Causal Basis of Biases in Visual Estimation

Supervisor: Paul Bays

Acknowledgement and Contributions

The behavioural experiment of this research project was designed and coded by the Bayslab (Computational Cognition Group, Department of Psychology, University of Cambridge). Data collection was collaboratively conducted by the author (on 5 participants), the author's project partner Ms. Neha Abraham (on 6 participants), and previous members of the Bayslab (on 17 participants).

Profound thanks to Dr. Ivan Tomić and Prof. Paul Bays for their kind instructions and supports on data collection, statistical analysis, and computational modelling. Profound thanks to Prof. Paul Bays, Mr. Jintong Hu, and Mr. Bowen Xiao for their reviews and comments on the draft of this report.

Abstract

Visual working memory (VWM) has been conceptualised as a collection of continuous, flexible, but limited neural resources, raising questions regarding the factors that influence its allocation among visual items. In this research project, I investigate the effect of item difficulty, i.e., how easily can an item be perceived and remembered, on the allocation of VWM neural resources among competing items. Through a combination of behavioural experiment and computational modelling, I show that human observers tend to biasedly allocate more VWM neural resources towards the easier item. In addition, I verify that this allocational bias is linked with the competition for attention among different items, and I demonstrate that human's allocational strategy seemingly deviates from the optimum.

Table of Contents

Key Abbreviations

VWM: Visual working memory

RDK: Random dot kinematogram

MAE: Mean absolute error

1. Introduction

Visual working memory (VWM) refers to the cognitive system that allows the temporary storage and retrieval of