondence, the thicker outline square as the recall cue switched from the first item to the other item for the second recall. To minimize a strategical recall of the non-target item at the first recall, a new response

color-wheel was presented for the second recall with a circular rotation of random degrees. Feedback was provided for both recalls as in Experiment 1A. Participants in Experiment 2 performed a total of 270 trials.

3.3.3. Data Analysis

Recall error was coded as the angular distance between the presented memory color and the reported color from the recall. Errors from CW and CCW trials were collapsed by reversing the sign of recall errors from CCW trials and consequently the directions of recall error for both conditions were identical, relative to the direction of the memory color rotation. That is, errors in positive values for both CW and CCW conditions are in the direction of the memory color rotation. Prior to this treatment, we identified that there was no systematic, statistically significant difference in the absolute circular mean of errors between CW and CCW trials in Experiment 1A, $t(26) = 1.50$, $p = .145$, Cohen's $d = 0.30$, Experiment 1B, $t(32) = 0.98$, $p = .334$, Cohen's $d = 0.17$, and Experiment 2 for the first recall, $t(24) = 1.07$, $p = .296$, Cohen's $d = 0.22$, and for the second recall, $t(24) = 0.27$, $p = .787$, Cohen's $d = 0.06$.

The recall errors were then modeled using a hierarchical Bayesian method, separately for two candidate models. First, to test the concurrent encoding hypothesis, *1VM+U* model, as an extended Zhang and Luck (2008) Mixture model, is mathematically a mixture of two distributions representing whether recall responses are based on noisy mnemonic evidence or random guesses (Figure 9A). The uniform distribution (U) captures recall responses with no mnemonic evidence (e.g., random guesses), represented by a single parameter, g, the height of the uniform probability distribution. The von

Mises distribution (*VM*; the circular analogue of the Gaussian distribution) captures recall responses resulting from noisy mnemonic representation with two parameters, mean (μ) representing the shift in the retained color hue relative to the first color of the memory color rotation and standard deviation (σ, converted from concentration κ) representing mnemonic imprecision. The shift and imprecision parameters manifest as the peak location and width of the bell-shaped von Mises distribution, respectively. In the present study, a systematic shift in the retained color toward the direction of color rotation captured by μ parameter is central for testing VWM encoding speed. Together, the *1VM+U* model is described by the following equation:

$$p(x|\mu) = (1 - g) \cdot VM(x; \mu, \kappa) + g(1/2\pi)$$

Second, *2VM+U* model is defined as the mixture of two von Mises distributions and the uniform distribution, testing our discrete encoding hypothesis (Figure 9B). The two von Mises functions were equally weighted in summation based on assumptions that each memory item is equally likely to be encoded earlier than the other, and more importantly, the chance that whether a test probe is the item that is encoded earlier or later is also equally likely on average across trials. The *2VM+U* model is described by the following equation:

$$p(x|\mu) = (1 - g) \cdot [0.5 \cdot VM_1(x;\ \mu_1, \kappa) + 0.5 \cdot VM_2(x;\ \mu_2, \kappa)] + g(1/2\pi)$$

, where VM_1 and VM_2 are two von Mises distributions sharing the same precision parameter κ. They are not arbitrarily ordered but only different by independent mean central tendency parameters, μ_1 and μ_2.

3.3.3.1. Hierarchical Bayesian Modeling

The two descriptive models are generated from and used to test the two competing theoretical accounts of VWM encoding. Critically, the magnitude of the separation of the two von Mises distributions, $\mu\, gap$ (μ_2 - μ_1), predicted by the discrete encoding hypothesis is expected to be numerically small, based on the previous literature (e.g., 20 to 100 ms; Vogel et al., 2006). This expected lag, in the context of the present experiment procedure, is equivalent to 2° to 12° gap between the peak locations of the error distributions, leading to two closely-overlapped distributions given the typical memory imprecision of about 20° standard deviation (Zhang & Luck, 2008). In addition, multiple sources of variability such as between-subject and between-trial variance can make it even more difficult to tease apart the two overlapping distributions, under the conventional maximum likelihood estimation method (Park & Zhang, 2019).

To address this issue, we adopted the hierarchical Bayesian model with MatlabStan (Stan Development Team, 2016), to assess the temporal characteristics of VWM encoding for two colors. In hierarchical Bayesian parameter estimation, multiple sources of variabilities are accounted for simultaneously through different levels of data structure (e.g., subject-by-subject and trial-by-trial variability) such that noisy parameter estimates can be shrunk toward the population mean (Shiffrin et al., 2008). As a result it provides posterior parameter estimates at population-level with a full range of possible values. We took a total of 12,000 samples after 12,000 warm-ups from four Markov Chain Monte Carlo chains. We chose reasonable to non-informative priors for all parameters to minimize biases due to the choice of priors, following existing literature that adopted hierarchical Bayesian modeling for continuous recall errors (Oberauer et al.,

2017; Park & Zhang, 2022). Model convergence assessed by $\hat{R}$ was found to be close or equal to 1.00 for all parameters (Gelman & Rubin, 1992).

The mean and the 95% credible interval (highest density interval, $HDI_{95\%}$) of the posterior distributions of the population-level parameters provide a point estimate and the strength/uncertainty of evidence, respectively. Statistical inference for the difference in parameter values (e.g., $\mu_2 - \mu_1$ representing the temporal lag between two encoded representations, μ_{gap}) can be made based on $HDI_{95\%}$. Specifically, if the lower and upper bounds of $HDI_{95\%}$ do not cross over zero, it can be considered as strong, credible evidence (Kruschke, 2014). Formal model comparison was performed to test the two competing models based on the widely applicable information criterion (WAIC; Vehtari et al., 2017; Watanabe, 2010), a robust measure of model fit for hierarchical models that takes the model complexity into account.

3.4. Experiment 1

In Experiment 1, we sought out to test our manipulation of memory color rotation in a continuous delayed estimation task. Unlike the presentation of a set of static memory items during encoding in conventional VWM recall tasks, the gradually changing physical appearance of the memory colors in the present study allows us to trace the temporal dynamics of VWM encoding. Accordingly, a central tendency measure of the error recall distribution (i.e., shift of the peak location) is predicted to reflect overall VWM encoding speed for two colors. A slower VWM encoding in this regard will be associated with greater shift of error distribution from the memory colors at their initial onset, toward the direction of the color rotation.

For the central peak of the recall error distributions, Experiment 1A further attempted to test the competing hypotheses about VWM encoding with a specific focus of the presence of temporal lag between encoding of two colors. A hierarchical Bayesian modeling and a formal model comparison using the likelihoods of the posteriors were conducted. If a discrete component is necessary for describing VWM encoding of multiple items, it will leave a dissociable temporal lag in the recall error distributions across probed items that are either encoded earlier or later than the other, under the experimental manipulation of the present study. In addition, Experiment 1B tested whether the participants could consciously notice the rotation of the memory colors and if their ability to detect a presence of color rotation strategically adjusted their recall responses accordingly.

3.4.1. Results and Discussion

3.4.1.1. Experiment 1A

The overall distribution of the recall errors showed a rightward shift toward the direction of the memory color rotation (Figure 10A). For the raw recall errors, the circular standard deviation was averaged at 30.3° [$CI_{95\%}$: 27.8°, 32.9°] across the participants. More importantly, the circular mean of the recall errors was averaged at +12.3° [$CI_{95\%}$: +11.6°, +12.9°] toward the direction of the memory color rotation, significantly deviated from the 0° initial color hue at the memory onset, $t(26) = 36.44$, $p < .001$, Cohen's $d = 7.15$.

Next, to test our hypotheses for the temporal relationship between VWM encoding of two dynamically rotating colors, a formal model comparison of the two

models, *1VM+U* and *2VM+U*, was performed using WAIC (widely applicable information criterion; Watanabe, 2010). The WAIC was calculated from the likelihood sets of the data given the estimated model parameters, integrated over all possible values from their posteriors (Oberauer et al., 2017). We found that WAIC largely favored the *2VM+U* model (WAIC: 85,379) over the *1VM+U* model (WAIC: 88,480). This indicates that the recall errors distributions for the two dynamically changing memory colors can be better described by the *2VM+U* model which postulates two separate central peaks for discrete encoding times for the two memory colors, than by the *1VM+U* model which postulates a single peak for the concurrent encoding of the two memory colors.

With the model comparison results, we then focused on the estimated parameters of the *2VM+U* model. The mean posterior of the population-level σ parameter (i.e., imprecision) was 18.9° [$HDI_{95\%}$: 18.5°, 19.4°] and *g* parameter (i.e., guessing rate) was 12.7% [$HDI_{95\%}$: 12.0%, 13.5%], consistent with the previous studies of VWM using static memory colors (e.g., Zhang & Luck, 2011). Of the primary interest, the posterior probability density function of the μ_1 and μ_2 parameters (Figure 10B) showed non-overlapping means and $HDIs_{95\%}$ (μ_1, 10.2° [$HDI_{95\%}$: 8.5°, 11.7°]; μ_2, 13.8° [$HDI_{95\%}$: 12.3°, 15.5°]). Therefore, the difference between the two shifts, μ_{gap} [$\mu_2 - \mu_1$], did not cross over zero, 3.6° [$HDI_{95\%}$: 0.4°, 6.7°], indicating a credible separability of the two discrete time points in VWM encoding for the two memory colors.

These μ_1 and μ_2 posteriors can then be converted to the temporal dimension. Given that 1° shift corresponds to 8.33 ms under our manipulation of color rotation velocity, two μs are equivalent to 85.0 ms [$HDI_{95\%}$: 70.8 ms, 97.8 ms] and 115.1 ms

[$HDI_{95\%}$: 102.1 ms, 129.2 ms], respectively, with a temporal gap of 30.1 ms [$HDI_{95\%}$: 3.6 ms, 55.7 ms] in between. Furthermore, the linear function for the two μs produced an intercept measure (i.e., μ_1 - μ_{gap}) at 55.5 ms [$HDI_{95\%}$: 12.3 ms, 96.7 ms], highly comparable with the previous report of 60 ms intercept which measures perceptual encoding time by Vogel and colleagues (2006).

3.4.1.2. Experiment 1B

Four out of the 33 participants repeatedly made the same response of either "Yes" or "No" for the first question across all trials and were therefore excluded from data analysis. Overall, the participants performed poorly on the detection of whether memory colors were changing or not during memory presentation (Figure 11A), with accuracy averaged at 50.0% [$CI_{95\%}$: 48.3%, 51.7%] and signal detection *d*-prime averaged at 0.02, [$CI_{95\%}$: -0.11, 0.14]. This performance was not significantly different from the chance-level (i.e., accuracy at 50% and *d*-prime at 0), *t*s(28) < 0.28, *p*s > .782, Cohen's *d*s < 0.05.

Across all trials, the ratio of subjective CW and CCW color change responses was comparable with each other, with a small preference for the CCW responses, 6.25% [$CI_{95\%}$: -3.41%, 15.9%] more than the CW responses, $t(28) = 1.27$, $p = .215$, Cohen's $d = 0.24$, indicating no systematic bias in the choice between CW and CCW responses. Critically, for the half of the trials when the memory color rotations actually happened, the participants' capability of detecting its direction remained at chance, 51.4% [47.0%, 55.8%], $t(28) = 0.\ 26$, $p = .540$, Cohen's $d = 0.12$. This near-chance performance was not significantly different regardless of whether they accurately detected the color change or not, $t(28) = 1.32$, $p = .100$, Cohen's $d = 0.25$.

To further evaluate whether or how the participants' detection of the memory color change affected their memory recall, we examined two circular descriptive statistics of the recall errors. A two-way repeated-measures analysis of variance (ANOVA) for the circular mean error (Figure 11B) as a function of presence/absence of the memory color rotation (change vs. no change) and participants' response type (Yes vs. No) revealed a significant main effect of color change, $F(1, 28) = 124.16$, $p < .001$, $\eta^2_p = .82$, whereas no significant main effect of the response type, $F(1, 28) = 0.14$, $p = .707$, $\eta^2_p = .01$, or the two-way interaction, $F(1, 28) = 0.86$, $p = .362$, $\eta^2_p = .03$. Planned *t*-tests between correct ("No") and incorrect responses ("Yes") for the color rotation absent condition, $t(28) = -0.51$, $p = .616$, Cohen's $d = -0.10$, and between correct ("Yes") and incorrect responses ("No") for color rotation present condition, $t(28) = 0.79$, $p = .436$, Cohen's $d = 0.15$, both failed to show significant differences.

Another two-way repeated-measures ANOVA was performed for the circular standard deviation of the recall errors (Figure 11B) with the same two factors yielded non-significant main effects of color change, $F(1, 28) = 1.59$, $p = .218$, $\eta^2_p = .05$, the response type, $F(1, 28) = 0.11$, $p = .739$, $\eta^2_p = .00$, and the two-way interaction, $F(1, 28) = 0.03$, $p = .872$, $\eta^2_p = .00$. Again, planned *t*-tests between correct and incorrect responses for the color rotation absent condition, $t(28) = 0.15$, $p = .884$, Cohen's $d = 0.03$, and the present condition, $t(28) = 0.32$, $p = .754$, Cohen's $d = 0.06$, produced nonsignificant effects. These results confirmed that our manipulation of memory color rotation was largely unnoticeable, and thus unlikely to elicit strategic responses such as biased recall responses toward the average of the dynamically changing memory colors.

Experiment 1A suggests that VWM encoding speed for two dynamically-changing colors can be measured by the extent of the circular shift in the remembered color toward the direction of memory color rotation. Furthermore, our Bayesian hierarchical modeling and formal model comparison revealed that the overall recall error distribution could be well explained by the *2VM+U* model postulating two credibly dissociated representations with the discrete central peaks, indicating a temporal lag between the discrete VWM encoding of two colors. Since the memory color rotation was largely unnoticeable and unrelated to the recall performance, tested in Experiment 1B, the systematic shift in the remembered color cannot be attributed to strategical adjustments to the recall responses.

3.5. Experiment 2

To perform a stronger test of the discrete-encoding hypothesis, Experiment 2 used a whole-report procedure in which both memory items were recalled in sequence. Critically, the relative difference between two recall errors (2^{nd} – 1^{st} recall) was modeled to test the validity of the discrete component in VWM encoding (see Figure 12 for simulation). For the discrete-encoding hypothesis, two recall responses should be mutually-exclusively sampled from different underlying mnemonic representations dissociated by the central peak locations, and consequently form a bi-modal distribution in the relative error difference between the two recalls. Contrarily, the concurrent-encoding hypothesis assumes that two recalls are *independently-and-identically* distributed in time, leading to a uni-modal distribution for the relative error difference with the central peak around zero.

3.5.1. Results and Discussion

Participants' average of the circular mean of recall errors was comparable between the 1st recall (11.9° [$CI_{95\%}$: 11.1°, 12.7°]) and the 2nd recall (12.0° [$CI_{95\%}$: 11.0°, 13.1°]), $t(24) = -0.22$, $p = .824$, Cohen's $d = -0.05$, and also comparable with the recall errors in Experiment 1A, ts < 0.36, ps > .719, Cohen's ds < 0.05. The circular mean for the difference in recall errors between the two recalls [2nd – 1st] was not statistically different from zero (0.3° [$CI_{95\%}$: -0.6°, 1.2°]), $t(24) = 0.65$, $p = .525$, Cohen's $d = 0.13$. For the circular standard deviation measure, the 2nd recall showed larger variabilities (37.4° [$CI_{95\%}$: 33.5°, 41.3°]) compared to the 1st recall (34.3° [$CI_{95\%}$: 30.6°, 38.0°]), $t(24) = 4.97$, $p < .001$, Cohen's $d = 1.02$, yielding 42.1° [$CI_{95\%}$: 39.0°, 45.1°] circular standard deviation of the error difference between the two recalls. The significant increase in variability of recall errors replicates the recall-order effect from the whole-report paradigm (Adam et al., 2017; Park & Zhang, 2022).

The overall distributions of the recall errors for the 1st recall, 2nd recall, and their difference [2nd – 1st] were well captured by the *2VM+U* model (Figure 13A-C). We compared the WAICs between the *1VM+U* and *2VM+U* models for the 1st and 2nd recall errors, separately. We again found that both sets of recall errors were better fitted by the *2VM+U* model ($WAIC_{1st}$: 50,502; $WAIC_{2nd}$: 52,736) than by the *1VM+U* model ($WAIC_{1st}$: 53,230; $WAIC_{2nd}$: 55,478). Critically, the relative difference between two recall errors [2nd – 1st recall] was also better accounted for by the *2VM+U* model ($WAIC_{diff}$: 56,718) than by the *1VM+U* model ($WAIC_{diff}$: 59,544). The model comparison results are consistent with our prediction that memory recalls for different

items were mutually-exclusively sampled from two mnemonic representations with separable central peak locations.

Consistent results were also obtained from the estimated posterior distributions of the *2VM+U* model parameters, specifically for the two shift parameters, μ_1 and μ_2, from the 1st and 2nd recalls (Figure 13D-E), and their difference μ_{gap} [$\mu_2 - \mu_1$], (Figure 13G-H). For the 1st recall, the posteriors of the μ_1, 10.0° [$HDI_{95\%}$: 8.3°, 11.8°], were credibly different from those of the μ_2, 14.0° [$HDI_{95\%}$: 12.2°, 15.7°], with non-overlapping HDIs, yielding a mean μ_{gap} [$\mu_2 - \mu_1$] at 3.9° [$HDI_{95\%}$: 0.4°, 7.1°]. Similarly, the 2nd recall errors resulted in a credible difference between the posteriors of the μ_1, 10.0° [$HDI_{95\%}$: 8.0°, 11.9°], and the μ_2, 14.5° [$HDI_{95\%}$: 12.7°, 16.9°], with μ_{gap} [$\mu_2 - \mu_1$] centered at 4.6° [$HDI_{95\%}$: 1.0°, 8.6°].

Critically, two distinct predictions were tested for the distribution of the relative error difference between two recalls, based on hypothetical principles of response generation for the 1st and 2nd recalls (see Figure 12). Specifically, if the separation of two μ parameters by μ_{gap} [$\mu_2 - \mu_1$] is psychologically meaningful, predicted by the discrete-encoding hypothesis, the 1st and 2nd recall responses should be mutually-exclusively sampled from the two von Mises distributions. That is, if the 1st recalled item is the one encoded earlier, then the 2nd recalled item has to be the one encoded later, and vice versa. Consequently, this hypothesis predicts that the distribution of the within-trial recall error differences would also have a similar bi-modal shape, with the μ_{gap} [$\mu_2 - \mu_1$] as the sum of the μ_{gap} from the 1st and 2nd recalls. On the contrary, if the separation of two μ parameters is merely a descriptive advantage of the *2VM+U* model over the *1VM+U*

model but nothing psychologically meaningful, each recall response should be independently and identically sampled from the same distribution. In other words, there should be no functional relationship between two recall errors. Consequently, the relative error differences should be centered at zero and uni-modally distributed, with no credible difference manifested in μ_{gap} [$\mu_2 - \mu_1$].

Consistent with the idea of discrete-encoding, the μ_1 and μ_2 posteriors estimated from the relative error difference between two recalls clearly showed two peaks (Figure 13F), contributing to the bi-modal shape of the error difference distribution. The posteriors of μ_1, -4.4° [$HDI_{95\%}$: -7.8°, -0.6°], were credibly separated from μ_2, 4.4° [$HDI_{95\%}$: 0.5°, 7.8°], with μ_{gap} [$\mu_2 - \mu_1$], 8.8° [$HDI_{95\%}$: 1.8°, 15.7°] nearly-doubled from those observed in the 1st and 2nd recalls, as predicted (Figure 13I). The 4.4° gap in the separation of two μs across recalls is equivalent to 36.7 ms.

These results confirm our hypothesis that two sequential recalls for different memory items within each trial were exclusively sampled from differentiable representational distributions with a credible separation between their peak locations, establishing psychological validity for the two μ parameters in the *2VM+U* model. Together, Experiment 2 provides strong evidence for the discrete component in VWM encoding for multiple items.

3.6. General Discussion

The present study investigated the temporal characteristics of VWM encoding with a novel method of estimating memory encoding time from the remembered color hues of continuously rotating memory colors. In three experiments, we found that

memory recalls were systematically shifted toward the direction of the memory color rotation, manifested as the shift of the central peak location of the recall error distribution. This shift in the circular color space can be mapped onto the temporal dimension based on the memory color rotation velocity, providing a proxy measure of VWM encoding speed. Our manipulation of dynamic color rotation was fairly unnoticeable to participants and their memory performance was unrelated to their subjective reporting of the presence of the color change (Experiment 1B).

Using this novel measure of VWM encoding speed, we further tested two competing hypotheses regarding the temporal relationship between VWM encoding of two memory colors by fitting the recall errors with two corresponding descriptive models under hierarchical Bayesian method. Our *1VM+U* and *2VM+U* models critically differed in the conceptualization of the temporal lag within encoding of two memory colors, each testing concurrent versus discrete encoding hypotheses, respectively. Formal model comparisons revealed that the *2VM+U* model outperformed the *1VM+U* model across experiments. In addition, the two central peak locations of the recall error distributions (μ_1 and μ_2) estimated by the hierarchical Bayesian *2VM+U* model were credibly separable each other with a statistically reliable gap in between (μ_{gap}) supported by non-overlapping boundaries of 95% HDIs. These results indicate an existence of the temporal lag between discrete VWM encoding of two colors.

Critically, Experiment 2 attempted to establish the validity of the separation of μ parameters for capturing discrete VWM encoding by modeling the relative error differences between two sequential recalls of both memory items with the whole-report

procedure (i.e., within-trial variability). The results confirmed a mutually exclusive sampling principle of the two recall responses from the two separable von Mises distributions. That is, if the 1st recall was made for the first encoded item, then the 2nd recall would be for the second encoded item. Conversely, if the 1st recall was made for the second encoded item, then the 2nd recall would be for the first encoded item. As a result, the distribution of the within-trial variability of the two recall errors can be better captured by the *2VM+U* model with its μ_{gap} size being equivalent to the summation of the μ_{gap} in the two recalls. Together, these findings provide strong evidence for a temporal lag within discrete encoding of the two memory colors into VWM.

It is important to note that this temporal gap by no means suggests that VWM encoding is a serial process in nature. Serial encoding means that discrete number of items are encoded into VWM in a strictly serial manner, so that only one item can be encoded into VWM at a given time (Townsend & Wenger, 2004). For example, Liu and Becker (2013) presented to-be-remembered stimuli either sequentially or simultaneously. The memory display duration for the simultaneous condition was fixed at a threshold interval for encoding a single item with a predetermined level of performance, whereas the presentation time for each one of the two sequential displays in the sequential condition was fixed at the threshold interval. They found the number of retained representations increased from the simultaneous condition to the sequential condition, whereas mnemonic precision of retained representations remained constant, which was taken as evidence for serial VWM encoding. A similar finding was reported previously by Zhang and Luck (2008) that the number, not precision, of encoded VWM

representations increased over a short period of time after memory array onset. However, they did not attribute the findings to serial VWM encoding, since a parallel process, especially a capacity-limited parallel process, can also mimic behavioral manifestation of the serial process (Townsend, 1990). That is, all memory items can be encoded into VWM simultaneously in parallel with a cost that VWM encoding slows down as the memory set size increases. The delay in parallel VWM encoding could result from decreases in available cognitive resources for each representation or increases in interference among representations as well as decisional noise (Oberauer & Lin, 2017). As such, the current finding of the lag between VWM encoding of two colors could also be consistent with capacity-limited parallel process.

The present study estimated about 4° lag between VWM encoding of two colors (averaged μ_{gap} across experiment 1A and 2), which is approximately 33.3 ms in our experimental context. This estimation is to some extent similar with the previous estimation of about 25 ms per item for alphanumeric stimuli (Gegenfurtner & Sperling, 1993; Shibuya & Bundesen, 1988) and 50 ms/item for four colors reported by Vogel and colleagues (2006). On more theoretical ground, the estimated temporal gap further corresponds to approximately 30 Hz cycle. This casts a natural speculation for the potential linkage to the gamma band oscillatory activities, considering the large literature on the functional role of gamma oscillatory brain activity in the encoding of attended sensory information and its maintenance in working memory (Bosman et al., 2012; Fries et al., 2001; Honkanen et al., 2015; Lundqvist et al., 2016; Miller et al., 2018). Therefore, the current findings raise several questions remained for future research to explore the

underling mechanisms for the discrete component in VWM encoding. A plausible account for the bottleneck that limits VWM encoding speed may be attentional in nature, given that VWM encoding is critically dependent on the availability and control of selective attention (Murray et al., 2011; Panichello & Buschman, 2021; Schmidt et al., 2002). That is, the speed of VWM encoding may be largely determined by the speed of focusing and shifting attention. Alternatively, the limit may arise from a bottleneck in conscious perception (Busch & VanRullen, 2010; Dehaene, 1993).

As our first step toward modeling the temporal dynamics of VWM encoding, some caveats of the present study need to be noted. First, it is analytically and experimentally challenging to generalize the *2VM+U* model to larger memory set sizes. Analytically, each addition memory item would require another VM component in the model, making it progressively more challenging to fit the limited amount of data reliably. Experimentally, larger memory set sizes would require extra presentation duration of the dynamically changing memory features values to accommodate the longer encoding time, leading to larger overall feature changes that may be perceptually noticeable. Second, the current findings need to be generalized across feature dimensions, given that temporal dynamics of VWM encoding may be different between different features (Liu & Becker, 2013; Miller et al., 2014). However, it may be difficult to implement the perceptually unnoticeable feature changes for some features (e.g., transparent motion with orientation). Third, the online feedback may have reduced the effect size because of the potential strategical adjustment of the recall responses due to the feedback. However, this effect would be minimal since color rotation of the memory items was not explicitly perceived,

and the CW and CCW color rotations were randomly mixed throughout the trials. Fourth, some alternative models of VWM, such as Variable-Precision (Van den Berg et al., 2012) and Target Confusability Competition (Schurgin et al., 2020) models, have not included μ as a free parameter to account for the current effects. Nonetheless, given that these models mainly focus on the alternative accounts of the capacity limit of VWM, which is conceptually and mathematically independent of the μ effects, similar findings are expected from these alternative models when they incorporate the μ parameter. Lastly, additional tests of the *2VM+U* model would be needed for independently validating the initial perceptual analysis time under linearity assumption (i.e., $\mu_1 - \mu_{gap}$, at ~60 ms) with experimental manipulation of perceptual noise (e.g., Zhang & Luck, 2009b).

In summary, with the combination of the novel experimental manipulation and the cognitive modeling approach, the present study provides strong evidence for discrete nature of VWM encoding for multiple items. Methodologically, it addresses some issues in the previous measure of VWM encoding and may provide better characterization of slower VWM encoding in various clinical populations using our method. Theoretically, it provides a novel perspective of VWM encoding and its underlying neural mechanisms.

References

Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79-97.

Barrouillet, P., & Camos, V. (2012). As time goes by: Temporal constraints in working memory. *Current Directions in Psychological Science, 21*, 413–419.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851-854.

Becker, M. W., Miller, J. R., & Liu, T. (2013). A severe capacity limit in the consolidation of orientation information into visual short-term memory. *Attention, Perception, & Psychophysics, 75*(3), 415-425.

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception, & Psychophysics, 77*(4), 1143-1158.

Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., & Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron, 75*(5), 875–888.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523-547.

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048-16053.

Cappiello, M., & Zhang, W. (2016). A dual-trace model for visual sensory memory. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1903-1922.

Cheung, V., Chen, E. Y., Chen, R. Y., Woo, M. F., & Yee, B. (2002). A comparison between schizophrenia patients and healthy controls on the expression of attentional blink in a rapid serial visual presentation (RSVP) paradigm. *Schizophrenia Bulletin, 28*(3), 443-458.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology, 21*(1), 109-127.

Cotton, K., & Ricker, T. J. (2022). Examining the relationship between working memory consolidation and long-term consolidation. *Psychonomic Bulletin & Review*, 1-24.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences, 24*(1), 154-176.

Dehaene, S. (1993). Temporal oscillations in human perception. *Psychological Science*, *4*(4), 264-270.

Dux, P. E., & Harris, I. M. (2007). On the failure of distractor inhibition in the attentional blink. *Psychonomic Bulletin and Review, 14*(4), 723-728.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*(12), 1321-1331.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*(1), 1-8.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*(4), 237-241.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*(5508), 1560-1563.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*(5), 673-679.

Fuller, R. L., Luck, S. J., McMahon, R. P., & Gold, J. M. (2005). Working memory consolidation is abnormally slow in schizophrenia. *Journal of Abnormal Psychology, 114*(2), 279-290.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 845-866.

Gold, J. M., Fuller, R. L., Robinson, B. M., McMahon, R. P., Braun, E. L., & Luck, S. J. (2006). Intact attentional control of working memory encoding in schizophrenia. *Journal of Abnormal Psychology, 115*(4), 658.

Habekost, T., & Starrfelt, R. (2009). Visual attention capacity: A review of TVA-based patient studies. *Scandinavian Journal of Psychology*, *50*(1), 23-32.

Hao, Y., Li, X., Zhang, H., & Ku, Y. (2021). Free-recall benefit, inhomogeneity and between-item interference in working memory. *Cognition*, *214*, 104739.

Honkanen, R., Rouhinen, S., Wang, S. H., Palva, J. M., & Palva, S. (2015). Gamma oscillations underlie the maintenance of feature-specific information and the contents of visual working memory. *Cerebral Cortex*, *25*(10), 3788-3801.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*(3), 420-456.

Jannati, A., McDonald, J. J., & Di Lollo, V. (2015). Individual differences in rate of encoding predict estimates of visual short-term memory capacity (K). *Canadian Journal of Experimental Psychology, 69*(2), 213-220.

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138-202.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199-225.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Landau, A. N. (2018). Neuroscience: A mechanism for rhythmic sampling in vision. *Current Biology*, *28*(15), 830-832.

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. *Neuropsychologia*, *49*(6), 1476-1486.

Liu, T., & Becker, M. W. (2013). Serial consolidation of orientation information into visual short-term memory. *Psychological Science, 24*(6), 1044-1050.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and beta bursts underlie working memory. *Neuron, 90*, 1–13.

Miller, J. R., Becker, M. W., & Liu, T. (2014). The bandwidth of consolidation into visual short-term memory depends on the visual feature. *Visual Cognition, 22*(7), 920-947.

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron, 100*(2), 463-475.

Murray, A. M., Nobre, A. C., & Stokes, M. G. (2011). Markers of preparatory attention predict visual short-term memory performance. *Neuropsychologia, 49*, 1458–1465.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 61-65.

Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21-59.

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H. Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5), 11:1-27.

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601-605.

Park, H. B., Zhang, W., & Hyun, J. S. (2017). Dissociating models of visual working memory by reaction-time distribution analysis. *Acta Psychologica, 173*, 21-31.

Park, H.-B., & Zhang, W. (2019). Hierarchical Bayesian modeling for testing representational shift in visual working memory. *Journal of Vision, 19*(10), 80.

Park, H., & Zhang, W. (2022). Trial-by-trial mouse trajectory predicts variance in precision across working memory representations: A critical reanalysis of Hao and colleagues (2021). *Psychonomic Bulletin & Review*. 29(4).

Pomper, U., & Ansorge, U. (2021). Theta-rhythmic oscillation of working memory performance. *Psychological Science*, *32*(11), 1801-1810.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*(3), 179-183.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception and Psychophysics, 64*(5), 754-763.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156-1172.

Shen, Z., Popov, V., Delahay, A. B., & Reder, L. M. (2018). Item strength affects working memory capacity. *Memory & Cognition, 46*(2), 204-215.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 591-600.

Sørensen, T. A., & Kyllingsbæk, S. (2012). Short-term storage capacity for visual objects depends on expertise. *Acta Psychologica, 140*(2), 158-163.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1-29.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*(1), 46-54.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*, 391–418.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary Memory. *Psychological Review, 114*(1), 104-132.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, 109*(22), 8780-8785.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences, 104*(49), 19204-19209.

VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences*, *20*(10), 723-735.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413-1432.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin and Review, 9*(4), 739–743.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436-1451.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571–3594.

Xie, W., & Zhang, W. (2017a). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition, 45*(4), 677–689.

Xie, W., & Zhang, W. (2017b). Familiarity speeds up visual short-term memory consolidation. *Journal of Experimental Psychology: Human Perception and Performance, 43*(6), 1207-1221.

Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience, 30*(1), 1-13.

Xie, W., Lu Sing, J. C., Martinez-Flores, A., & Zhang, W. (2022). Induced negative arousal modulates the speed of visual working memory consolidation. *Emotion*, 22, 179-197.

Xie, W. & Zhang, W. (2022). Pre-existing long-term memory facilitates the formation of visual short-term memory. In T. F. Brady & W. A. Bainbridge (Eds.). *Visual Memory* (pp. 84–104). New York, NY: Routledge.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.

Zhang, W., & Luck, S. J. (2009a). Sudden death and gradual decay in visual working memory. *Psychological Science, 20*(4), 423-428.

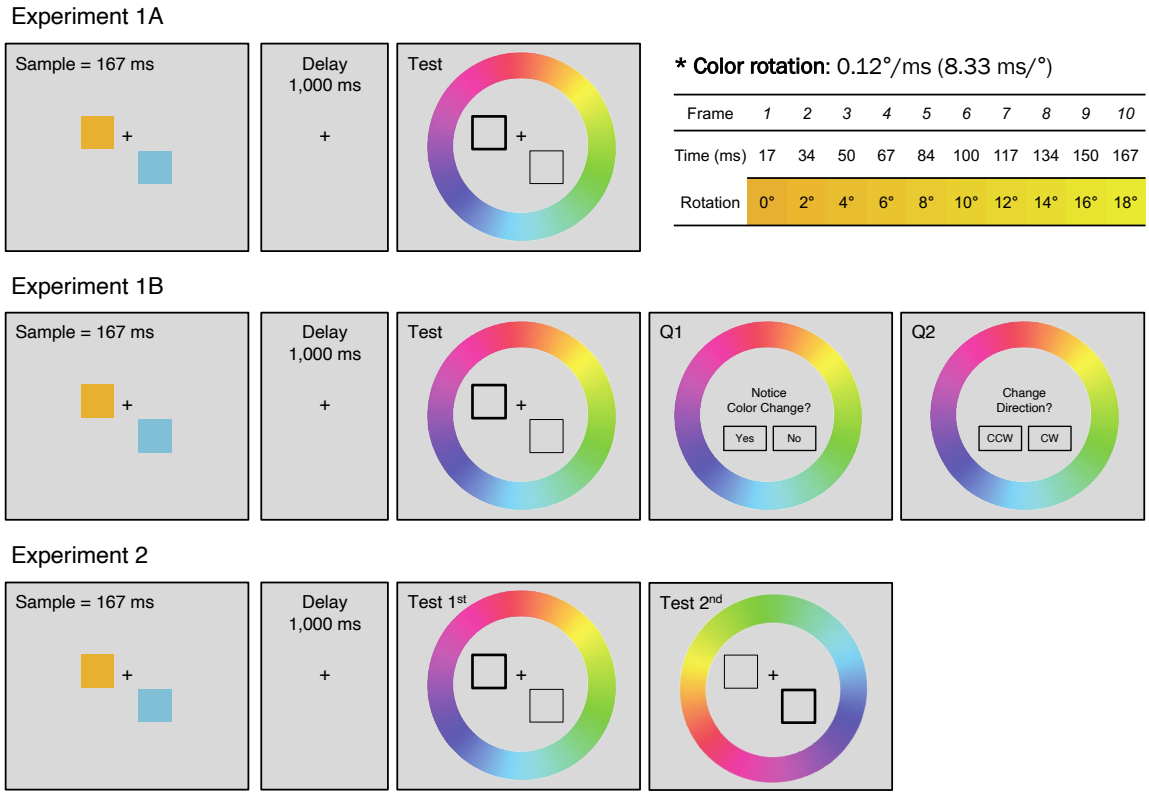
Zhang, W., & Luck, S. J. (2009b). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24-25.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science, 22*(11), 1434-1441.

Figures

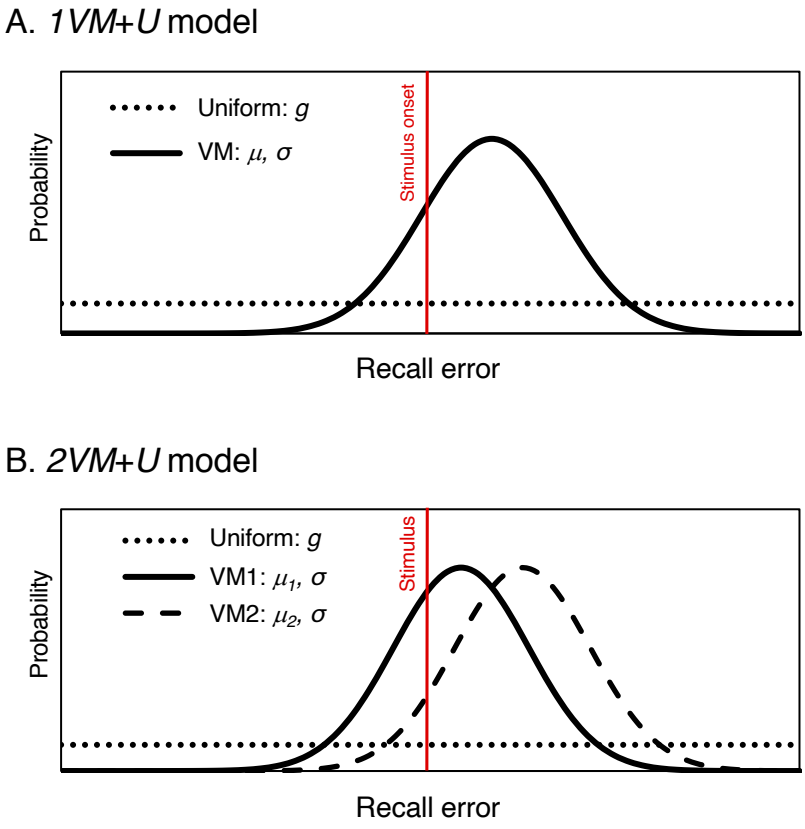
Figure 8.

Stimuli and procedure for the delayed estimation tasks in Experiment 1A, 1B, and 2.



Note. In all three experiments, two colored squares in the 167 ms memory array continuously changed color hues, either clockwise or counterclockwise, by 2° every 16.7 ms, resulting in a total of 18° change throughout the memory array (0.12°/ms or 8.33 ms/°, see top right table). In Experiment 1B, each recall test was followed by two additional questions for the subjective judgment of the presence of color change during the memory array and its direction. Participants reported their responses by clicking on one of the options presented below the question using a computer mouse. In Experiment 2, both memory items were tested in sequence, with the response color wheel randomly rotated between the two recalls.

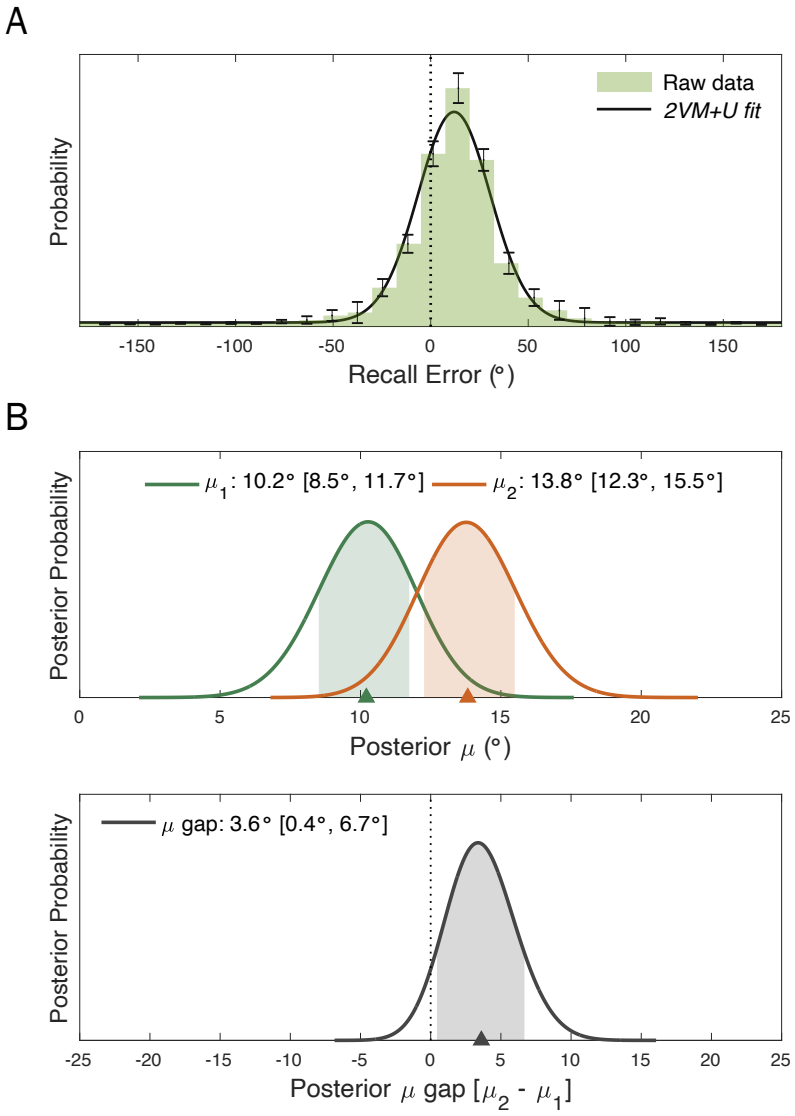
Figure 9.
The descriptive candidate models for the two competing hypotheses.



Note. (A) The *1VM+U* model for the concurrent encoding hypothesis is a mixture of a single von Mises distribution and uniform distribution. (B) The *2VM+U* model for the discrete encoding hypothesis consists of two von Mises distributions with different peak locations but the same variability terms, in conjunction with the uniform distribution. In both models, the μ and σ parameters represent bias and imprecision, respectively, and the g parameter represents guessing rate.

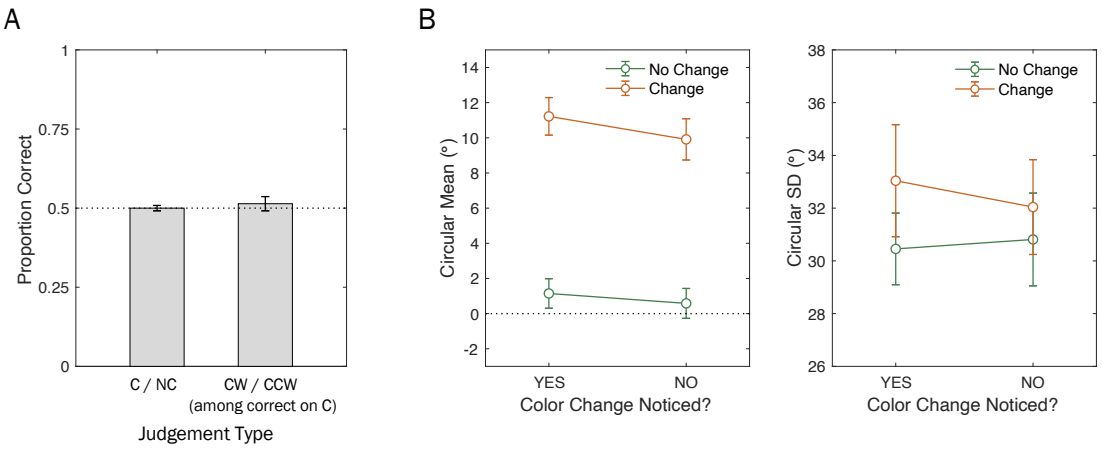
Figure 10.

Experiment 1A results of recall errors and the hierarchical Bayesian posteriors of the μ parameters.



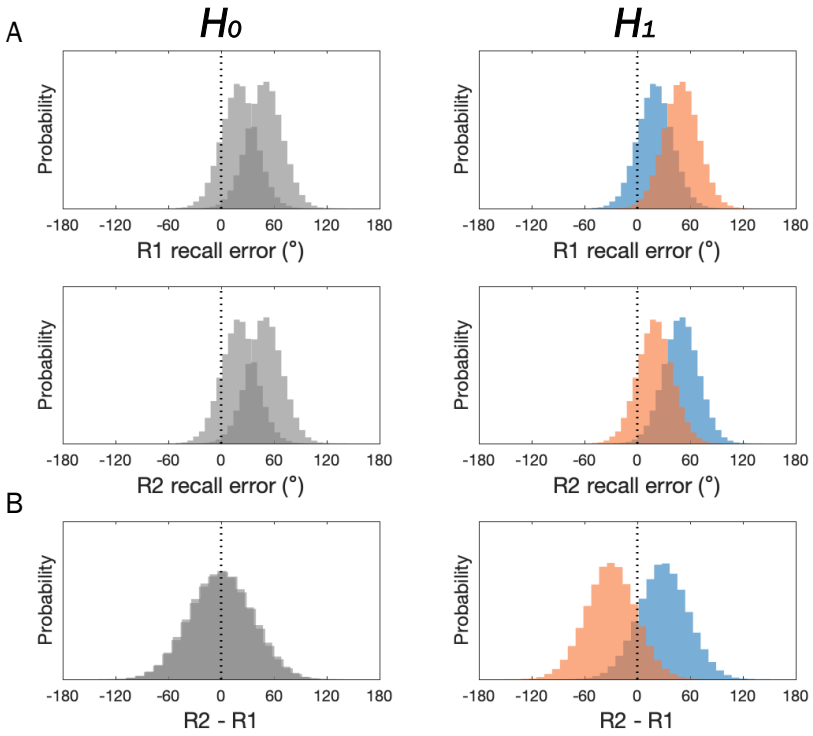
Note. (A) The histogram of the raw recall error (green bars) overlapped with the fit from the hierarchical Bayesian *2VM+U* model (black curve). Error bars represent the standard error of the mean. (B) The posterior probability of the μ_1 and μ_2 parameters from the *2VM+U* model (top), and their difference, μ_{gap} [$\mu_2 - \mu_1$] (bottom). The numeric values in the legends represent the mean and 95% highest density intervals (HDI) of the posterior parameter values, and the shaded area under the curves indicate the region between the lower and upper boundaries of HDIs.

Figure 11.
Experiment 1B results in subjective judgments and recall errors.



Note. (A) Accuracies for reporting the color Change versus No-Change (*C/NC*) of the memory items (left bar) and the color change direction (clockwise CW versus counter-clockwise CCW) from only color change trials when the change was correctly reported (*CW/CCW among correct on C*, right bar). (B) Participants' average of circular mean error (left) and circular standard deviation (right) measures as a function of their reporting of the color change (YES) versus no-change (NO), separately for color change and no-change trials (filled-circles indicate correct responses for each case). Error bars represent the standard error of the mean.

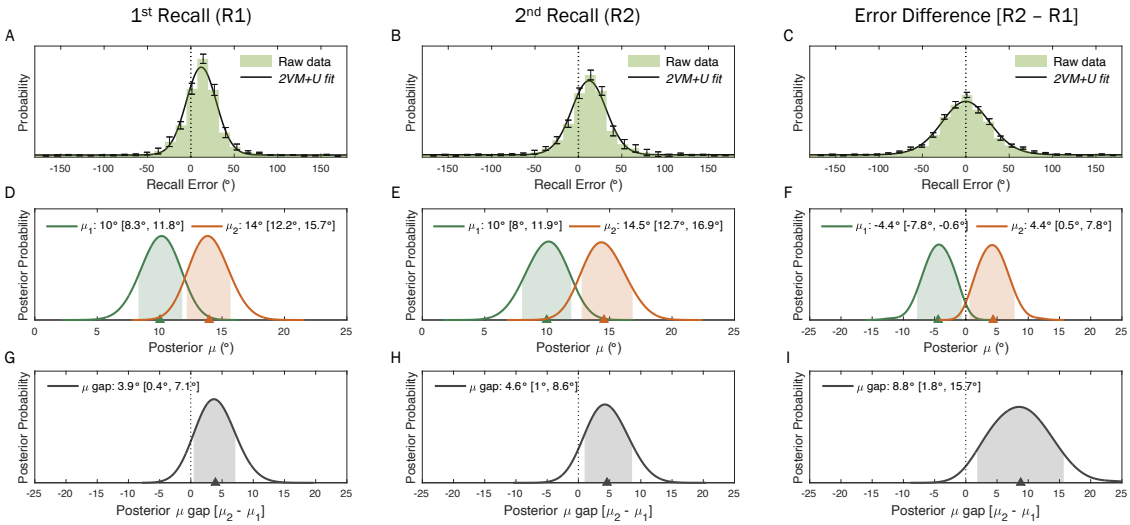
Figure 12.
Simulation of whole-report recall errors generated by the 2VM+U model and the hypothesis-driven predictions for the relative error difference.



Note. (A) Hypothetical 1st and 2nd recall errors (R1 and R2) simulated from the *2VM+U* model. The two von Mises (*VM*) distributions were sampled with a fixed set of parameters, VM_1 (20°, 20° for μ and σ) and VM_2 (50°, 20°), with a 30° gap in between. Note, the guessing component (*U*) was kept as zero in this simulation for simplicity and the representational shift parameters (μ_s) are exaggerated for illustration purpose. (B) Distinct predictions for the distribution of relative error difference (R2 – R1) can be generated based on competing hypotheses regarding the psychological validity of the separation of two shift parameters, μ_1 and μ_2. Specifically, if the separation of two shift parameters is merely a descriptive advantage for the trial-by-trial variability without psychological meaning (H_0; left panel), two recall responses would be independently-and-identically sampled from the same representational distribution (i.e., overall grey histograms). Consequently, this will result in a *uni*-modal distribution for the relative error differences (R2 – R1), with no credible separation of the two shift parameters. On the contrary, for the separation of two shift parameters to be valid (H_1; right panel), two recall responses should be mutually-exclusively sampled from two representational distributions that differ in central peak location (i.e., encoding time; color-coded). Therefore, the relative error difference (R2 – R1) should also be *bi*-modally distributed across trials, and subsequently captured by the *2VM+U* model with a credible separation of the two peak locations.

Figure 13.

Experiment 2 results of recall errors and the hierarchical Bayesian 2VM+U model posteriors of the μ parameters.



Note. (A-C) Raw recall error histograms (green bars) overlapped with the fitted curves from the posterior mean parameters of the 2VM+U model (black curves), separately for the 1st recall (R1), 2nd recall (R2), and their relative difference [R2 – R1], respectively. Error bars represent the standard error of the mean. (D-F) The posterior probability of the μ_1 and μ_2 parameters, and (G-I) their difference, μ_{gap} [$\mu_2 - \mu_1$] estimated from the corresponding data at each column. The numeric values in the legends represent the mean and 95% highest density intervals (HDI) of the posterior parameter values, and the shaded area under the curves indicates the lower and upper boundaries of $HDI_{95\%}$.

Chapter 4. μ Shift Testing the Influence of Long-Term Memory on Working Memory

4.1. Chapter Abstract

Prior stimulus familiarity has a variety of effects on visual working memory representations and processes. However, it is still unclear how familiarity interacts with the veridical correspondence between mnemonic representation and external stimuli. Here, we examined the effect of familiarity on two aspects of mnemonic correspondence, precision and accuracy, in visual working memory. Specifically, we used a hierarchical Bayesian method to model task performance in a change detection task with celebrity lookalikes (morphed faces between celebrities and non-celebrities with various ratios) as the memory stimuli. We found that familiarity improves memory precision by sharpening mnemonic representation but impairs memory accuracy by biasing mnemonic representation toward familiar faces (i.e., celebrity faces). These findings provide an integrated account of the puzzling celebrity sighting phenomena with the dissociable effects on mnemonic imprecision and bias and further highlight the importance of assessing these two aspects of memory correspondence in future research.

4.2. Introduction

Celebrities can be easily spotted on the street, although the internet is also flooded with celebrity lookalikes and hilarious "celebrity sightings" where strangers were mistaken as celebrities. These seemingly contradictory "celebrity sightings" phenomena bring up a question, are celebrity faces remembered *more precisely* (and hence more

recognizable), but *less accurately* (with biases where similar faces are easily misidentified as celebrities)?

At the center of this question is representational correspondence, the relationship between external stimuli and internal representations (Koriat et al., 2000). As the examples of celebrity sighting suggest, there can be dissociable effects of familiarity on two different aspects of memory correspondence. First, mnemonic *precision* can be conceptualized and operationalized as the inverse of variability in overt memory responses, reflecting the noisy nature of the underlying memory representations (Bays & Husain, 2008; Zhang & Luck, 2008). Second, mnemonic *accuracy* can be operationalized as the inverse of the deviation or bias in the mnemonic representations from the external stimuli. With the noisy memory representation modeled as a bell-shaped distribution in the memory feature space (e.g., faces, Figure 14A), mnemonic precision and accuracy manifest as the width (dotted green line) and offset (solid green line) of the peak, respectively (for details, see Methods).

First, the effects of familiarity on mnemonic precision are mixed in the literature. Some studies showed that familiarity can improve mnemonic precision in visual working memory (VWM, Lorenc et al., 2014; Montefusco-Siegmund et al., 2018; Scolari et al., 2008). For example, upright faces, more familiar than inverted faces to ordinary observers, can be remembered more precisely than inverted faces in VWM (Lorenc et al., 2014). In contrast, Xie and Zhang (2017b) showed that VWM precision is comparable for familiar and unfamiliar Pokémon characters, although a greater number of familiar Pokémon characters can be maintained in VWM (i.e., larger capacity) than unfamiliar

Pokémon characters. It is possible that the previous precision effects may result from the use of different stimuli across familiarity levels (e.g., difference in perceptual encoding of familiar versus unfamiliar stimuli such as configural face encoding; Azer & Zhang, 2019). Consistent with this account, Zhou and colleagues (2018) found that the VWM precision effect for familiar faces (own-race faces) relative to unfamiliar faces (other-race faces) only occurs with limited encoding time, whereas the capacity effect dominates when encoding is uninterrupted. Although either the capacity effects (Nishimura et al., 2021; Xie & Zhang, 2017b; Zhou et al., 2018) or precision effects (Lorenc et al., 2014; Montefusco-Siegmund et al., 2018; Scolari et al., 2008) can account for better VWM performance for familiar than unfamiliar stimuli (Buttle & Raymond, 2003; Curby & Gauthier, 2007; Jackson & Raymond, 2008), it is important to resolve the conflicting findings on the effects of familiarity on VWM precision.

Second, the effect of familiarity on mnemonic accuracy is largely unknown. However, several previous studies indicate mnemonic bias that leads to distortion of mnemonic representation can be functional (Mallett et al., 2020; Panichello et al., 2019; Teng & Kravitz, 2019). For instance, memory representations for faces or objects can be exaggerated to boost distinctiveness (Lewis & Johnston, 1998; Mauro & Kubovy, 1992; Rhodes et al., 1987; Tomita et al., 2014). Similar to this face caricature effect, the internal representation of a celebrity lookalike may bias toward the celebrity. Furthermore, this memory bias may provide an alternative account of the null effects of familiarity on VWM precision. Specifically, a model that does not capture memory accuracy may lead

to an underestimate of precision (Figure 14B) and subsequently cancel out the experimental effect of improved precision.

The present study has thus assessed the hypothesis that familiarity improves mnemonic precision but with reduced memory accuracy. We adopted a change detection task with faces morphed between American celebrity and non-celebrity faces with varied ratios and compared the task performance, using hierarchical Bayesian modeling (HBM), for the same face between participants who are familiar versus unfamiliar with a given celebrity. We predicted that familiarity of celebrity faces would produce attractive distortion of mnemonic representation of a morphed face toward the more familiar faces (i.e., celebrity faces) and more precise mnemonic representation (Figure 14C).

4.3. Method

4.3.1. Participants

Fifty-two subjects (19.0 ± 1.2 years; 30 women, 21 men, 1 other gender) recruited from UC Riverside Psychology Subject Pool participated in an online experiment hosted at testable.org for course credits. Sample size was determined by a priori power analysis for a two-by-four mixed-effect repeated measures analysis of variance (ANOVA; Faul et al., 2009, with $\alpha = .05$ and $\beta = .01$, under predefined effect sizes of $\eta_p^2 = .04$ and an intercondition correlation as $r = .70$ using G*Power (3.1). This analysis suggested that a sample size of 44 should provide sufficient power for the predefined effect sizes. We chose to err on the side of caution and collect a larger sample size. All participants had normal or corrected-to-normal vision and provided informed consent. The experimental protocol was approved by UC Riverside Institutional Review Board.

4.3.2. Stimuli

A total of 8 face stimuli were collected from Google images, including 4 faces from American celebrities (Anne Hathaway, Brad Pitt, Zendaya Coleman, and George Clooney) and 4 other faces. These celebrities were selected for the large individual differences in the undergraduate students' familiarity for these celebrities based on an informal survey. Each celebrity face was paired with a non-celebrity face with matching gender and race. All images were cropped, resized, and transformed to black and white. Especially, images of hair, ears, and neck parts were cropped to minimize the peripheral non-identity-related features (e.g., hairstyle). The faces in each pair (e.g., from Anne 100% to Anne 0%) were then used to produce 8 morphed images that represented gradual transitions from a celebrity face to a non-celebrity face, in steps of 11%, using Fantamorph 5.6.2 (Abrosoft). Given that the present study focuses on the effects of familiarity with celebrity, we only used the five images from each set that celebrity face prevailed over non-celebrity face (100:0%, 89:11%, 78:22%, 67:33%, and 56:44% in the ratio for celebrity : non-celebrity, respectively, Figure 15A). The mean size of a face stimulus was 253 x 400 in pixels. Note that the actual stimulus size and resolutions might vary depending on the participant's environment (e.g., monitor, video card, screen specifications, and lighting conditions) considering that this was an online experiment.

4.3.3. Procedure

4.3.3.1. Face Change Detection Task

The data was collected on Testable (http://testable.org), an online experiment platform. The experiment began with the task instructions and four practice trials

followed by the main experiment. On each trial of the face change detection task (for an example, see Figure 15B), a face was shown at the center of the memory display for 500-msec, followed by a delay interval for 1000-msec, and then another face at the center of the test display. Participants were asked to report whether the face in the test display was the same as or different from the face in the memory display by pressing a "Z" key (same) or an "X" key (different) on their keyboard. The test face stayed on the screen until the participants responded. Because this study was partially focused on memory bias, no immediate performance feedback was provided to the participants, to avoid strategical correction of the memory biases.

The participants completed a total of 320 trials, which consisted of 160 "same" face trials and 160 "different" face trials in random orders. In the "same" face trials, the same face repeated from the memory display to the test display. The face was randomly sampled from the five face morphs for each celebrity set that were repeated eight times (4 celebrity sets x 5 morphs x 8 repetitions). In the "different" face trials, the face in the memory display and the face in the test display were different face morphs from the same set (e.g., 100% Zendaya in the memory display changed to either 89%, 78%, 67%, or 56% Zendaya). That is, once the memory face was selected, a different face was randomly sampled from the four remaining face morphs for each celebrity set that were repeated two times (4 celebrity sets x 5 morphs for memory x 4 morph for test x 2 repetitions). The participants received three short breaks evenly distributed across the task session.

4.3.3.2. Familiarity Evaluation Questionnaire

After the face change detection task, two questions for each celebrity face were presented to the participants, rendering a categorical and a continuous self-report measure of celebrity familiarity, respectively. The first question was "*Have you seen this person on TV?*" with a forced choice between Yes or No. The second question was "*If you have seen this person on TV, how familiar are you with this person? If you haven't seen the person, please rate it as 0.*" The participants used a slide bar (0-10) to report their ratings. Participants were given unlimited time for each question. The participants were then divided into Seen versus Unseen groups based on their responses to the first question, and high familiarity group versus low familiarity group based on the median split of their responses to the second question. The subsequent between-group comparisons of the memory performance for the same task stimuli could minimize the physical confounds and differences in perceptual encoding of task stimuli compared to that for two different sets of faces (e.g., familiar faces vs. non-familiar faces as in Scolari et al., 2008, same-race versus other-race faces as in Zhou et al., 2018, and upright versus inverted faces as in Lorenc et al., 2014).

4.3.4. Data Analysis

In the face change detection task, the "different" face trials had a test face that consisted of a different morph ratio from the memory face within the same celebrity face set. This led to varying change magnitudes (differences in face morph ratio from memory to test) across the specific morph ratio of the memory faces (100% to 56% celebrity face), ranging from negative four steps toward the celebrity face to positive four steps away from the celebrity face (i.e., test – memory: -4, -3, -2, -1, 0, +1, +2, +3, +4 steps with 11%

change in morph ratio at each step, where 0 is "same" trial and all others are "different" trials). For instance, when the memory face was a 100% celebrity face, the test face could be the same 100% celebrity face for the "same" trials (0), or one of the 89% (+1), 78% (+2), 67% (+3), and 56% (+4) celebrity faces for the "different" trials. Similarly, trials with a memory face of a 67% celebrity morph ratio could have 56% (+1), 67% (0), 78% (-1), 89% (-2), or 100% (-3) celebrity faces for the test.

Thirteen out of 52 subjects who showed signal-detection theory *d*-primes (the difference between standardized hit rate and false-alarm rate, *z(Hit) – z(FA)*) below *zero* from one or more conditions of celebrity face set in the face change detection task were excluded from data analysis due to poor performance (i.e., chance-level discriminability). Overall, rejected subjects' mean *d*-prime was 0.23 (± 0.28 standard deviation) and mean accuracy of 53.8% (± 4.6% standard deviation). To precisely examine participants' change detection performance, we first calculated the proportion of the "same" responses as a function of the change magnitudes for each % celebrity memory face (Figure 16). We then modeled these proportion same response curves using the Gaussian function consisting of three parameters:

$$f(p(Same)|CM) = Amp \times e^{-\left(\frac{CM-\mu}{\sigma}\right)^2}$$

where *CM* represents the change magnitude and the three Gaussian parameters, Amplitude (*Amp*), Mu (μ), and Sigma (σ), represent the height of the central peak of the curve, the horizontal shift of the peak from zero, and the width of the curve, respectively. In the context of the present study, this Gaussian function is used as a measurement model, and μ and σ parameters are interpreted to capture representational *bias* and

imprecision properties, respectively. The *Amp* parameter is a scaling factor for the response curve that is largely driven by the response tendency of reporting "same" across change magnitudes. Consequently, the Result section will focus on the imprecision and bias parameters given the primary testing hypothesis.

This analytic method allowed us to estimate the model parameters at each % celebrity memory face and each subject, even though the to-be-fitted data at each condition had a restricted and varying set of observations (i.e., five change magnitudes for each memory face morph type, over nine possible data points ranging from negative four to positive four steps). However, the number of responses was highly limited for all the data points on the proportion same response curves, making conventional curve-fitting that minimizes the least-squares error less reliable. To address this issue, we adopted HBM method for fitting the Gaussian function to the proportion of the same responses. This method simultaneously accounts for multiple levels of variance for each parameter, specifically a population-level, a participant-level, a stimulus-level of the celebrity face set (i.e., four celebrities), and another stimulus-level of the memory face type (i.e., five morph ratios), and further provides population-level posteriors of each parameter. The population-level parameter posteriors were estimated through a general linear model, sampling values from the normal distribution whose mean is a sum of the fixed and random effects, and the variability term is a sum of the interaction effects across levels (Park et al., 2021). This HBM method can produce reliable parameter estimation and recovery of numerically and statistically subtle effects on mnemonic

biases, for limited data that is challenging for non-hierarchical model fitting methods (Park & Zhang, 2019).

We used Markov Chain Monte Carlo simulations (20,000 samples after 20,000 warm-ups). To minimize biases driven by the choice of priors, reasonably informative priors were chosen for all parameters [e.g., $Amp_{fixed} \sim Beta(1, 1)$; $Amp_{random} \sim Normal(0, Var_{Amp})$; $Var_{Amp} \sim Gamma(1, 1)$]. Statistical inference was made based on the resulting posterior means and their 95% credible intervals (highest density interval, HDI), as a point estimate and an error term (similar to the frequentist 95% confidence interval), respectively. Specifically, the range of HDIs can indicate the strength of evidence depending on whether the upper or lower bound of 95% HDI for the experimental effect crosses over zero or not (Kruschke, 2013). Change detection performance and the estimated posteriors of parameters were then compared between two participants groups (i.e., Seen vs. Unseen; high familiarity and low familiarity) to examine whether or how familiarity of celebrity faces affects mnemonic representations of faces.

4.4. Results

As expected, the proportion of the same responses in the change detection task decreased with the change magnitudes (Figure 16) and the celebrity familiarity ratings showed reasonable individual differences (Figure 17A). Overall change detection performance, measured by signal-detection theory *d*-prime, was better for high familiarity group than low familiarity group (Figure 17B). Consistent with this observation, a two-by-four factorial analysis of variance (ANOVA) for *d*-prime as a function of the categorical familiarity group (*Seen* vs. *Unseen*), and the celebrity set (*Anne*, *Brad*,

Zendaya, & *George*), revealed a significant main effect of the categorical measure of familiarity, $F(1, 148) = 19.57$, $p < .001$, $\eta^2_p = .12$. The main effect of the celebrity set, $F(3, 148) = 1.92$, $p = .130$, $\eta^2_p = .04$, and the interaction effect were not significant, $F(3, 148) = 0.52$, $p = .672$, $\eta^2_p = .01$. In addition, the continuous measure of familiarity rating positively correlated with the *d*-prime measure across the participants and celebrity face sets, $r(154) = .30$ [95% CI: 0.15, 0.44], $p < .001$. These findings suggest that familiarity improves VWM performance, consistent with the literature (Lorenc et al., 2014; Montefusco-Siegmund et al., 2018; Scolari et al., 2008).

To assess how familiarity affects mnemonic precision and accuracy, we modeled the proportion same responses with HBM of Gaussian function (see *Data Analysis* for detail and Figure 16 for the overall fits). The population-level posteriors for the bias (μ) and imprecision (σ) measures showed large differences between the categorical measure of familiarity (between the *Seen* and *Unseen* groups, Figure 18), manifested in posterior mean and 95% HDIs that do not cross over zero (bias: μ, -0.144, [-0.233, -0.052]; imprecision: σ (-0.384 [-0.517, -0.254]). Similar patterns were obtained for each celebrity face set (see Figure 19). Consistent with our predictions, these findings indicate that VWM representations for faces were more precise for familiar facial identities, but at the same time, biased toward the celebrity faces.

4.5. Discussion

The present study developed an HBM method to model change detection task performance for morphed faces that look alike celebrities to assess whether and how prior familiarity affects VWM precision and accuracy. We found that familiarity improves

memory precision by sharpening mnemonic representation but impairs memory accuracy by biasing mnemonic representation toward familiar faces.

These findings add to the growing literature on the effects of familiarity on VWM. Familiarity has been found to improve VWM capacity (Jackson & Raymond, 2008; Xie & Zhang, 2017b), speed up VWM consolidation (Xie & Zhang, 2018), and increase VWM encoding efficiency by prioritizing novel sensory information (Bruning & Lewis-Peacock, 2020). Our findings suggest that familiarity also affects memory correspondence, consistent with the previous findings that various factors, such as co-occurred stimulus (Bae & Luck, 2017; Golomb, 2015), emotional context (Xie et al., 2022; Xie & Zhang, 2016, 2017a), and prior knowledge (Hansen et al., 2006) can modulate mnemonic precision and accuracy of internal representations.

Although the benefits of high mnemonic precision for memory performance are straightforward (e.g., vivid memory retrieval with confidence and details, Xie & Zhang, 2017a), the functional roles of mnemonic bias are less clear. Theoretically, mnemonic biases can maximize memory performance and minimize memory losses (Alvarez, 2011; Chunharas et al., 2022; Golomb, 2015; Huttenlocher et al., 2000; Schacter, 2012). For example, when two items are similar in feature space (e.g., color or orientation), their VWM representations tend to repulse each other to reduce their confusability (Bae & Luck, 2017; Scotti et al., 2021). However, when two items are dissimilar, their VWM representations tend to attract each other to increase the probability of successful memory retrieval (i.e., ensemble representation, Alvarez, 2011). Canonical information also biases WM representation. For example, an oblique bar (e.g., 85°) close to the cardinal

orientation (e.g., 90°) tends to be represented toward the cardinal orientation to reduce working memory load (Bae & Luck, 2019; Crawford et al., 2016; Hardman et al., 2017).

In addition to mnemonic biases in isolation, the current observation of the precision and accuracy tradeoff has also been observed across cognitive processes such as categorization (Briscoe & Feldman, 2011; Huttenlocher et al., 2000), decision making (Yaniv & Foster, 1995, 1997), and eyewitness memory (Evans & Fisher, 2011). For example, prior knowledge such as categories can be used to adjust inexactly represented stimuli, leading to memory bias, but improved memory precision (Huttenlocher et al., 2000). This precision-accuracy tradeoff may be driven by a decisional strategy to improve the confidence and informativeness from mnemonic precision at the cost of memory accuracy under uncertainty (Yaniv & Foster, 1995, 1997), which could also account for the present findings.

Some further caveats of the present findings need to be noted. First, the face feature space generated from the face morphing procedure has evenly distributed steps of change in physical space. However, this feature space may not be represented linearly in perception and memory. Nonetheless, the relative effects of bias should still be present with nonlinear models of VWM (e.g., Schurgin et al., 2020). Furthermore, given the small memory set size of one, the present precision effect could in principle manifest to the precision measure in some alternative models of VWM (van den Berg et al., 2014) or the memory strength measure (Schurgin et al., 2020). Second, the present findings suggest that the previous null results of familiarity on VWM precision (e.g., Xie & Zhang, 2017b) may result from the underestimate of precision due to memory biases canceling

out the true precision effect. However, it is still possible that the larger VWM capacity for familiar information (Jackson & Raymond, 2008; Xie & Zhang, 2017b) may lead to better precision for the single memory item in the current study due to capacity and precision tradeoff (the "averaging' component of Zhang & Luck 2008 model, Zhang & Luck, 2008). Lastly, mnemonic imprecision and bias are conceptually and mathematically independent. However, their underlying neurocomputational mechanisms are still largely unknown. Future research needs to explore whether the underlying neurocomputational mechanisms for mnemonic precision and bias can be dissociated.

In conclusion, the present study provides an integrated account of the puzzling celebrity sighting phenomena with the dissociable effects on imprecision and bias. In addition, the finding of precision-accuracy tradeoff suggests an alternative account of the conflicting literature regarding the effects of familiarity on VWM precision. Our findings further highlight the theoretical and empirical importance of assessing both memory imprecision and bias in hypothesis testing and memory performance measurement in future studies. It is also pivotal to test the extent to which VWM precision and accuracy are affected by affective, aging, and mental health factors.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Azer, L., & Zhang, W. (2019). Composite face effect predicts configural encoding in visual short-term memory. *Frontiers in Psychology, 10*, 2753.

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, and Psychophysics, 79*(8), 2376–2395.

Bae, G. Y., & Luck, S. J. (2019). Reactivation of previous experiences in a working memory task. *Psychological Science, 30*(4), 587–595.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851–854.

Briscoe, E., & Feldman, J. (2011). Conceptual complexity and the bias/variance tradeoff. *Cognition, 118*(1), 2–16.

Bruning, A. L., & Lewis-Peacock, J. A. (2020). Long-term memory guides resource allocation in working memory. *Scientific Reports, 10*(1), 1–10.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & Psychophysics, 65*(8), 1296–1306.

Chunharas, C., Rademaker, R. L., Brady, T. F., & Serences, J. T. (2022). An adaptive perspective on visual working memory distortions. *Journal of Experimental Psychology: General*. Advance online publication.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review, 14*(4), 620–628.

Elizabeth Crawford, L., Landy, D. H., & Salthouse, T. A. (2016). Spatial working memory capacity predicts bias in estimates of location. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*(9), 1434.

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology, 25*(3), 501–508.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology. Human Perception and Performance, 43*(1), 30-54.

Huttenlocher, J., Hedges, L. v., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology. General, 129*(2), 220–241.

Jackson, M. C., & Raymond, J. E. (2008). Familiarity Enhances Visual Working Memory for Faces. *Journal of Experimental Psychology. Human Perception and Performance, 34*(3), 556.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*(2), 573–603.

Lewis, M. B., & Johnston, R. A. (1998). Understanding caricatures of faces. *The Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology, 51*(2), 321–346.

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, and Psychophysics, 76*(7), 1975–1984.

Mallett, R., Mummaneni, A., & Lewis-Peacock, J. A. (2020). Distraction biases working memory for faces. *Psychonomic Bulletin & Review, 27*(2), 350.

Mauro, R., & Kubovy, M. (1992). Caricature and face recognition. *Memory & Cognition*, *20*(4), 433-440.

Montefusco-Siegmund, R., Toro, M., Maldonado, P. E., & Aylwin, M. de la L. (2018). Unsupervised visual discrimination learning of complex stimuli: Accuracy, bias and generalization. *Vision Research, 148*, 37–48.

Nishimura, Y., Tsuda, H., & Ogawa, H. (2021). Own-race advantage in visual working memory for faces reflects enhanced storage capacity and quick encoding. *Japanese Psychological Research*. Advance online publication.

Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Communications, 10*(1), 1–11.

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.

Park, H.-B., & Zhang, W. (2019). Hierarchical Bayesian modeling for testing representational shift in visual working memory. *Journal of Vision, 19*(10), 80.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive Psychology, 19*(4), 473–497.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *The American Psychologist, 67*(8), 603–613.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour, 4*(11), 1156–1172.

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15*(1), 215–222.

Scotti, P. S., Hong, Y., Leber, A. B., & Golomb, J. D. (2021). Visual working memory items drift apart due to active, not passive, maintenance. *Journal of Experimental Psychology: General, 150*(12), 2506–2524.

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour, 3*(8), 827–836.

Tomita, A., Yamamoto, S., Matsushita, S., & Morikawa, K. (2014). Resemblance to familiar faces is exaggerated in memory. *Japanese Psychological Research, 56*(1), 24–32.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121*(1), 124–149.

Xie, W., Sing, J. C. L. L., Martinez-Flores, A., & Zhang, W. (2022). Induced negative arousal modulates the speed of visual working memory consolidation. *Emotion, 22*(1).

Xie, W., & Zhang, W. (2016). Negative emotion boosts quality of visual working memory representation. *Emotion, 16*(5), 760–774.

Xie, W., & Zhang, W. (2017a). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition, 166*, 73–83.

Xie, W., & Zhang, W. (2017b). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory and Cognition, 45*(4), 677–689.

Xie, W., & Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: Electrophysiological evidence from contralateral delay activities. *Journal of Cognitive Neuroscience, 30*(1), 1–13.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124*(4), 424–432.

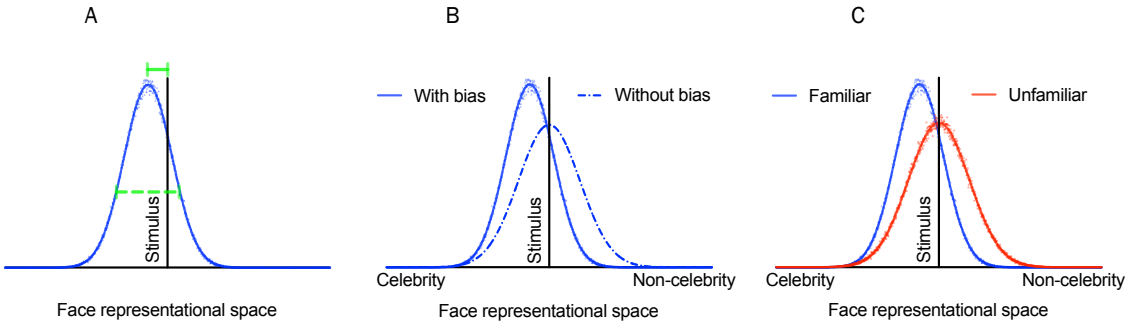
Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10*, 21–32.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233.

Zhou, X., Mondloch, C. J., & Emrich, S. M. (2018). Encoding differences affect the number and precision of own-race versus other-race faces stored in visual working memory. *Attention, Perception, and Psychophysics, 80*(3), 702–712.

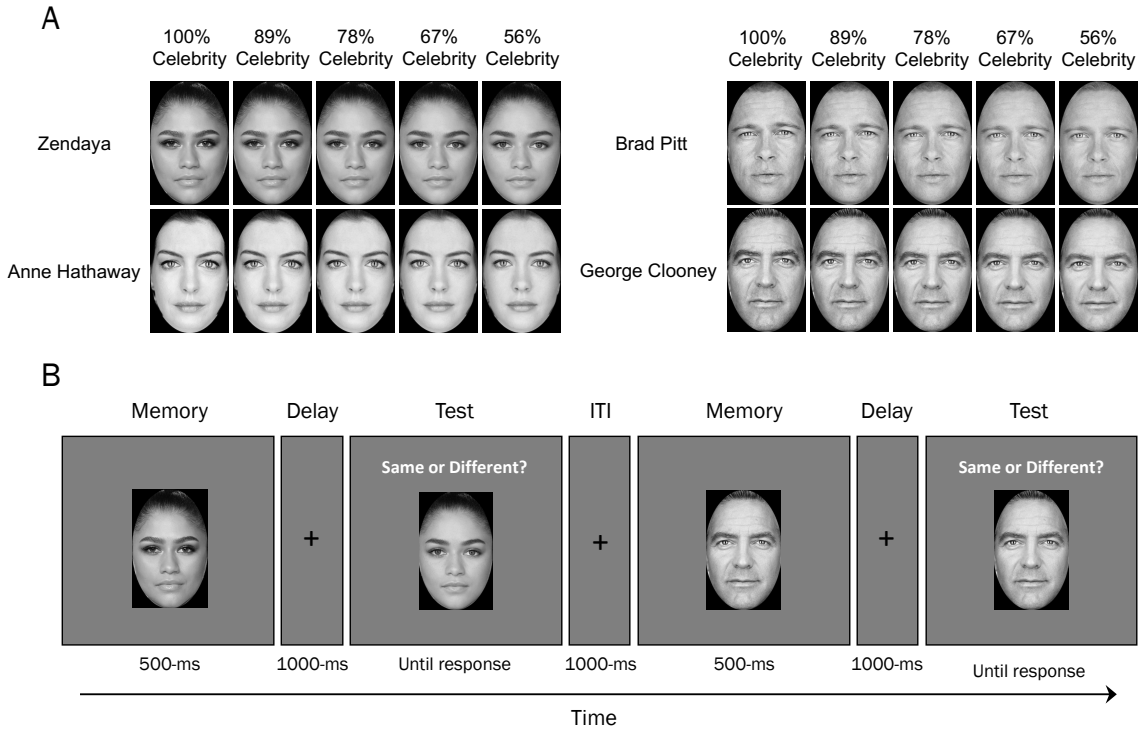
Figures

Figure 14.
Model and the hypotheses.



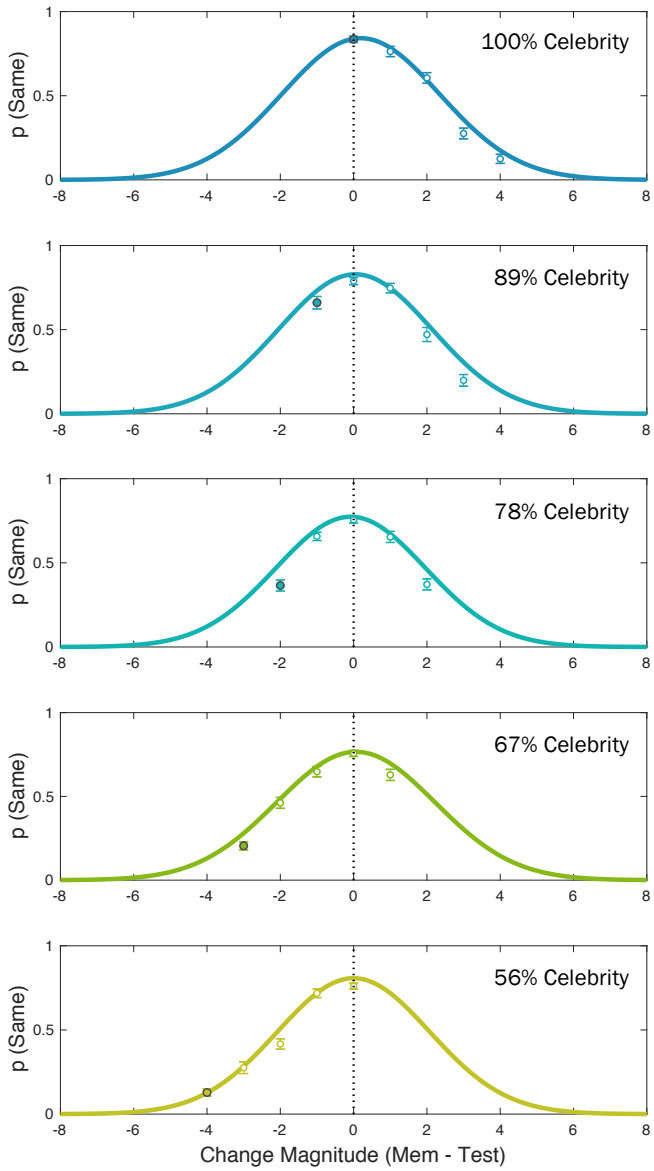
Note. (A) Noisy memory representation manifested as a Bell-shaped distribution in the memory feature space. Mnemonic precision and accuracy are represented by the width (dotted green line) and offset (solid green line) of the peak, respectively. (B) Model showing an underestimate of precision by not capturing memory accuracy. (C) The hypothesized attractive distortion of mnemonic representation toward the celebrity faces and more precise mnemonic representation.

Figure 15.
Stimuli and procedure of the change detection task.



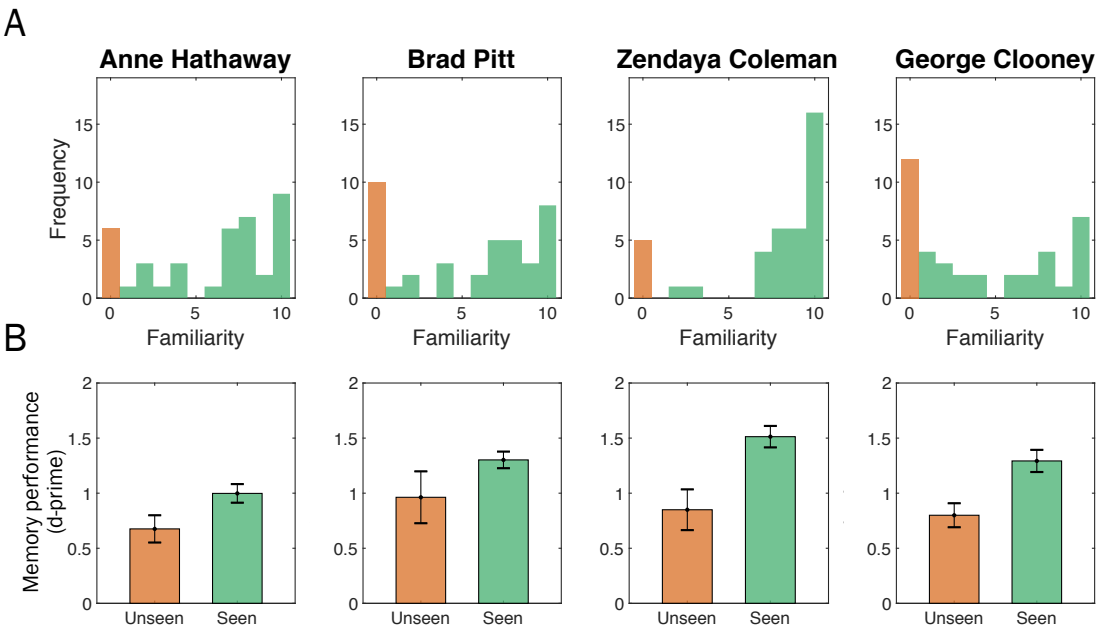
Note. (A) The four sets of five face morphs used in the face change detection task. (B) In the face change detection task, participants tried to remember a briefly presented memory face and, after a retention interval, reported whether a test face was the same as or different from the memory face. On a change trial, the test face could be another face with a different morph ratio, from the same face set, than the one for the memory face. For example, in the first trial in the figure, the memory face (100% Zendaya) changes to 56% Zendaya in the test (i.e., a "different" trial). In the second trial, the memory face and the test face morph are both 78% George (i.e., a "same" trial). The proportions of the images are modified for illustration purpose.

Figure 16.
A summary of the face change detection task performance.



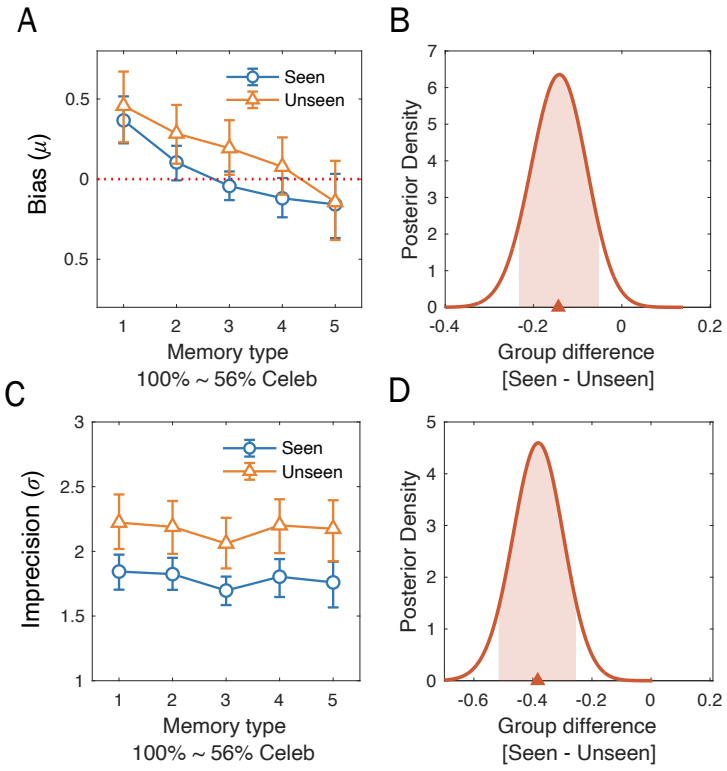
Note. Proportion of the same responses (circles), averaged across celebrity face sets and participants, as a function of the change magnitude, separately for each memory face %celebrity morph ratio, along with the hierarchical Bayesian model (HBM) fits (solid lines). The possible change magnitudes on a given trial are constrained by the memory face % celebrity morph ratio. When mapped onto the nine-step change magnitude, zeros on the x-axis represent "same" face trials, whereas negative and positive change steps represent the face changes in a direction towards the celebrity and non-celebrity faces, respectively. The filled circles represent the 100% celebrity faces at the test. The error bars represent the standard error of mean.

Figure 17.
Overall celebrity familiarity ratings and the change detection performance.



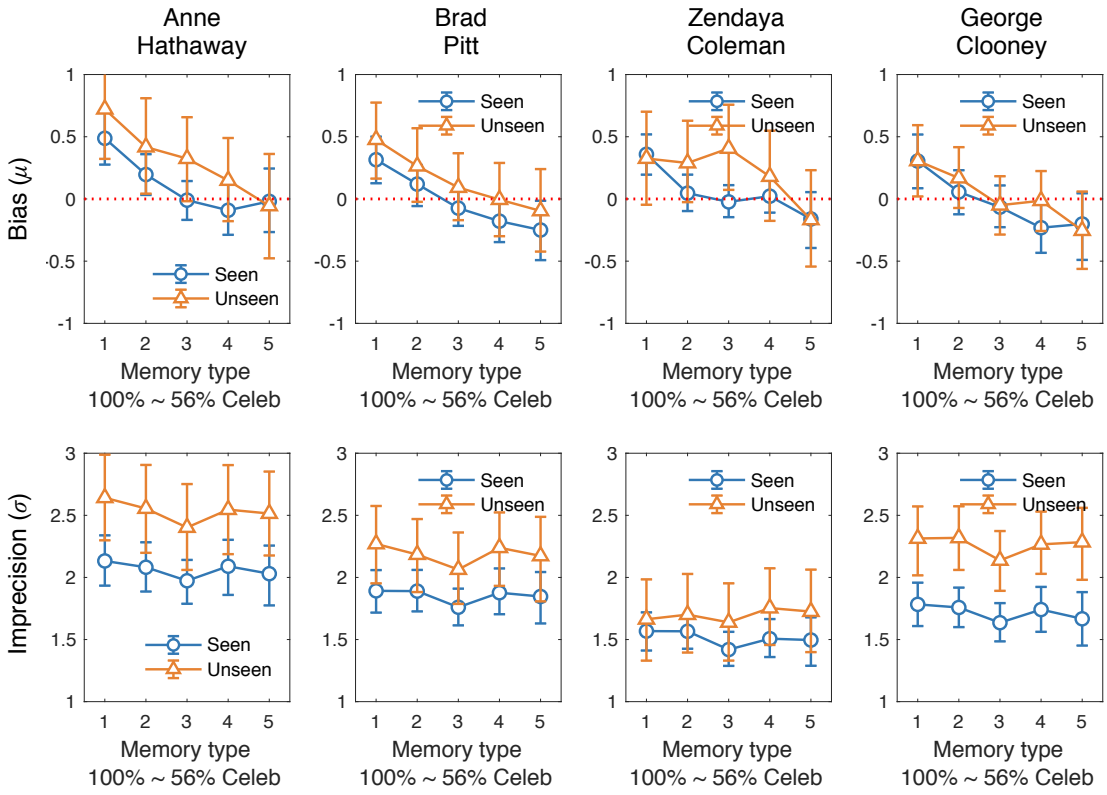
Note. (A) Histograms of familiarity ratings for celebrity faces, where the zero familiarity rating was from participants who reported that they have *not seen* a given celebrity face before (*Unseen* group) and 1-10 familiarity ratings were from those with the *seen* reports (*Seen* group). Note, participants could have different *Seen*/*Unseen* responses and familiarity ratings (and hence different subject grouping) for different celebrity faces. Overall, 84.6%, 79.5%, 87.2%, and 64.1% of participants gave "*Seen*" responses for the faces of *Anne Hathaway*, *Brad Pitt*, *Zendaya Coleman*, and *George Clooney*, respectively. (B) Overall change detection performance, measured by *d*-prime, between Unseen and Seen groups. The error bars represent the standard error of mean.

Figure 18.
Population-level posteriors for bias (μ) and imprecision (σ) and the group difference effects.



Note. (A, C) The population-level posteriors for the bias (μ) and imprecision (σ) parameters as a function of memory face %celebrity morph ratio, separated by the categorical measure of familiarity (*Seen* vs. *Unseen*). (B, D) The population-level parameter posteriors for the group difference effect, with the shaded area under the curve representing 95% highest density intervals.

Figure 19.
Celebrity-by-celebrity results of the population-level posteriors for bias (μ) and imprecision (σ) measures.



Note. Error bars represent 95% highest density intervals (HDIs).

PART III.

The traditional information-processing framework posits that cognitive processes, action planning, and motor execution are operated in a serial manner (Sternberg, 1969; Tversky & Kahneman, 1981). However, many studies have challenged this view by proposing that cognition and action are rather interactive and parallel processes (Cisek, 2007; Spivey et al., 2005). The action-oriented view of cognition has encouraged many studies on oculomotor control coupled with endogenous attention and has provided a rich understanding of how ongoing cognition is reflected in motor action and movement trajectory (Awh et al., 2006; Howard & Tipper, 1997; van Ede et al., 2019; Welsh & Elliott, 2004).

Behavioral responses in cognitive tasks involve sequences of discrete movements and leave continuous trajectories. The analysis of movement trajectory can provide unique insights to look into temporal changes in internal cognitive processes before a final response is made (Song & Nakayama, 2009). For example, studies have shown that computer mouse trajectories can be indicators of conflicts between multiple choices during decision-making (Freeman et al., 2011). Similarly, studies of movement trajectory in different domains, saccadic eye movements or manual reaching, have shown that online-movement trajectories to a target item were systematically deviated due to the impact of distractors (Tipper et al., 1997; Menceloglu et al., 2021). Moreover, changes in the direction of trajectory curvature suggest temporal dynamics of selective and inhibitory attentional signals (Erb et al., 2016; Welsh & Elliott, 2004).

A quite number of studies with movement trajectory analysis have primarily focused on the deviation property of the trajectory pattern (see for review, Van der Stigchel et al., 2006). The direction of trajectory deviation and its magnitude can be mapped onto the underlying spatial attention mechanisms such as inhibition against distractors in the visual field (Doyle & Walker, 2001), especially when combined with the temporal dynamics of such deviations (McSorley et al., 2006). Therefore, one of the appealing aspects of including movement trajectory analysis in this dissertation is that trajectory deviations can be thought to reflect biases in the underlying representation. Moreover, trajectory data can be measured on a fine temporal scale, thus offering insights into the temporal dynamics of how the central tendency changes over time (i.e., moment-by-moment evolution of shift errors). In the following chapters, I took advantage of the mouse cursor movement trajectory analysis in testing the dynamics of representational shift during decision-making and motor responses, specifically testing the influence of the current WM contents on the subsequent processes of attention and perception. The mouse trajectories in these studies were obtained in a specialized task paradigm where the mouse cursor movements were bounded within the circular response window (e.g., heading toward a certain color value on a color wheel).

Chapter 5. μ Shift Testing Between-Item Interference Arising from Variance of Precision Across Multiple Working Memory Representations

5.1. Chapter Abstract

Multiple representations in visual working memory (VWM) can vary in mnemonic precision. This inhomogeneity of VWM precision has received support from recent studies with the whole-report procedure, in which all memory items are recalled in free or forced orders. Recently, Hao and colleagues (2021) 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 to forced-recall, which was taken as evidence for less between-item interference. Here, we reanalyzed the original dataset with a different analytic approach and attempted independent hypothesis testing focusing on trajectory deviations. We found that the direction of 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 arising from variance of precision of VWM representations, and further highlights the novel methodological benefits of probing memory decisional processes using mouse trajectory data.

5.2. Introduction

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; Fougnie 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 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; Rademaker et al., 2012).

Hao and colleagues (2021) recently revised the whole-report task by adding a separate item-selection stage immediately before each recall (Figure 20). 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 non-preferred locations. 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 1st 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 1st 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 inter-item dynamics can be explained by variable mnemonic quality across items. The rich mouse trajectory dataset from Hao and colleagues (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 and colleagues (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.

5.3. Method

5.3.1. Data Reanalysis and Rationales

The original data from Hao and colleagues (2021) were retrieved from the authors' deposit of the data at Open Science Framework (https://osf.io/67upz/). Figure 21 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 ± 11.2%, 3.4 ± 1.4%, and 3.3 ± 1.2%

normalized time from display onset to click, for the 1st, 2nd, and 3rd item-selection, respectively).

Second, to map the raw mouse cursor positions in Cartesian coordinates (x, y) to polar coordinate (-π to π), Hao and colleagues (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 and colleagues (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 Figure 21 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 and colleagues (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 Figure 21, 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 Figure 20B, 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 and colleagues (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 1st recall item was placed within four displacements from the other two items on the opposite side among eight possible locations (i.e., 1st target placed somewhere in the middle of 2nd and 3rd items). For example, from Figure 22A, when the 1st recall item was placed at the location marked as '1', only those trials where the 2nd and 3rd items were placed at the

[1] Note 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.

opposite side between location '1' and '5' were included for data analysis. In other words, all other types of trials where 2nd and 3rd recall items were placed on the same side relative to the 1st recall item were excluded. We allowed uneven clockwise and counter-clockwise displacements from the 1st item to maximize the trial number (e.g., 2nd and 3rd item placed three-steps clockwise and two-steps counter-clockwise to 1st 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 warming-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).

5.4. Results

5.4.1. Endpoint Bias and Its Impact on Between-Item Interference

In Hao and colleagues (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 (Figure 22A). The extent of endpoint variability was estimated for individual participants by calculating circular *SD* as a function of recall order (1st, 2nd, vs. 3rd) and recall type (forced vs. free), separately (Figure 22B). A two-way repeated-measures analysis of variances (ANOVA) revealed significant main effects of recall order, $F(2, 30) = 15.36$, $p < .001$, $\eta^2_p = .51$, and recall type, $F(1, 15) = 10.24$, $p = .006$, $\eta^2_p = .41$, without significant interaction effect, $F(2, 30) = 0.06$, $p = .943$, $\eta^2_p = .00$. Surprisingly, endpoint bias was greater in free-recall than forced-recall, which is opposite to what can be inferred from Hao and colleagues (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 23 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 1st 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 2nd 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 3rd 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 (Figure 23B), with AUC differences [forced – free] in the 1st, 2nd, and 3rd 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, ds > 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, $F(2, 30) = 15.71$, $p < .001$, $\eta^2_p = .51$, and endpoint,

$F(1, 15) = 26.69$, $p < .001$, $\eta^2_p = .64$. However, there was no significant interaction between them, $F(2, 30) = 1.04$, $p = .361$, $\eta^2_p = .07$.

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.

5.4.2. Within-Trial Trajectory Predicts Relative Precision of VWM Items

Hao and colleagues (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 1st item-selection are related to subsequent item recall performance. We predict that mouse

trajectory to the 1st forced-target would be attracted toward one of the two non-targets, where its direction may depend on the relative precision of those items. Specifically, on two-thirds probability, the 1st forcedly-chosen target would not be of the *best* precision and mouse movement to this non-best 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 1st item-selection trajectory deviated toward 2nd or 3rd recall item (by the sign of trajectory AUC), whereas the other dimension categorizes the relative recall performance (2nd – 3rd absolute recall errors; a negative value indicates better recall for the 2nd item than 3rd item). Note, for the first dimension of the sign of trajectory AUC, AUCs from trials with the 3rd item presented counterclockwise (–) to the 1st item were reversed. Hence, negative and positive AUCs represent 1st item-selection trajectory deviated toward the 2nd item (–) and the 3rd item (+), respectively. For the second dimension of the absolute [2nd – 3rd] recall error differences, the overall main effect of recall order (i.e., 2nd recall more precise than 3rd 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 3rd recall more precise) while subtracting it from the bottom row cells (i.e., when 2nd recall more precise). This was to equate the null probability of top and bottom cells at 50% each. Trials with the absolute relative [2nd – 3rd] 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 1st item-selection trajectory bias was predictive of the relative recall precision of the 2nd and 3rd 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 1st recalled item had the best precision.

The results were consistent with these predictions (Figure 24). 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 1st 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 1st item-selection trajectory predicts representational quality of the subsequently recalled items, we fitted 2nd and 3rd recall errors with the extended Mixture model (Zhang & Luck, 2008) using hierarchical Bayesian method,

separately for different 1st item-selection trajectory bias direction (toward 2nd or 3rd item; left and right cells in Figure 24). Figure 25 summarizes the resulting population-level posteriors of *μ, SD,* and *guessing* parameters*,* showing their difference between two recalls (2nd – 3rd). Further analyses based on posterior mean with 95% HDI for the difference between two cases when the 1st item-selection trajectory deviated toward the 2nd or 3rd 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 1st item-selection under forced-recall.

5.5. Discussion

The present study reanalyzed the behavioral and mouse trajectory data from Hao and colleagues (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 1st 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 non-target 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.

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 1st 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 2nd and 3rd item locations to the 1st 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 2nd and

3rd 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 moue 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.

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., 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., & Luck, S. J. (2012). Simultaneous control of attention by multiple working memory representations. *Psychological Science, 23*(8), 887.

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.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851-854.

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., Dosher, 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. (2016). 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. Dzharafarov, & 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.

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.

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.

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., Cappiello, 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.

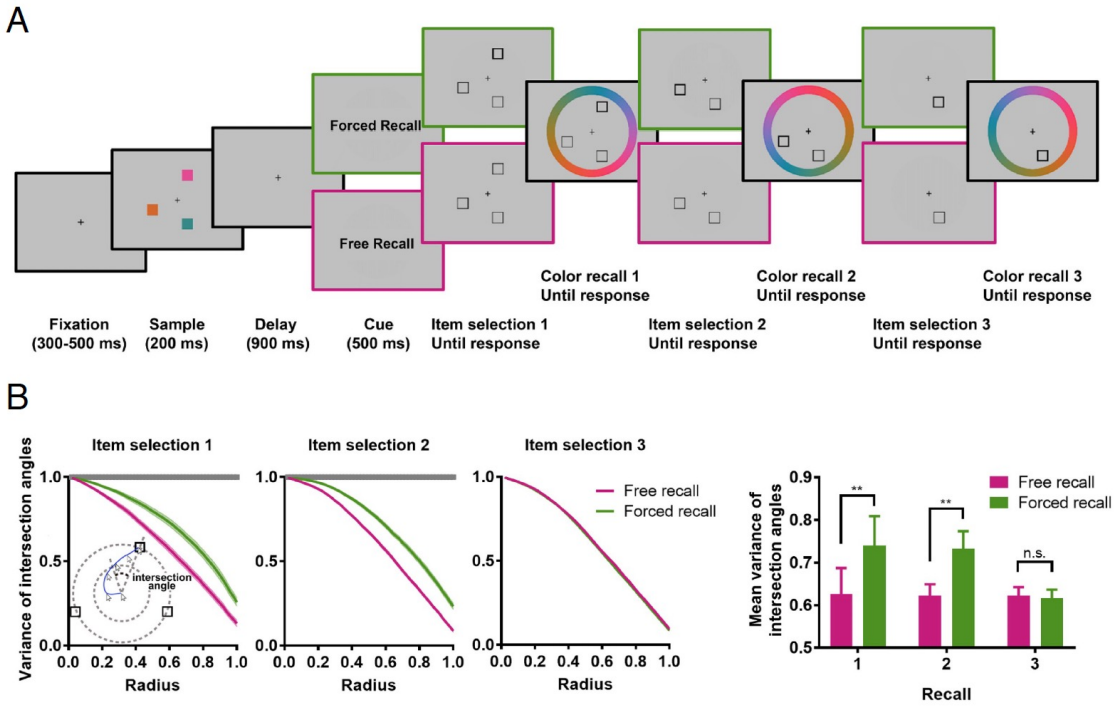
Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24-25.

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.

Figures

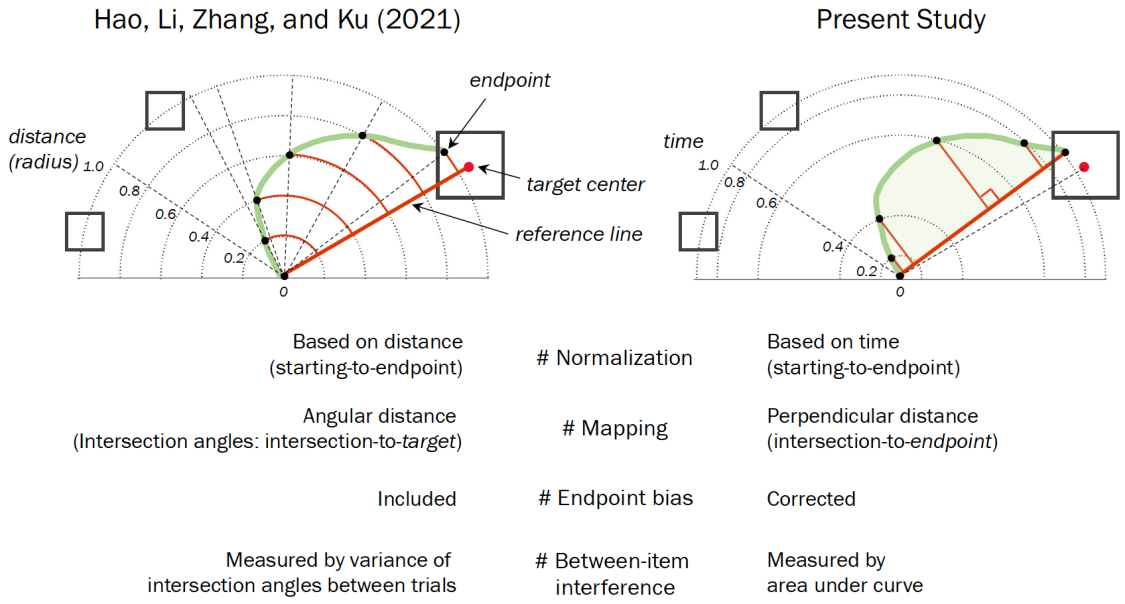
Figure 20.

Task procedure and primary results from Hao and colleague (2021).



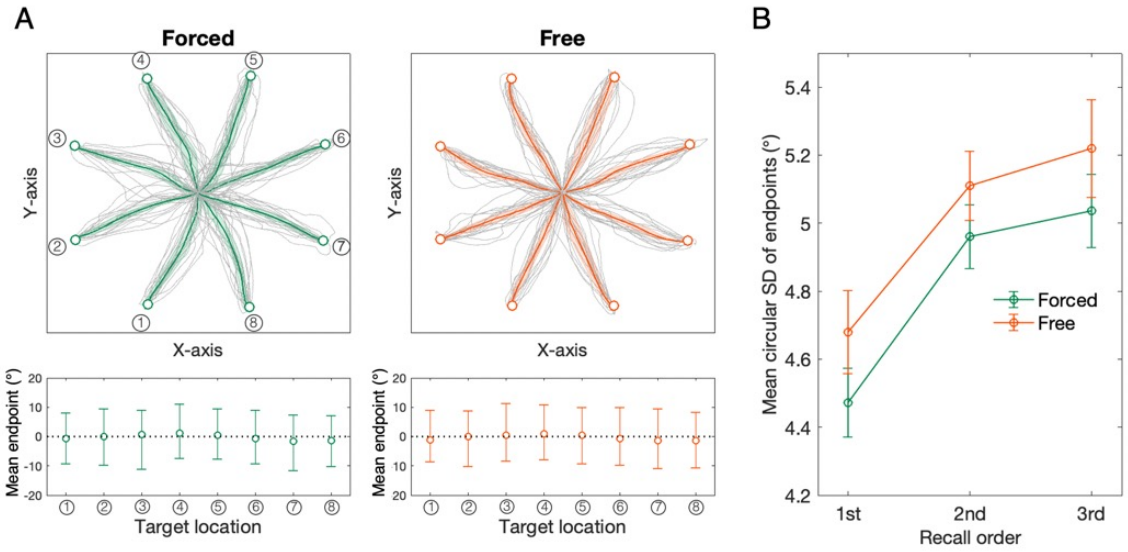
Note. Figures reconfigured from the original Figure 1 and 2. (A) Participants remembered three color squares briefly presented. After a short delay interval, a center cue appeared to indicate the recall type of the trial. Afterward, participants were asked to click the to-be-recalled item (item-selection; 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 1st, 2nd, and 3rd 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.

Figure 21.
Differences in the analytic method for the mouse trajectory data from the original and the present study.



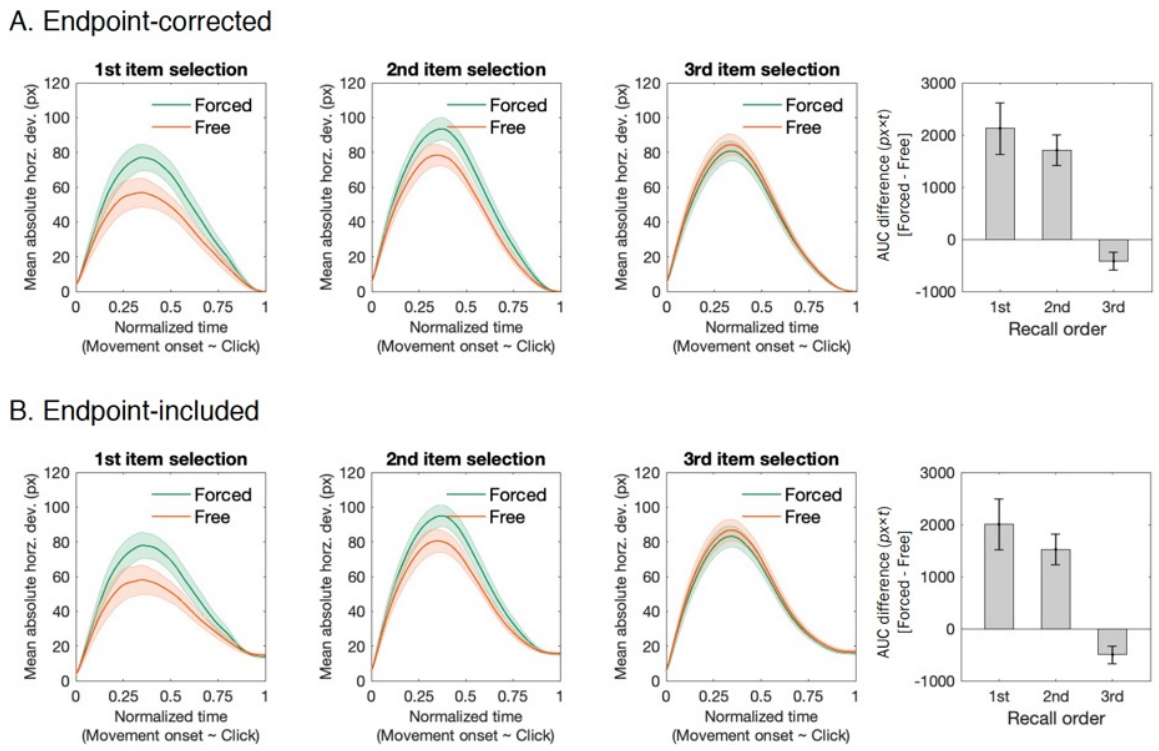
Note. The green curve is an example trajectory from the mouse movement onset at the display center to the mouse click endpoint, depicted in a reconstructed setting of the item-selection stage. During this stage, three placeholders appeared at the locations of the three memory items and the participants selected one for the next memory retrieval. Four key differences are summarized below: 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.

Figure 22.
Overall mouse trajectories of the item-selection stage.



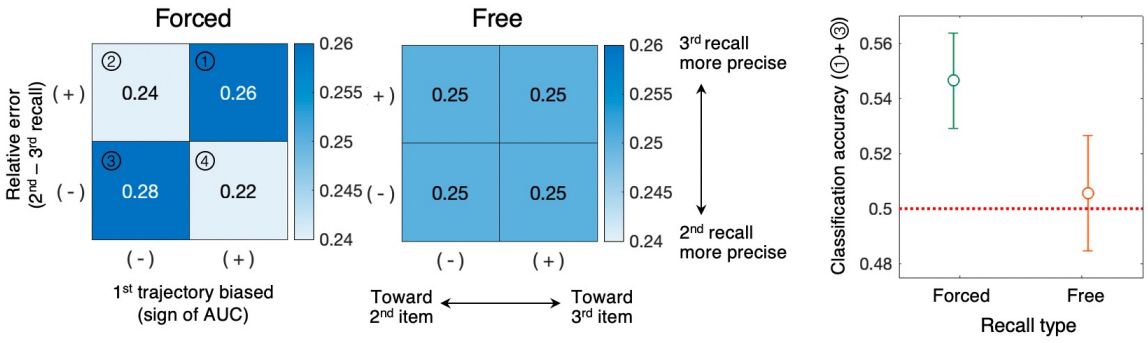
Note. (A) reconstructed item-selection mouse trajectories over eight possible target locations for forced and free-recall, collapsed across participants and selections for the three items. Colored curves and shades represent group mean trajectory and 95% CI, whereas grey curves are individual participant 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 ± 0.99°). (B) The mean and standard error of the endpoint variability measured by circular standard deviation (SD) as a function of recall order (1st, 2nd, and 3rd item-selection) and recall type (forced vs. free-recall). The endpoint variability was systemically larger under free-recall than forced-recall.

Figure 23.
Comparison of the mouse trajectory results between endpoint correction methods.



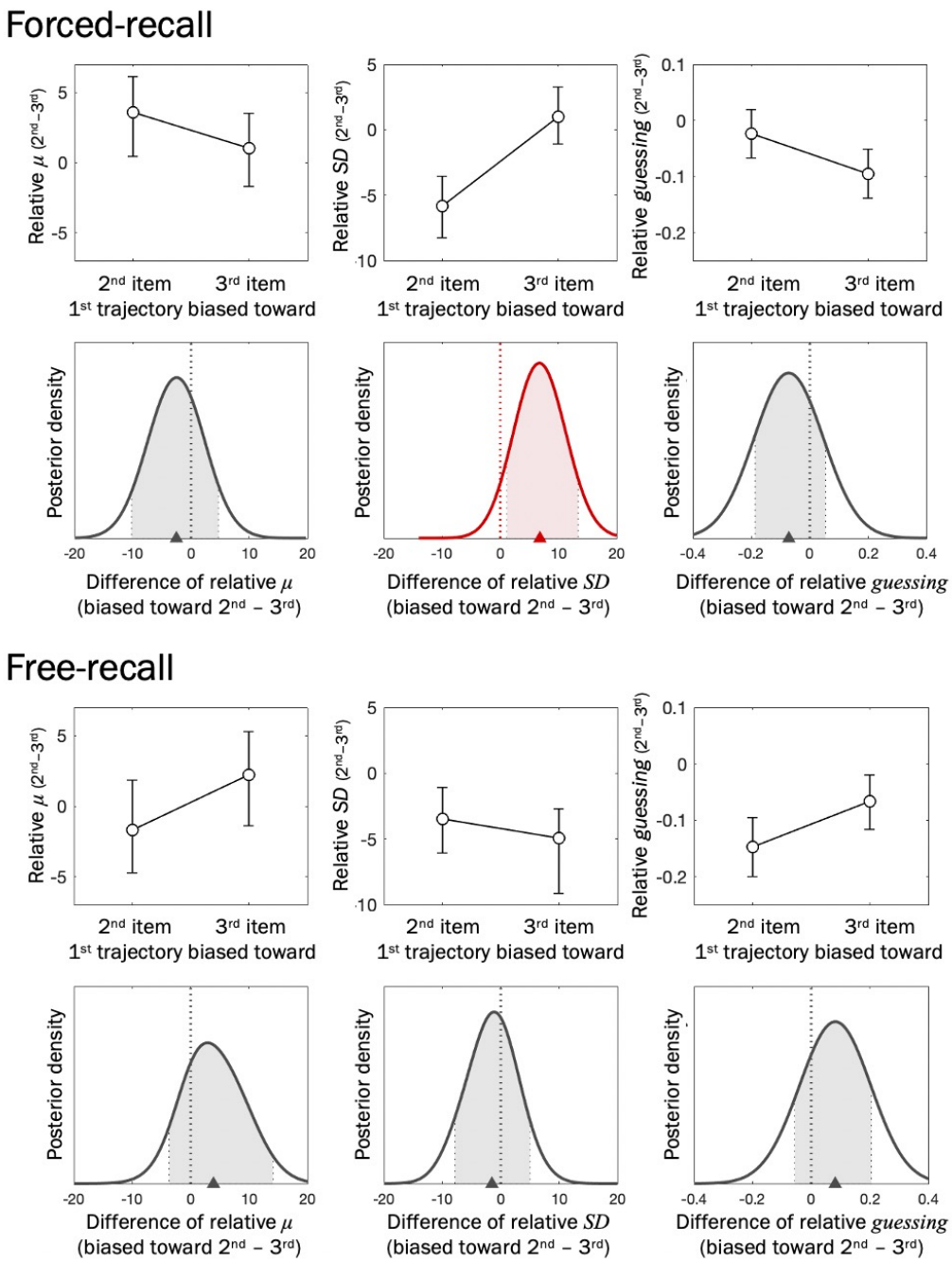
Note. Mean absolute horizontal deviations as a function of time normalized from the mouse movement onset to click endpoint, separately for recall type (forced vs. free-recall) and recall order (1st, 2nd, and 3rd item-selection), respectively. Either trial-specific endpoint (A) or target probe center (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.

Figure 24.
Trial-categorization results



Note. Proportion of trials as a result of trial-categorization consisting of a 2-by-2 matrix, where the horizontal dimension categorizes whether the 1st item-selection mouse trajectory deviated toward the 2nd or 3rd recalled item (by the sign of area under curve; AUC), whereas the vertical dimension categorizes by the sign of the relative absolute recall errors (2nd – 3rd; negative sign indicates the 2nd recall being more precise than the 3rd recall). The first (+, +) and the third quadrants (–, –) are consistent with the prediction that the direction of 1st item-selection mouse trajectory under forced-recall will be predictive of the relative precision of the 2nd and 3rd 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.

Figure 25.
The posteriors of the hierarchical Bayesian mixture model parameters.



Note. Hierarchical Bayesian posterior distributions of the three parameters (μ, *SD*, and *guessing*) from the extended Zhang & Luck (2008) Mixture model, representing the difference between 2nd and 3rd 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 (2^{nd} – 3rd recall) as a function of the 1^{st} item-selection trajectory bias direction (toward 2^{nd} or 3^{rd} item). The bottom row illustrates the difference between these two conditions (trajectory biased toward 2^{nd} – 3^{rd} 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 $HDI_{95\%}$. These resulting posteriors in other words represent the interaction effect between recall-order (2^{nd} and 3^{rd}) and trajectory bias (toward 2^{nd} and 3^{rd}). For only the precision parameter (*SD*) under forced-recall reveals its $HDI_{95\%}$ 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 gave a rise to the trial-specific directional mouse trajectory bias during the 1^{st} item-selection under forced-recall.

Chapter 6. μ Shift for Testing a Unit of Attentional Guidance by the Contents of Working Memory

6.1. Chapter Abstract

Working memory (WM) contents can guide attention towards matching sensory information in the environment. However, there are mixed findings regarding whether only a single prioritized item or multiple items held in WM can effectively guide attention. The present study aims to precisely examine the unit of WM-guided attention by using a novel task procedure and mouse trajectory analysis. Specifically, a perceptual-matching task was inserted into the maintenance interval of a WM task with the memory set size of one or two color squares. In this perceptual-matching task, a target color was presented at the screen center until it was matched on a continuous color-wheel with a mouse-click response. Participants' mouse cursor trajectories were recorded to estimate the moment-by-moment influence of WM on the subsequent perceptual process. We found that when a single item was remembered, the perceptual matching response and the mouse trajectory were robustly biased toward a location of the WM color on the color-wheel (i.e., attraction bias). When two items were remembered, the same trial-average measures did not show systematic bias toward either memory color. However, a closer examination of the mouse trajectories using hierarchical Bayesian modeling revealed two separable central peaks from the trajectory distributions for both memory set sizes. A novel trial-categorization method further revealed that the curved mouse trajectories, as a proxy measure of attentional guidance by WM items, are related to the precision of the memory representations. Together, these results support the single-item template account

and highlight the utility of mouse trajectory analyses in hypothesis testing in the domain of visual attention and WM.

6.2. Introduction

Working memory (WM) and attention are closely intertwined in their operations for guiding our adaptive behaviors. For example, they cooperate for selectively prioritizing relevant information, maintaining coherent internal representations across time, and prospectively modulating the process of upcoming sensory signals (Awh & Jonides, 2001; Oberauer, 2019; Nobre & Stokes, 2019). Grounded on the conceptualization of WM as a major source of top-down proactive attention (Desimone & Duncan, 1995; Gazzaley & Nobre, 2012), researchers have extended the idea of by which mechanisms selective attention gates access to WM (Chun & Potter, 1995; Mayer et al., 2007; McNab & Klingberg, 2008; Vogel et al., 2005) to how the contents of WM reversely guide where to attend (Hutchinson & Turk-Browne, 2012; Wolfe & Horowitz, 2004).

Studies have shown that visual information matching the WM content of the moment can automatically guide attentional selection, even when such attentional capture is detrimental to the immediate task goal (Downing, 2000; Olivers et al., 2006; Woodman et al., 2013). These studies typically employ a dual-task paradigm, where an independent attentional task such as a visual search is performed while holding a set of simple objects in WM. When distractors in the intervening attentional task match the representations held in WM, the search goes slower and/or inaccurate compared to when there are no WM-match distractors presented. Further studies found that the WM-driven search cost is

often accompanied by overt eye fixations over the matching distractors (Hollingworth et al., 2013; Houtkamp & Roelfsema, 2006).

Despite the large body of research on the prospective role of WM in attentional guidance, there still has been an ongoing debate over the *unit* of WM-guided attentional selection. A central question is how many discrete memory representations are active at a moment that in turn interact with ongoing attentional selection. Two major competing hypotheses were proposed in this regard. The single-item template (SIT) hypothesis proposes that only one item among memoranda can be in an active state and readily interact with attentional selection (Houtkamp & Roelfsema, 2006; Olivers et al., 2011). Substantial empirical evidence was reported in favor of the SIT hypothesis, for example, showing that multiple WM-match distractors in the secondary attention task do not produce any additional distractor interference, or even diminish the WM-match distractor effect due to mutual canceling of attentional biases as a product of competition between WM-match distractors (Van Moorselaar et al., 2014). On the contrary, other recent studies found evidence supporting the multiple-item template (MIT) hypothesis, which proposes that multiple representations, as long as successfully maintained in WM, can simultaneously serve as active attentional templates (Bahle et al., 2018; Beck & Hollingworth, 2017; Fan et al., 2022).

The discrepancy in the literature could be attributed to various factors. For instance, task paradigms and types of WM-matching distractors could determine whether or how multiple WM-match distractors interfere ongoing attentional processes (Frătescu et al., 2019). Moreover, a lack of precise measures of memory-driven capture could be

problematic (Beck & Vickery, 2019; Zhou et al., 2020). The majority of previous studies have primarily focused on the trial-average estimates such as mean reaction time (RT) or accuracy between experimental conditions (e.g., one vs. two WM-match distractors). Given that the WM load-dependent capture is thought to be a key feature in testing the SIT and MIT hypotheses, the trial-average single point estimates might not be sensitive enough to capture what actually happened when one, two, or none of the WM-match distractors captured attention at a given trial. Instead, such dynamics may vary across trials or even across items due to variable precision and interference between memoranda (Hao et al., 2021; Williams et al., 2022), which could only manifest in the trial-by-trial variability (Beck & Vickery, 2019; Park & Zhang, 2022). Lastly, here we claim that distractors in the subsequent visual search tasks that *exactly* match the perceptual appearance of the WM contents could be another experimental artifact. That is, the re-appearance of the preceded memory features may allow additional processes to play besides memory-driven capture, such as temporary suppression of mnemonic trace or even a second-chance for supplementary WM encoding.

To overcome these potential experimental and analytic limitations, the present study employed a novel dual-task paradigm in which a simple perceptual matching task inserted in the maintenance period of the delayed WM recall task. Participants first saw either one or two colored squares to remember (i.e., WM set size). Shortly after a blank interval, participants were presented with a perceptual matching array consisting of a colored circle at the screen center. This perceptual target color had different shape and no spatial overlap with the preceded memory items. Participants were instructed to simply

match this target color on a surrounding color-wheel by moving their mouse cursor from the center and clicking at the best-match color. We utilized the continuous estimation paradigm so that the WM-match sensory feature was not presented as an explicit distractor but only occupied a slice of the response window along a circular feature space. Notably, the perceptual target color was kept available on screen until the mouse click response was made. All of one or two remembered colors were then recalled in sequence at the end of the trial on color-wheel randomly rotated at each recall.

Critically, participants' mouse trajectory during the perceptual matching task was recorded to estimate the moment-by-moment influence of WM on the subsequent perceptual process. Following a well-established literature on the distractor interference effect manifested in movement trajectories via different motor actions (Song & Nakayama, 2009; Van der Stigchel, 2010), we operationalized a systematic WM-dependent attractive trajectory bias (i.e., deviated toward a location of the WM-match color on color-wheel) as a proxy measure of memory-driven capture. To normalize the strength of bias across trials and two set sizes, we manipulated a relative feature distance of the perceptual target from the memory items, either clockwise (CW) or counter-clockwise (CCW). For the delicate examination of the mouse trajectory data, here we introduced a novel analytic approach, *destination-vector transformation*, that leverages the moment-by-moment direction of trajectory progression. We then modeled their circular trajectory bias distributions at different time points using the extended Zhang and Luck (2008) mixture model implemented in a hierarchical Bayesian method (see *Method* for detail).

To test whether or how multiple WM representations simultaneously guide attention, we utilized the perceptual matching trajectory pattern under memory set size 1 as a baseline and compared it with the trajectories under set size 2. Distinct hypothesis-driven predictions are proposed for the perceptual matching mouse trajectory under memory set size 2. These predictions largely benefited from the trial-by-trial variability in the circular trajectory bias distributions (see Figure 26). For a summary, the SIT hypothesis predicts a *bi-modal* trajectory distribution with two dissociable central peaks at a given moment. The bi-modal distribution is hypothetically a mixture of two trajectory distributions, each of them are from distinct proportion of trials when attention is captured by a single WM-match color located either CW or CCW relative to the perceptual target. In other words, the CW and CCW memory-match colors would exclusively capture attention and attract mouse trajectory on a trial-by-trial basis. Across all trials, the set size 2 mouse trajectories should result in a bi-modal distribution that is nearly identical to the trajectory distribution from set size 1 when the trial-specific WM-match color location is ignored. On the contrary, the MIT hypothesis assumes both WM representations to be active and thus ready to guide attention at the moment of perceptual-matching task. Because they compete each other for selection, however, the attractive momentum from two WM-match colors will cancel each other out as a product of competition (Sogo & Takeda, 2007). Therefore, the MIT-based prediction holds no systematic and asymmetric trajectory bias toward one of the two WM-match colors over the other. Consequently, the trajectory distribution will be better described by a *uni-modal* function with its peak location centered at around the perceptual target color.

Lastly, we further sought to investigate whether asymmetric trajectory bias between WM-match distractors could possibly be accounted for by the variable precision across items. Consistent with a recent argument from Williams and colleagues (2022), the idea proposes that representational precision is a determinant factor for which WM items guide attention whereas others do not. Our recent work has provided a partial support for this idea using mouse trajectory data, where a trial-categorization method revealed that the relative recall performance between items could be predicted by the direction of mouse trajectory bias during the forced-selection of to-be-recalled item (Park & Zhang, 2022). However, this association was examined to explain between-item interference arising from inhomogeneity of WM precision rather than WM-driven attentional capture. Therefore, the present study adopted the same analysis to test if the curved mouse trajectories, as a proxy measure of attentional guidance, are related to the relative precision of WM representations. Note that the analysis exploits both the within-trial variability (i.e., trajectory bias and subsequent memory recall performance) and the between-trial variability (i.e., variability in recall performance). It is thus predicted that for both memory set sizes, the more precisely represented item will exhibit the stronger attractive trajectory bias during the subsequent perceptual matching task.

6.3. Method

6.3.1. Participants

Forty-five college students (19.1 ± 1.3 years old, 20 male) at the University of California, Riverside participated in the study for course credit. All participants reported

normal or corrected to normal visual acuity and normal color vision. Informed consent was obtained at the beginning of the experiment.

6.3.2. Stimuli and Procedure

Stimuli were presented on a 60 Hz LCD monitor (calibrated with a X-Rite I1Pro spectrophotometer) with a grey background (15.1 cd/m^2) at a viewing distance of 60 cm. Figure 27 illustrates the stimuli and procedure of the study. Each trial began with an 800 ms central fixation, followed by a sample array for 500 ms. The sample array consisted of either one or two colored squares (2.0° × 2.0° in visual angle). One memory color was randomly selected from a set of 180 evenly distributed on a circle in the CIRLAB color space (see Zhang & Luck, 2008 for details), and the other memory color for the set size 2 trials was picked at either ± 120° from the other. All colors had comparable luminance and varied mainly in hue and slightly in saturation. The locations of colors were randomly selected from a set of six possible locations equally spaced on an imaginary circle with a radius of 5.3° from the center. Participants were instructed to remember the one or two memory colors as accurately as possible.

After a short delay of 800 ms, the perceptual matching array appeared. It contained a colored circle presented at the center (2.0° × 2.0°), in addition to a continuous color-wheel (radius of 8.2°, thickness of 2.2°) with 180 evenly distributed colors. This perceptual target color had a relative distance of ± 60° to the memory colors. This made the perceptual target color in the middle of the memory colors for the set size 2 condition, whereas either clockwise (CW) or counterclockwise (CCW) for the set size 1 condition (hereafter, WM_{CCW} and WM_{CW} indicates memory-match color at -60° and +60° from the

perceptual target on color-wheel, respectively). This relative featural distance between items was unknown to participants. Participants were required to report the color of this target item using a computer mouse to click a matching color on the color-wheel. Note, the target item was kept visible until the mouse click response was made. The performance was measured as the error in degrees between the color chosen by the participant and the actual color shown in the sample array. Immediately after the recall response, a 500 ms online feedback was presented with a cross marking the color the participants picked and an arrow marking the actual color.

After another 800 ms delay, a memory test array followed. The test array consisted of outlined squares at the corresponding memory item locations and another color-wheel. One square was thicker than the other, indicating a probe item that needs to be recalled. For the trials of memory set size 1, only a single thick outline was presented. The color-wheel had a circular rotation of random degrees to prevent a strategical mapping of stimulus to color-wheel location during the previous perceptual matching array. Participants were again required to reproduce the remembered color of the probed item by clicking on the color-wheel as in the perceptual matching task, except that no colors were displayed on the screen. For those trials of memory set size 2, another test array was presented for the remaining memory item, with a second delay after the first memory recall. The response procedure was identical to the first memory recall and another randomly rotated color-wheel was provided. Visual feedback was provided for both recalls. All participants completed three blocks of 66 trials, yielding 198 trials in total. Trials for the memory set size 1 and 2 conditions were randomly mixed across trials

and took the one-third (66 trials) and two-thirds (132 trials) of total trials, respectively. For every mouse click response made on the color-wheel, the cursor position was reset to the screen center as a starting point. Participants' mouse trajectories were recorded.

6.3.3. Data Analysis

6.3.3.1. Mouse Trajectory

The trial-by-trial mouse trajectory during the perceptual matching task was first normalized in time, from its movement onset (0%) to a final click (100%). The movement onset was detected by cursor offset by three pixels in any direction from the center point set to (0, 0) in x-y coordinates. The target-match color location on color-wheel was rotated to be perpendicular to the x-axis (0, *r*), where *r* is a radius of color-wheel. The extent of trajectory deviation in all trials can thus be represented by area under curve (AUC), a definite integral of the horizontal cursor offsets over the course of trajectory. The sign of AUC indicates which direction trajectory deviated toward (i.e., which WM-match color on color-wheel), and the magnitude tells how strong WM-driven capture was. We also calculated maximum deviation (MD), the greatest offset along with trajectory curvature for sanity check. Participants' AUCs and MDs were highly correlated, $r(43) = .90$, 95% CI [.82, .94], $p < .001$. The main result pattern using the AUCs held the same when done with the MDs.

These measures of trajectory bias, however, are based on distance on the raw cartesian coordinates and possess inherent artifacts due to asynchronous variability in movement factors along time. For example, initial mouse cursor movements tend to deviate weakly in distance relative to mid-flight movements. This may cause under-

and/or over-estimation of trajectory bias, may even lead to a risk of misinterpretation in hypothesis testing. Imagine a hypothetical case wherein a mouse cursor is incorrectly directed straight to a distractor. Even though its movement target (i.e., distractor) is constant across time, the distance-based measures would show linearly increasing trajectory bias over time. Furthermore, AUC provides a single point estimate for a given trial that neglects the moment-by-moment dynamics in decision making.

In the present study, mouse movement was bounded within the circular feature space, therefore the direction of trajectory projected to the corresponding color on color-wheel is critical in testing the dynamics of WM-driven capture, especially when two items in WM compete each other for a selection. Here we benefited this circular nature of mouse movement destinations to characterize how the moment-by-moment moving direction of mouse trajectory tracks the influence of WM content on ongoing perceptual process. One simple way to map raw trajectory position onto meaningful feature space is to convert cartesian to polar coordinates with a starting position as a reference point (Hao et al., 2021). However, it comes with another limitation from its insensitivity of measuring the actual moving *direction* and its changes over time, especially when a movement target is subject to alter between attractors at work.

Here we developed a novel solution for the mapping of raw position onto underlying feature at the moment, *destination-vector transformation*, that focuses on the progression of movement targets based on two consecutive time points. Figure 28 demonstrates the example destination vectors in comparison with the cartesian-to-polar conversion output. Instead of computing angular position of the cursor at each moment,

the destination vectors compute a series of where on color-wheel the cursor headed toward from its previous position. As a consequence, this results in $N - 1$ data points in circular degree from N cursor positions varying in time. We treated this moment-by-moment circular bias as a proxy measure of WM-driven attentional capture. Furthermore, it produces a probability distribution of the circular bias across trials at each time point, of which the distributional shape can be critical to test competing hypotheses.

6.3.3.2. Hierarchical Bayesian Modeling of Trajectory Bias

To test whether or how multiple WM representations guided attention, we modeled the circular bias measures of mouse trajectory under a hierarchical Bayesian method. Two candidate models were constructed for the distribution of circular trajectory biases across trials. One model testing the MIT hypothesis is a *uni*-modal function, as an extended Zhang and Luck (2008) Mixture model, predicting the trajectory bias distribution to be a mixture of the uniform distribution and a single von Mises distribution (*VM*; circular analogue of the Gaussian distribution). The proportion of the uniform component (g) captures data points that are not driven by target-relevant process (i.e., randomly directed mouse cursor). In mouse trajectory data, a cursor occasionally wanders rapidly at early movements due to motor noise and thus needs to be separated. The von Mises distribution has two parameters, mean shift (μ) and variability (σ, converted from concentration κ). The present study has a primary interest of the μ shift parameter as it represents the systematic WM-dependent bias in mouse trajectory. This MIT-predicted uni-modal function is described by:

$$p(x|\mu) = (1 - g) \cdot VM(x; \mu, \kappa) + g(1/2\pi) \quad (1)$$

Another model testing the SIT hypothesis is a *bi*-modal function, consisting of a mixture of the uniform distribution (g) and two equal-variance (κ) von Mises distributions (VM_1 and VM_2). The two von Mises functions are dissociable by their central peak locations, μ_1 and μ_2, and designated to capture different proportions of trials with which the perceptual matching mouse trajectory deviated toward WM_{CCW} or WM_{CW} on color-wheel, relative to the perceptual target. According to the SIT hypothesis, when two items are remembered in WM, only one of the two WM-match colors will be guiding attention during the perceptual matching. Regardless of which WM-match color captured attention on a given trial, the trial-by-trial variability in trajectory bias will result in a bimodal distribution with two separable central peaks. In this model, two von Mises were equally weighted in summation based on assumptions that each WM representation is equally likely to be able to capture attention across trials. This SIT-predicted bi-modal function is described by:

$$p(x|\mu) = (1-g)\cdot[0.5\cdot VM_1(x;\ \mu_1,\kappa) + 0.5\cdot VM_2(x;\ \mu_2,\kappa)] + g(1/2\pi) \qquad (2)$$

Modeling and parameter estimation were done under the hierarchical Bayesian method. The hierarchical structure of variabilities and the choice of reasonably informative to non-informative priors were determined following previous studies with similar modeling approach (Oberauer et al., 2017; Park & Zhang, 2022). We took a total of 4,000 samples after 4,000 warm-ups from four chains of Markov chain Monte Carlo sampling. Statistical inference was made based on the mean and 95% credible interval (highest density interval, $HDI_{95\%}$) of the posterior distributions of the population-level parameters (Kruschke, 2014). To conduct a formal model comparison for our two

competing models with varying numbers of parameters, we computed the widely applicable information criterion (WAIC; Watanabe, 2010), which is a robust measure of goodness-of-fit for hierarchical models by taking the model complexity into account. A model with lower WAIC indicates a better fit to the observed data.

6.3.3.3. Trial-Categorization

We hypothesized that the WM-directed trajectory bias during perceptual matching would reflect variable precision across multiple WM representations. To test this we adopted a binary trial-categorization method, recently introduced by Park and Zhang (2021). Mouse movements could be subject to various motor noises that are either jittering alongside target-directed movement or even random wandering of the cursor. Nonetheless, the potential association between variable precision and asynchronous mouse trajectory bias across WM representations could still remain detectable when the effects are simplified into the binary codes.

In the present study, for example, every trial of the WM set size 2 condition could be categorized into a 2-by-2 binary matrix: 1) one dimension categorizes which side of the WM-match distractors (i.e., -60° CCW or +60° CW to a perceptual target) the perceptual matching mouse trajectory was deviated toward (e.g., by the sign of AUC; negative and positive values indicate WM_{CCW} and WM_{CW} capture trials, respectively). 2) The other dimension categorizes the relative recall performance between two WM items for the set size 2 condition (2^{nd} – 1^{st} absolute recall errors; positive value indicates 1^{st} recalled item was more precise than the 2^{nd} item). For the set size 1 condition, this second

dimension categorizes whether the absolute recall error was within or outside of the circular standard deviation of individual participant.

Importantly, the association between the direction of trajectory bias and the relative WM precision could manifest as an interaction effect in the proportion of trials at the *diagonal positions* over the 2-by-2 categorization matrix, where the signs of the two dimensions match each other (–/– and +/+), indicating a correspondence between the trajectory bias direction and the WM item more precisely reported. To make this diagonal interaction clear, the main effects were corrected by up- and down-weighting a half-size of the difference between rows and columns. As a consequence, the null trial proportions across the quadrant cells can be equated at 25%, and thus a statistical significance of the interaction effect can be tested by a one-sample *t*-test for the proportion of the diagonal cells (–/– and +/+) to 50% null proportion.

6.4. Results

Overall performance of the continuous reports for perceptual matching and WM recalls was estimated by the Zhang and Luck (2008) mixture model *sd* parameter, representing imprecision of reports (Figure 29). Perceptual matching performance was statistically comparable between the WM set size 1 (11.1° [$CI_{95\%}$ hereafter: 10.3, 12.0]) and set size 2 (11.5° [10.7, 12.2]) conditions, $t(44) = -0.95$, $p = .350$, Cohen's $d = -0.14$. WM recall performance showed the set size effect and the recall order effect. The 1st WM recall was more precise for the set size 1 (17.1° [16.2, 18.1]) than for the set size 2 (20.2° [19.2, 21.3]), $t(44) = -5.71$, $p < .001$, Cohen's $d = -0.86$. In the set size 2 condition, the

2nd recall (25.2° [23.9, 26.6]) was less precise than the 1st recall, $t(44) = 10.66$, $p < .001$, Cohen's $d = 1.61$.

Of the primary interest, we focus on performance for the perceptual matching task inserted in the middle of the WM maintenance period. First, to see how the conventional WM-driven attentional capture manifested in our novel paradigm of perceptual matching on a continuous scale, we analyzed reaction time (RT) and mouse onset latency calculated from the trajectory data as a function of WM set size (Figure 30). On one hand, overall mean RTs were comparable between the WM set size 1 (2196.4 ms [1961.1, 2416.5]) and set size 2 conditions (2285.0 ms [2049.5, 2520.4]), $t(44) = -1.47$, $p = .149$, Cohen's $d = -0.22$. On the other hand, however, mean movement onset latency for WM set size 2 condition (823.1 ms [761.0, 885.2]) was significantly slower than set size 1 (657.2 ms [599.3, 715.0]), $t(44) = 10.08$, $p < .001$, Cohen's $d = 1.52$. The same pattern was observed with the normalized movement onset latency in percentage of RT in each trial (set size 1: 32.4% [29.1, 35.6] vs. set size 2: 38.9% [35.5, 42.4]; $t(44) = -5.55$, $p < .001$, Cohen's $d = -0.84$). It suggests a trade-off between movement onset latency and 'in-flight' duration, in a way that delayed response selection was compensated by expedited mouse cursor movement and clicking response. Together, these results replicate the previous mixed findings at different aspects of response speed measures. An extra WM-match distractor did not produce additional distractor interference on the overall perceptual matching RT as predicted by the SIT hypothesis, whereas delayed the initiation of mouse movement, potentially due to additional competition between multiple

WM-match distractors for a selection as predicted by the MIT hypothesis. Yet, it is not clear whether the observation is simply due to task difficulty.

We next analyzed the perceptual matching mouse trajectories, separately for WM set size 1 and 2 trials. Figure 31 illustrates the horizontal trajectory deviations, as well as the circular biases calculated from the destination-vector transformation along the normalized time from movement onset to final click (see *Method* for detail). For the WM set size 1 condition, there was a robust, systematic trajectory bias during perceptual matching toward the location of the single WM-match color appeared on color-wheel (i.e., attraction bias), either at CW or CCW to the perceptual target color, respectively. Specifically, a paired *t*-test revealed a significant difference in the mean AUCs between the CW (27.9 px/t [19.5, 36.2]) and the CCW conditions (-29.2 px/t [-37.3, -21.1]), $t(44) = -7.20$, $p < .001$, Cohen's $d = -1.08$, confirming a WM-dependent attentional capture. When the sign of the CCW trial AUCs were flipped to be matched with the CW trial AUCs, their magnitudes were found comparable each other, $t(44) = 0.46$, $p = .646$, Cohen's $d = 0.07$, further supported by a strong correlation in individual participants' AUCs between CW and CCW conditions, $r(43) = .78$ [.63, .87], $p < .001$.

When these raw trajectory coordinates are converted to the circular bias estimates at every 10^{th} quantile along the normalized time, the degree to which the reconstructed color feature (i.e., movement target on color-wheel at the moment) was biased from the perceptual target was most evidence in the first 10% of movement time, then corrected over time toward the perceptual target. These circular biases at every time point were again systematically shifted toward the WM-match color, yielded significant difference

between CW and CCW trials, ps < .001, Cohen's ds > 0.77, even at the last moment of the perceptual target report, $t(44) = 6.41$, $p < .001$, Cohen's $d = 0.97$.

It is important to note, however, that these trial-average measures of mouse trajectory are inadequate to test the predictions of the SIT and MIT hypotheses for the perceptual matching mouse trajectories under WM set size 2. Only the set size 1 condition had a condition code for the location of a single WM-match color, whereas in the set size 2 condition both WM colors appeared at CW and CCW to a perceptual matching target on color-wheel. In fact, when the condition code for WM-match location is ignored from set size 1, the trial-average AUCs were comparable between set size 1 (-1.4 px/t [-3.8, 1.0]) and set size 2 (0.6 px/t [-3.4, 2.1]) conditions, $t(44) = 0.44$, $p = .662$, Cohen's $d = 0.07$. The same pattern was observed from the circular bias measures, the set size 1 and 2 trials were not significantly different at all 10 time points, ps > .288, Cohen's $|d|$s < 0.16.

For a closer examination of the mouse trajectories, we focused on the trial-by-trial variability in trajectory and tested the distinct predictions for the shape of the circular bias distributions, derived from the SIT and MIT hypotheses each predicts a bi-modal and a uni-modal function, respectively (see Figure 32). For the first step, the set size 1 circular biases at every 10^{th} normalized movement time without their condition codes of WM_{CW} or WM_{CCW} location were successfully modeled by the bi-modal function, where the two central peak locations could be credibly dissociated. It is not surprising given the meaning of their condition-coded trajectories we already examined (Figure 31A). In the next step, more importantly, we applied the competing models (i.e., *bi-modal SIT* vs. *uni-*

modal MIT) to the circular bias distributions from set size 2 trials. A formal model comparison revealed that the bi-modal SIT model (WAIC: 217,696) robustly outperformed the uni-modal MIT model (240,587), with WAIC difference of 22,891 (9.51% in WAIC ratio). Figure 33 summarizes the hierarchical Bayesian posteriors of the two μ shift parameters, μ_1 and μ_2, in our SIT model, at each of the 10th mouse movement time. The overall pattern of the μ shift posteriors show a remarkable resemblance between WM set sizes, with the most credible separation of $\mu_{gap}\ [\mu_2 - \mu_1]$ at those early mouse movements. Furthermore, individual participants' $\mu_{gap}\ [\mu_2 - \mu_1]$ means between the set size 1 and 2 conditions were strongly correlated at all time points, rs(43) > .85 [.74, 91], ps < .001. These results suggest evidence strongly favored by the SIT hypothesis. In other words, even when two items are remembered in WM at a given trial, only one representation could effectively capture attention and guide mouse trajectory toward it.

Finally, we tested whether such imbalanced trajectory bias exclusively toward a single WM-match distractor could be accounted for by WM precision. The trial-categorization was conducted separately for the WM set size 1 and 2 conditions, over the two-by-two matrix consisting of the trajectory bias direction (AUC toward WM_{CW} vs. WM_{CCW}) and the WM recall precision (for set size 1: within- vs. over-circular *SD*; for set size 2: 2nd – 1st absolute recall errors). Figure 34 illustrates the resulting proportion of trials across the quadrants (1st: +/+, 2nd: –/+, 3rd: –/–, and 4th: +/–) and the categorization accuracy (1st + 3rd quadrants) for both set sizes. For set size 1 trials, we found the categorization accuracy at 61.6% [58.5, 64.7%], significantly greater than the chance-level at 50%, one-tailed one-sample $t(44) = 7.39$, $p < .001$, Cohen's $d = 1.11$, suggesting

that WM-directed trajectory bias was associated with how precise the memory item was represented in WM. Similar pattern was obtained for set size 2 trials, the categorization accuracy was 53.7% [51.3%, 56.1%], significantly above the chance-level at 50%, one-tailed one-sample $t(44) = 3.03$, $p = .002$, Cohen's $d = 0.46$. This suggest that although the numerical effect size was decreased from the set size 1 trial categorization, the relative recall performance between WM items could be predicted by the direction of mouse trajectory bias during the preceded perceptual matching task.

6.5. Discussion

A debate has continued on whether multiple representations held in WM can concurrently guide attention or only a single representation in an active state can do so while others cannot. These MIT and SIT hypotheses both have empirical supports and the mixed findings are thought to originate from the different experimental factors across studies (Frătescu et al., 2019). The present study investigated the unit of WM-guided attention with a novel method of estimating attentional bias by WM-match information on the secondary perceptual matching task. In particular, we examined the mouse trajectory to assess the moment-by-moment influence of the WM contents on the subsequent perceptual process.

Three major findings are reported as follows. First, remembering a single color item caused a robust WM-dependent attraction bias in mouse trajectory during the subsequent perceptual matching task. The trajectories were deviated systematically toward a side of WM-match color on a continuous color-wheel response window. We introduced a novel destination-vector transformation that reconstructs the underlying

feature value based on a cursor movement target on the response window on a moment-by-moment basis. These converted circular trajectory bias revealed that the extent of WM-toward trajectory bias was at peak during early moments after cursor movement onset, then corrected toward the original perceptual target. This corrective movement shortly after initial movement toward distractor is consistent with the previous studies using manual reaching (Song & Nakayama, 2008) or saccadic eye-movements (McPeek et al., 2000). Second, for the critical testing of the SIT and MIT hypotheses, the overall pattern of the set size 1 mouse trajectory served as a baseline and compared with that of the set size 2 trajectories. A closer examination of the circular trajectory bias distributions with the hierarchical Bayesian modeling and formal model comparison provided convincing evidence in favor of the SIT hypothesis. In particular, we identified two credibly separable central peaks from the trajectory distributions for both set sizes, where each peak distribution represents a half proportion of trials either WM_{CW} or WM_{CCW} color on color-wheel captured trajectory. Separate model fits for each WM set size condition revealed a remarkable resemblance in the estimated separation of the μ shift parameters in between, throughout every 10^{th} movement time points. These results indicate that even when two items are remembered, only one WM representation was able to capture attention and guide mouse cursor movement. Lastly, we showed a partial support for the idea that attentional guidance by the WM contents and their representational fidelity are not independent constructs. A novel trial-categorization adopted from Park and Zhang (2022) identified this interaction effect, statistically reliable though numerically weak, between the direction of trajectory bias and the subsequent

WM recall performance, manifested in diagonal cells on the two-by-2 categorization matrix.

The present study speaks to some important notions in the literature of WM-guided attention in several ways. We attributed the discrepancy in the discussion of the SIT and MIT hypotheses to two major points, experimental design and precision of estimates. In typical dual-task paradigms, memory-match distractors in the secondary visual search task are the exact copies of the memory items. Though it helps granting a robust guidance effect, it may call for other processes such as global feature suppression or encouraging supplementary memory encoding. Here we instead introduced the perceptual matching task in which WM-match sensory features are not presented as explicit distractors to be suppressed but only occupied a slice of circular response window along with all color palettes. Therefore, it serves a more natural setup for testing any influence of the task-irrelevant WM-match visual information on independent attentional process, with minimal differences in task demand over distractors and the number of distractors across studies. In addition, it is noteworthy that a perceptual target color in our paradigm was always in view until it was matched on the color-wheel by a mouse clicking response. We nevertheless observed a strong guidance effect by WM-match color on the mouse trajectory pattern. This grants the validity of our paradigm and adds to the existing literature on the robustness of WM-driven capture (Olivers et al., 2006; Soto et al., 2005).

Moreover, the discussion of SIT and MIT has tapped into a simple comparison of trial-average estimates of attentional capture. Regardless of the types of measure whether

it is WM load-dependent cost in RT or oculomotor capture, this approach neglects the trial-by-trial variability that may otherwise convey essential information of whether one or more WM representations or even none of them guided attention on a particular trial (see for a similar notion, Beck & Vickery, 2019). Instead, the present study leveraged the distributional properties of mouse trajectory biases and modeled them after distinct predictions from the competing hypotheses under the hierarchical Bayesian method.

Considering a growth of interests on the role of mnemonic precision in determining whether a certain WM representation is sufficient to guide attention (Olivers et al., 2011; Williams et al., 2022), the dichotomy of *single* versus *multiple* in current theories might be an oversimplification of the dynamics in WM-guided attention that vary trial to trial and item to item. The present study adds a partial support to this notion from the trial-categorization results showing that the relative precision among the concurrent WM representations could somewhat predict which side of WM-match color the mouse trajectory would deviate toward during independent task. Even for set size 1, we found a statistically reliable association between recall precision for the single remembered item and whether the preceded perceptual mouse trajectory was deviated toward the WM-match color or not. This weighs the idea beyond the division of active versus latent states (Williams et al., 2022). However, we only tagged binary codes for each trial based on the sign of measures, therefore cannot answer the gradient nature of precision-dependent WM guidance of attention. In fact, we failed to see a meaningful pattern when we attempted more complex trial-categorization matrix, 3-by-3 or 4-by-4 matrices, to take the strength of estimates into account. Considering numerically weak

categorization accuracy we obtained from binary codes, the failure of the strength-based account could possibly be due to random noise in trial-level measures. Further study would be needed to develop more precise and strength-sensitive trial-level marker of WM-driven trajectory bias.

To the best of our knowledge, this study is the first to investigate the nature of WM-guided attention using mouse trajectory. Mouse trajectory data contains common features with other goal-directed movement trajectories in different modalities such as saccadic eye-movements (Van der Stigchel, 2010) and manual reaching (Song & Nakayama, 2009) providing continuous measures before the final choice is made. Compared to saccadic movements, on the other hand, mouse trajectory can provide more reliable trajectories and thus have relative benefits in a read-out of internal cognitive decisions against conflicts (Song & Nakayama, 2009). In addition, the mouse trajectories in the present study were specifically designed to be bounded within a circular feature space (i.e., color-wheel response window). This allowed us to reconstruct underlying sensory features through a series of movement targets across moments (Spivey et al., 2005; but see, van der Wel et al., 2009) by the destination-vector transformation we proposed.

A caveat of mouse trajectory, however, is that a measurement window could be much lagged compared to saccades. Importantly, this leads to a caution that the current findings do not necessarily mean that only a single WM representation can exclusively guide attention throughout the entire attentional process. Alternatively, it is possible that multiple WM representations are maintained with overall comparable accessibility but

being cycled in and out of the active and latent states (Barrouillet et al., 2011; Camos et al., 2018; Gilchrist & Cowan, 2011). We found a load-dependent delay in movement onset latency, though the overall RTs were comparable between memory set sizes. One would not expect such delay if a certain WM representation was exclusively in an active state at the onset of perceptual matching task. Although it could totally be driven by general task difficulty increased with memory set size, another theoretically plausible account would be that a competition occurred between memoranda for a selection but then resolved before the motor response began. The guidance by this 'winning' WM representation was then followed by inhibition and during ongoing decision making and motor response, became detectable by our mouse trajectory recording. Further examination of mouse-tracking in conjunction with a concurrent eye-tracking method would be recommended for testing these theoretically important questions in future work (see Koop & Johnson, 2013).

Taken together, the present study took several novel steps toward resolving an ongoing debate on the unit of attentional guidance by WM. A thorough examination of the mouse trajectory data through computational modeling of trial-by-trial variability concludes that, at least for its manifestation at the behaviorally meaningful level, a single WM item can guide attention and interact with the subsequent perceptual processes. In line with growing consensus on the dynamic interaction between motor action and ongoing cognitive processes (Olivers & Roelfsema, 2020; van Ede et al., 2019), the current work further sheds light on the utilization of mouse trajectory in critical hypothesis testing in the domain of visual attention and WM.

References

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119-126.

Bahle, B., Beck, V. M., & Hollingworth, A. (2018). The architecture of interaction between visual working memory and visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 992-1011.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175-192.

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., & Vickery, T. J. (2019). Multiple states in visual working memory: Evidence from oculomotor capture by memory-matching distractors. *Psychonomic Bulletin & Review*, *26*(4), 1340-1346.

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory?. *Annals of the New York Academy of Sciences*, *1424*(1), 19-32.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental psychology: Human Perception and Performance*, *21*(1), 109-127.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193-222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467-473.

Fan, L., Diao, L., Xu, M., & Zhang, X. (2022). Multiple representations in visual working memory can simultaneously guide attention. *Current Psychology*, Advance online publication.

Frătescu, M., Van Moorselaar, D., & Mathôt, S. (2019). Can you have multiple attentional templates? Large-scale replications of Van Moorselaar, Theeuwes, and Olivers (2014) and Hollingworth and Beck (2016). *Attention, Perception, & Psychophysics*, *81*(8), 2700-2709.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129-135.

Gilchrist, A. L., & Cowan, N. (2011). Can the focus of attention accommodate multiple, separate items?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1484-1502.

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 low-level saccade target selection: Evidence from rapidly generated saccades in the global effect paradigm. *Journal of Vision*, *13*(13), 4.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 423-442.

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, *16*(12), 576-579.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, *67*(4), 151-185.

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York, NY: Academic Press.

Mayer, J. S., Bittner, R. A., Nikolić, D., Bledowski, C., Goebel, R., & Linden, D. E. (2007). Common neural substrates for visual working memory and attention. *Neuroimage*, *36*(2), 441-453.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103-107.

McPeek, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research*, *40*(18), 2499-2516.

Nobre, A. C., & Stokes, M. G. (2019). Premembering experience: A hierarchy of time-scales for proactive attention. *Neuron*, *104*(1), 132-146.

Oberauer, K., Stoneking, C., Wabersich, D., & Lin, H. Y. (2017). Hierarchical Bayesian measurement models for continuous reproduction of visual features from working memory. *Journal of Vision*, *17*(5), 11.

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

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243.

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.

Park, H. B., & Zhang, W. (2022). Trial-by-trial mouse trajectory predicts variance in precision across working memory representations: A critical reanalysis of Hao et al.(2021). *Psychonomic Bulletin & Review*, Advance online publication.

Sogo, H., & Takeda, Y. (2007). Saccade trajectory under simultaneous inhibition for two locations. *Vision Research*, *47*(11), 1537-1549.

Song, J. H., & Nakayama, K. (2008). Target selection in visual search as revealed by movement trajectories. *Vision Research*, *48*(7), 853-861.

Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360-366.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248-261.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393-10398.

Van der Stigchel, S. (2010). Recent advances in the study of saccade trajectory deviations. *Vision Research*, *50*(17), 1619-1627.

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 Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477-483.

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.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2005). Pushing around the locus of selection: Evidence for the flexible-selection hypothesis. *Journal of Cognitive Neuroscience*, *17*(12), 1907-1922.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571–3594.

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*. Advance online publication

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it?. *Nature Reviews Neuroscience*, *5*(6), 495-501.

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention?. *Journal of Vision*, *13*(3), 1.

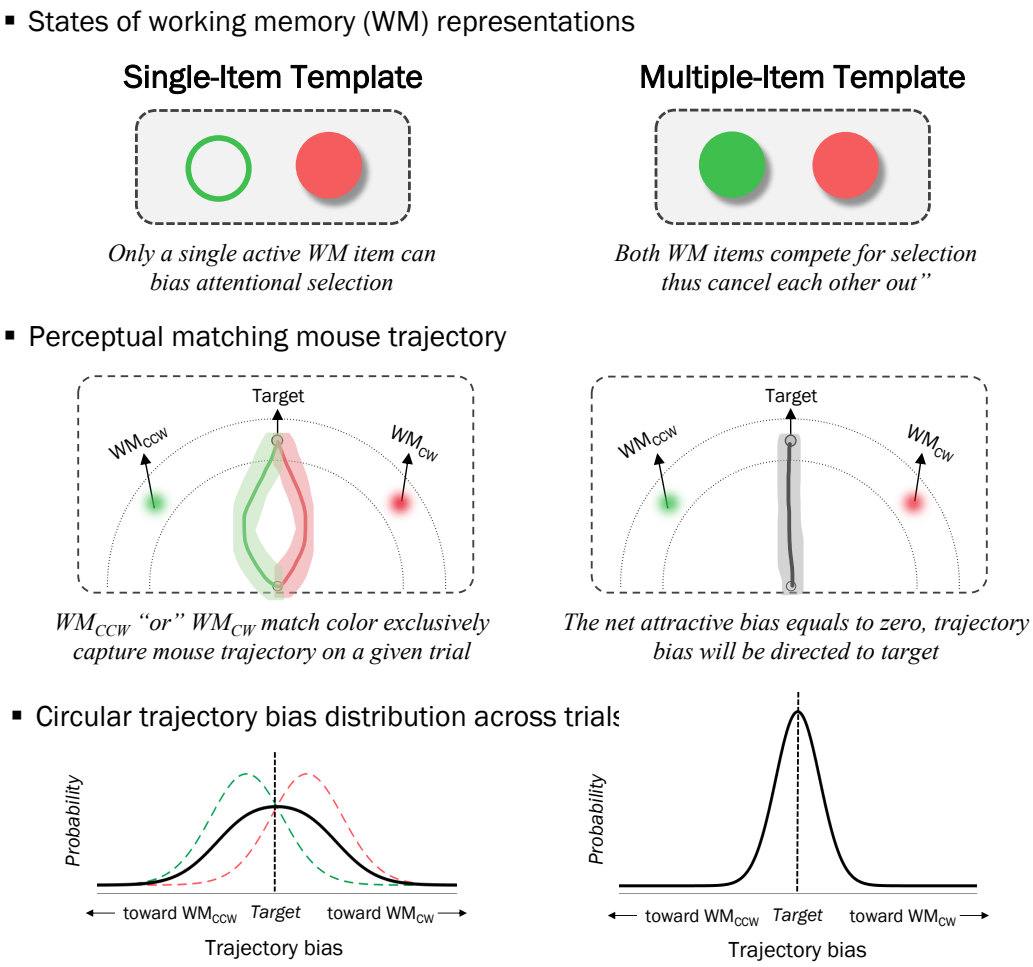
Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233-235.

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.

Figures

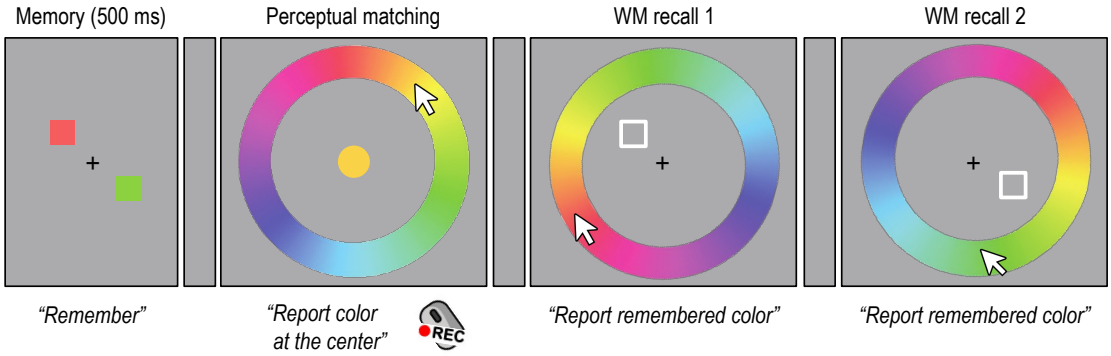
Figure 26.

Predictions for mouse trajectory based on the single-item template (SIT) and the multiple-item template (MIT) hypotheses.



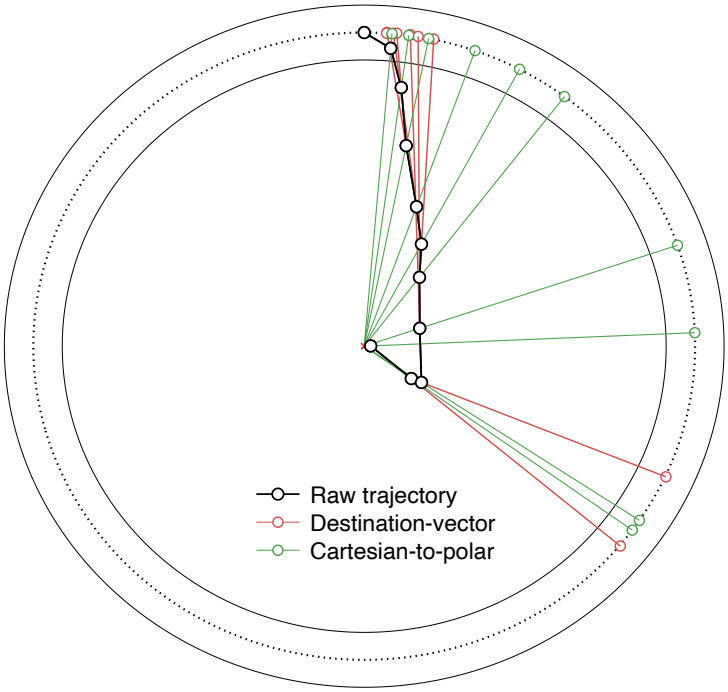
Note. The SIT hypothesis proposes that only a single item among two WM representations can be in a focus of attention and thus able to guide attention. This predicts that perceptual mouse trajectory on a given trial will be deviated toward a particular WM-match color that is either clockwise (WM_{CW}) or counterclockwise (WM_{CCW}) to a target color on a color wheel response window. Across over trials, the trajectory bias distribution could be best characterized by a bi-modal function. Contrarily, the MIT hypothesis proposes that both WM items are active and thus compete for attentional selection. The competition cancels out attractive biases and makes the net attraction effect at zero on average. Consequently, the trajectory bias will be best characterized by a uni-modal function centered around the straight target-directed trajectory.

Figure 27.
Stimuli and procedure of the working memory and perceptual matching dual task.



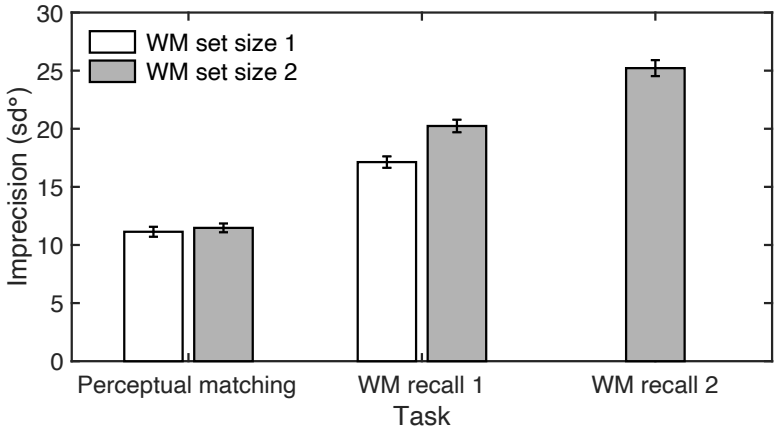
Note. Either one or two colored-squares were presented for memory items. After a short delay, a secondary perceptual matching task followed. A perceptual target colored-circle was presented at the display center, with no spatial overlap with the preceded memory items. Participants matched a target color on a surrounding color wheel. The perceptual target color had a fixed feature distance of 60° from the WM colors. Mouse trajectory during the perceptual matching response was recorded. After response, either one or two working memory (WM) recall arrays followed to test memory item(s). The three color wheels in sequence were randomly rotated to prevent strategic mapping of stimulus to color-wheel location.

Figure 28.
Example of destination-vector transformation of mouse trajectory.



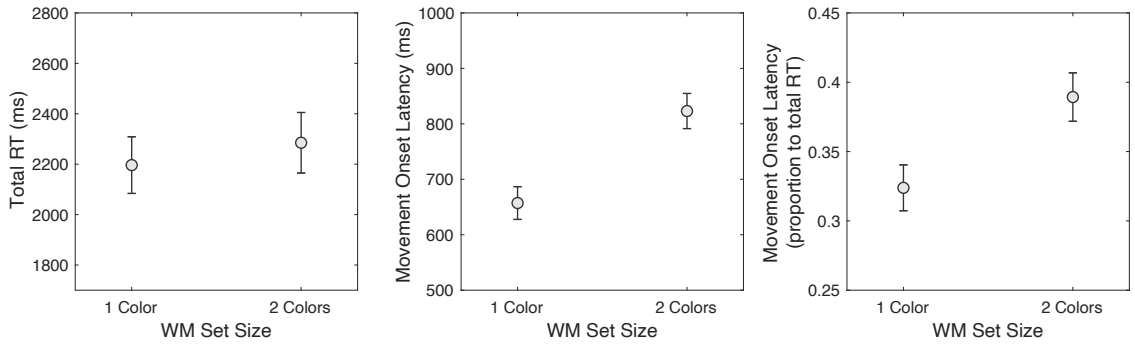
Note. A solid black line and circle markers represents an actual example mouse trajectory (participant #1, trial #19) at every 10th normalized time points. The surrounding big circles indicate a color wheel contour (solid lines) and a centerline (dotted-line). A radius from the center point to the color wheel centerline was used for angular transformations. Red lines are the calculated destination vectors reflecting the moment-by-moment mouse cursor movement targets along the color wheel centerline, based on two consecutive time points. For a comparison, green lines are the conventional cartesian-to-polar transformation results.

Figure 29.
Imprecision of the reports for perceptual matching and working memory recall tasks.



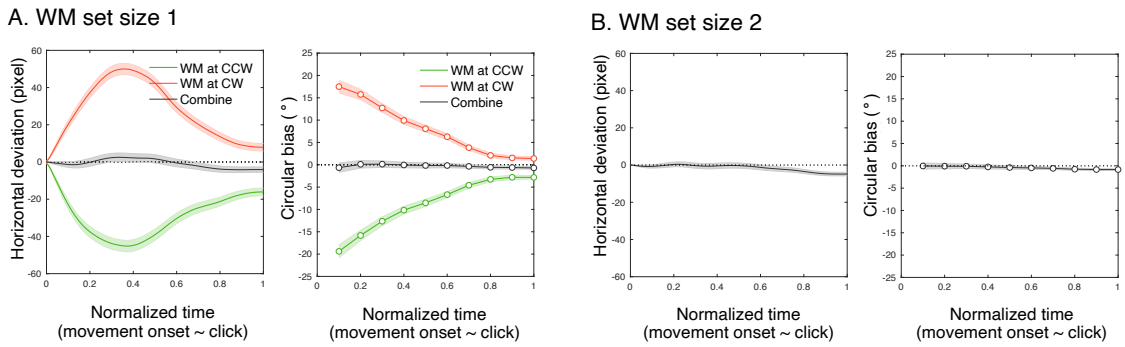
Note. Estimation errors for each type of reports were fitted to the Zhang and Luck (2008) mixture model. The imprecision estimates on the y-axis is the *sd* parameter of the model reflecting the width of the error distribution. Error bars represent standard error of mean.

Figure 30.
Mean reaction times and mouse onset latencies from perceptual matching task.



Note. Error bars represent standard error of mean.

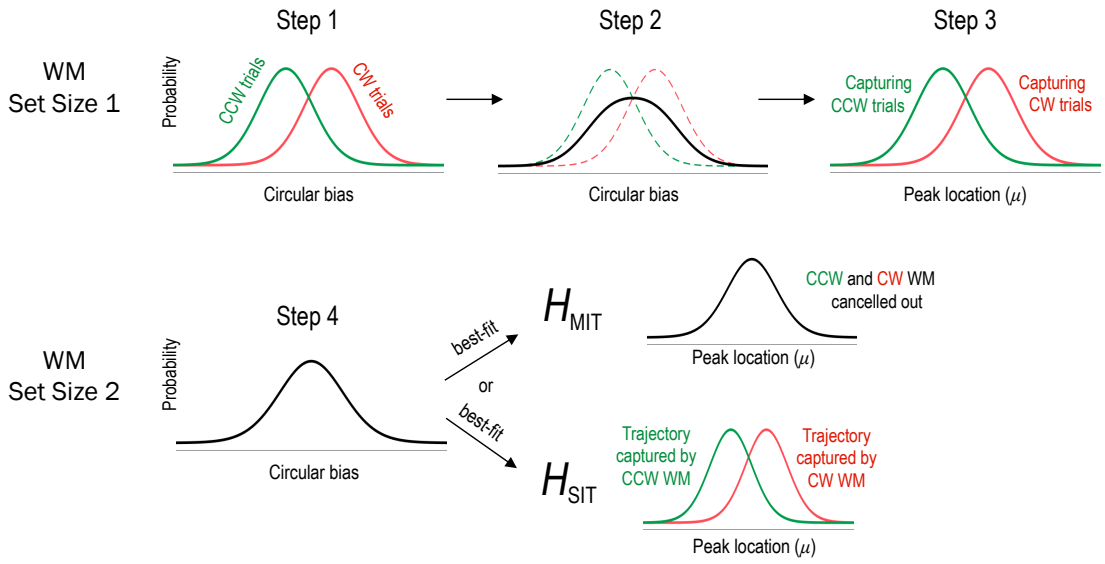
Figure 31.
Perceptual matching mouse trajectory results.



Note. The horizontal deviation of the mouse cursor positions (left in panel) and the circular bias calculated from the destination vector transformation (right in panel) as a function of the normalized time between movement onset and final click response. (A) Working memory (WM) set size 1 trajectory results. Green and Red curves indicate the mean and the standard error of mean of the measures from the counterclockwise (WM at CCW) and clockwise WM-match color (WM at CW) on color wheel, respectively. Gray curves are all trials combined between WM CW and WM CCW trials. (B) The same measures from WM set size 2 trials. In the set size 2 condition, the color wheel contained both WM-match colors. Therefore, there was no condition code (e.g., CW or CCW).

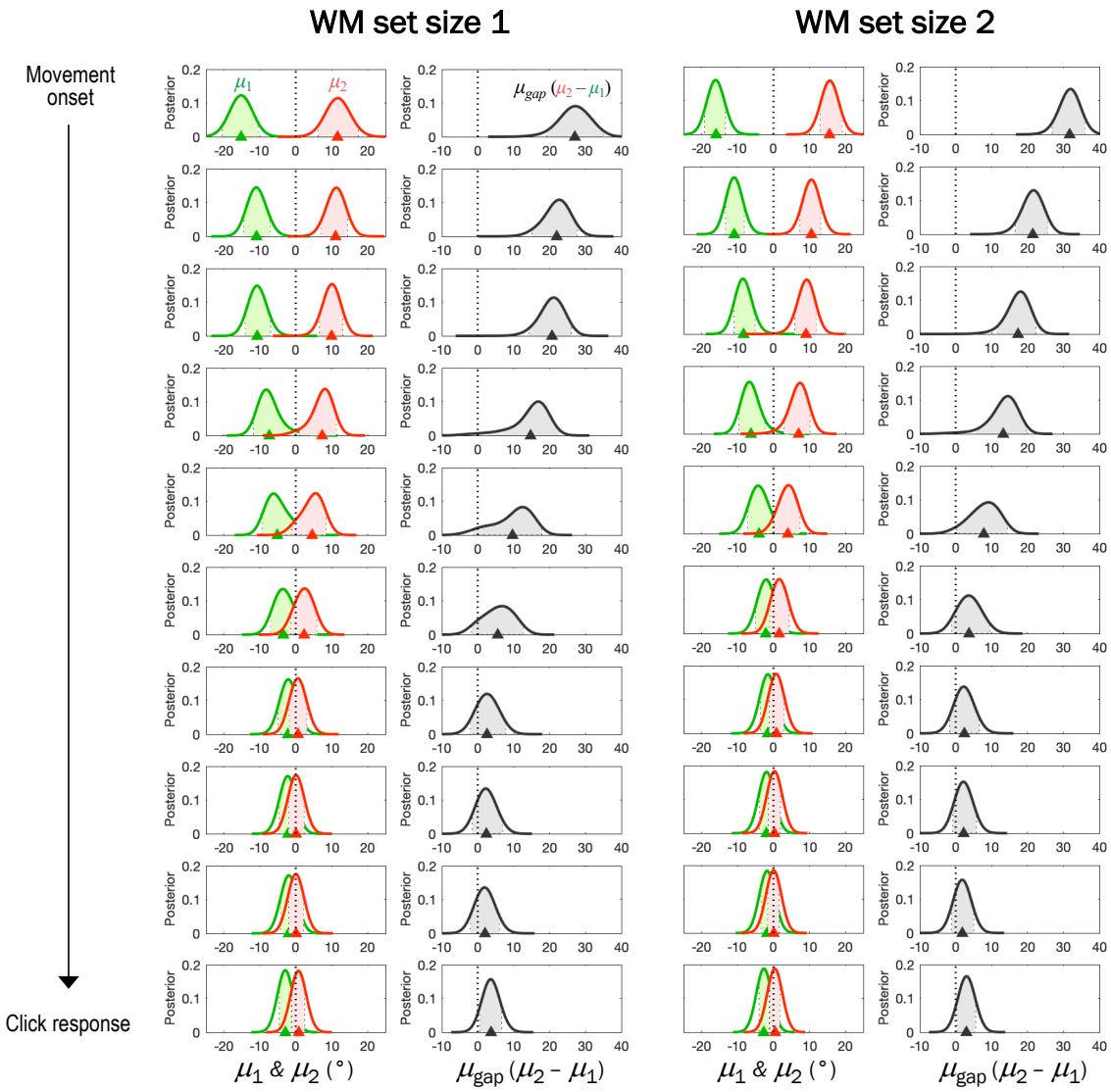
Figure 32.

Analytic approach for testing the predictions for the circular trajectory bias distributions based on the single-item template hypothesis (H_{SIT}) and multiple-item hypothesis (H_{MIT}).



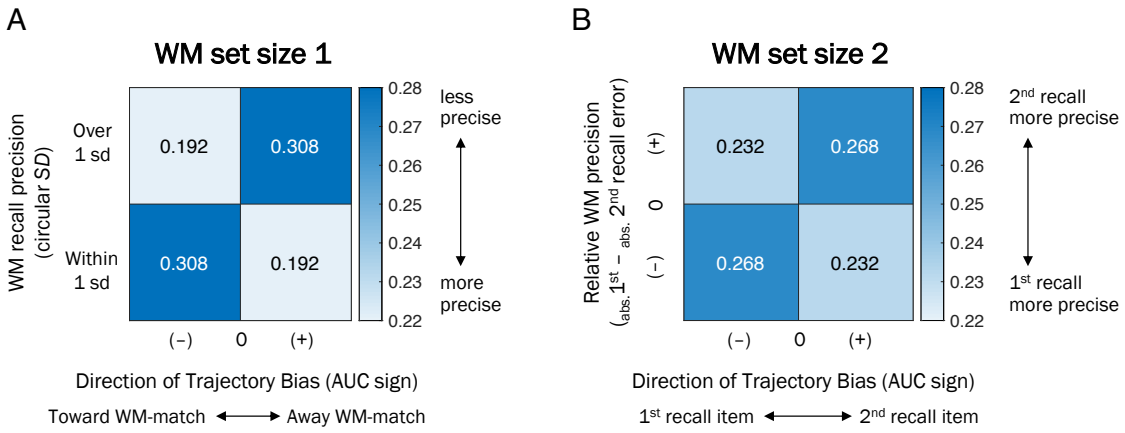
Note. In step 1, we identified that the set size 1 circular trajectory biases were systematically deviated toward the corresponding WM-match color on color wheel, either clockwise (CW trials) or counterclockwise from the perceptual target (CCW trials). In step 2, these set size 1 circular biases at every 10th normalized movement time were ignored their CW or CCW condition codes. In step 3, these condition-ignored circular bias distributions were successfully modeled by the bi-modal function (i.e., recovery of the two dissociable peaks). In step 4 and 5, importantly, the set size 2 circular trajectory bias distributions were fitted to the competing models (i.e., bi-modal SIT vs. uni-modal MIT) to see if they are better described by a model predicted by the H_{SIT} or H_{MIT}.

Figure 33.
The hierarchical Bayesian posteriors of μ shift parameters.



Note. The posterior probability densities of the two μ shift parameters μ_1 and μ_2 in the bi-modal SIT model and their differences μ_{gap} (μ_2 - μ_1) along every 10th normalized movement time from movement onset to final click response. Shaded area indicates 95% highest density interval.

Figure 34.
Trial-categorization results for each memory set size condition.



Note. The results of proportion of trials from the trial-categorization consisting of a 2-by-2 matrix. (A) For working memory (WM) set size 1, the horizontal dimension categorizes whether the perceptual matching mouse trajectory deviated toward or away from the WM-match color on color wheel (by the sign of area under curve; AUC), whereas the vertical dimension categorizes whether WM recall error for a given trial was within or over 1 circular standard deviation of errors for total trials. (B) For WM set size 2, the horizontal dimension categorizes whether the mouse trajectory deviated toward the 1st or 2nd recalled WM-match color on color wheel, whereas the vertical dimension categorizes by the sign of the relative absolute recall errors (1st – 2nd recalls; negative sign indicates the 1st recall being more precise than the 2nd recall). For both set sizes, the first (+, +) and the third quadrants (–, –) are consistent with the idea that WM-directed trajectory bias was associated with how precise the memory item was represented in WM.

Chapter 7. General Discussion

7.1. Establishing a Psychological Meaning of μ Shift

The dissertation research devoted to characterizing the central tendency of the errors from continuous WM reports as a behavioral manifestation of the underlying representational appearance, and further establishing accuracy as a psychologically valid aspect of memory quality that is independent of precision. Several lines of work are presented in support of the argument. Part I sets a theoretical ground for the necessity of representational shift in characterizing noisy WM presentations. I elucidated the erroneous nature of WM and how the recall error distribution can be dissected into qualitatively distinct components reflecting different states of mnemonic representations by probabilistic mixture modeling. Despite this widely-accepted division of the quality and quantity of WM representation conceptualized as precision and capacity, the correspondence between external stimulus and the central tendency of internal representation (i.e., accuracy) has largely been ignored in the previous WM literature.

Several previous studies have looked into the central tendency of the WM errors, but within a restricted domain of research such as statistical systematic memory distortion toward the ensemble average of the stimuli (Brady & Alvarez, 2011) or category means (Bae et al., 2015; Hardman et al., 2017). Otherwise, the central tendency is often neglected based on assumption that the overall WM errors will eventually be centered around the true feature value of the physical stimulus. Contradict to this assumption, however, the data simulation and parameter recovery tests in Chapter 1 showed that random drifts added to individual sampling points contributed to the variability of the

overall simulation set in an additive manner, thus resulted in a robust underestimation of precision (i.e., the width of the sampling distribution). This casts a doubt on the conventional operationalization of mnemonic precision estimated by the width of the between-trial error distribution. From the simulation results, I derived a *root-sum-squared* relationship of variances, where one type of variance arises from the assumed overall variability of the representation (i.e., *precision*), whereas the other type of variance emerges from the trial-specific random drift of the representation (i.e., *accuracy*). The recent advances in continuous measurements enabled us to examine the quality aspect of memory representation which is difficult with binary response methods. Nonetheless, the level of analysis anchored to the between-trial variability has inherent limitations in accurately characterizing the noisy mnemonic representation convoluted by independent sources of variance, inaccuracy and imprecision.

Using the advantages for reliable parameter estimation of the hierarchical Bayesian mixture model introduced in Chapter 2, Part II introduced the many faces of representational shift with two empirical studies. In these studies, the μ shift parameter of the mixture model was utilized as a critical variable for testing theoretically-grounded hypotheses in the domain of WM research. Another set of empirical studies in Part III further combined the μ shift with the mouse-tracking method. These studies demonstrated the potential for mouse trajectory data to investigate how the moment-by-moment evolutions of shift errors are manifested during decision-making and motor actions. The results from the dynamic mouse trajectory patterns provide supporting evidence for a hierarchical structure of WM representations via attentional weights.

Altogether, this dissertation presents a theoretical framework, a quantitative modeling approach, and a collection of empirical studies to validate representational shift as an independent source of noise that contributes to the overall mnemonic quality. The novel methodological approaches for the mouse trajectory analyses in conjunction with the measures of representational shift further shed light on how the trial-wise level of analysis can be conducted and offer a different perspective for characterizing mnemonic quality in WM tasks.

7.2. Future Research Directions

The collective empirical work in this dissertation highlights how the representational shifts can emerge through a variety of factors such as stimulus property, categorical encoding, and interactions between representations or between internal representations and external sensory inputs. For this representational shift to be fully psychologically valid, however, further evidence need to be fulfilled by inspecting the neurally reconstructed representation using neuroimaging techniques and computational modeling. For example, although the precision and shift parameters of the probabilistic mixture model are mathematically independent, their underlying neurocomputational mechanisms are largely unknown. Therefore, testing whether the behaviorally estimated representational shift and precision are directly derived from the corresponding properties of the underlying mental representation in the form of population brain activity would be necessary. On the one hand, the behavioral manifestation of representational shift is expected to reflect the biased central tendency peak of the underlying neural representation. On the other hand, behavioral responses are inherently confounded by

several types of noise due to unstable motor actions or strategic response bias. Therefore, the ultimate goal of future research would be identifying the correspondence in representational shifts between the brain and behavior.

Quantifying representational shift at the neural level could be achieved by adopting an inverted encoding model (IEM) approach to reconstruct trial-by-trial feature-specific information retained in WM from population neural response profile recorded by functional magnetic resonance imaging (fMRI) blood-oxygenation-level-dependent (BOLD) signals. The IEM now become one of the popular standard methods for reconstructing the population channel responses by modeling hypothesized population-level tuning functions (Ester et al., 2015; Sprague & Serences, 2013). Some recent studies using the IEM techniques have shown that the reconstructed channel response profile could successfully predict trial-by-trial sensory features being maintained (Kim et al., 2020; Li et al., 2021). Since the modeled channel tunning functions are typically comprised of a continuous feature dimension, the reconstructed channel responses could also be modeled by the same hierarchical Bayesian mixture model introduced in the present work to reliably recover the representational shift.

Taken altogether, the theoretical framework of representational shift in WM in the present dissertation provides a different perspective of understanding the blurred nature of our mental representation of the external world. There are several motivations to apply this perspective to the future WM research. First, a shift in the central peak location in WM error distribution can serve a useful means to investigate the fundamental structure of WM and its interaction with ongoing visual processing. It is widely accepted that

multiple concurrent WM representations are organized in a hierarchical structure via attentional accessibility (D'Esposito & Postle, 2015). Although debate continues on the specific mechanisms (Stokes et al., 2020), general consensus in the field is that only a single representation in a focus of attention can directly interact with ongoing visual processing in early visual cortices whereas the other representations are maintained in a latent, activity-silent state (Serences et al., 2009; but see Xu, 2017). This sensory recruitment theory predicts a common use of early visual areas to support both ongoing perception and WM maintenance (Rademaker et al., 2019). Therefore, examining the representational shift of WM representations in response to new perceptual inputs could be central to testing the sensory recruitment hypothesis and a dynamic interplay between WM and other cognitive processes that guide our behaviors (Boettcher et al., 2021; Zokaei et al, 2019).

Second, the degree of representational shift may vary across individuals which could potentially contribute to the individual differences in WM function. There is only a handful of studies that examined how different experimental manipulations such as between-item relationship or memory/cognitive load affect the extent of attractive or repulsive representational bias (Allred et al., 2016; Chunharas et al., 2019). Further examining individual differences in representational shift by various factors such as affective, aging, and health-related conditions would help to understand dissociable components in WM functions in addition to the known dissociation between capacity and precision (Gold et al., 2010; Wee et al., 2013; Xie & Zhang, 2018).

Lastly, although subtle shifts in the correspondence between the external stimulus and internal representation may seem like minor errors of memory, a systematic tendency of distorting internal representations of the world can be extended to promote our understanding of general cognition and its interaction with broader factors such as prior knowledge, personality, attitude, political stance, and so forth.

References

Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review, 23*(6), 1825-1831.

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*(4), 744-763.

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

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384-392.

Chunharas, C., Rademaker, R. L., Brady, T., & Serences, J. (2019). Adaptive memory distortion in visual working memory. *PsyArXiv*.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*, 115-142.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron, 87*(4), 893-905.

Gold, J. M., Hahn, B., Zhang, W., Robinson, B. M., Kappenman, E. S., Beck, V. M., & Luck, S. J. (2010). Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. *Archives of General Psychiatry, 67*(6), 570–577.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology. Human Perception and Performance, 43*(1), 30-54.

Kim, I., Hong, S. W., Shevell, S. K., & Shim, W. M. (2020). Neural representations of perceptual color experience in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, 117*(23), 13145-13150.

Li, H.-H., Sprague, T. C., Yoo, A., Ma, W. J., & Curtis, C. E. (2021). Joint representation of working memory and uncertainty in human cortex. *BioRxiv.*

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience, 22*(8), 1336-1344.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science, 20*(2), 207-214.

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience, 16*(12), 1879.

Stokes, M. G., Muhle-Karbe, P. S., & Myers, N. E. (2020). Theoretical distinction between functional states in working memory and their corresponding neural states. *Visual Cognition*, 28(5-8), 420-432.

Wee, N., Asplund, C. L., & Chee, M. W. L. (2013). Sleep deprivation accelerates delay-related loss of visual short-term memories without affecting precision. *Sleep, 36*(6), 849–856.

Xie, W., & Zhang, W. (2018). Mood-dependent retrieval in visual long-term memory: Dissociable effects on retrieval probability and mnemonic precision. *Cognition and Emotion, 32*(4), 674–690.

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences, 21*(10), 794-815.

Zokaei, N., Board, A. G., Manohar, S. G., & Nobre, A. C. (2019). Modulation of the pupillary response by the content of visual working memory. *Proceedings of the National Academy of Sciences, 116*(45), 22802-22810.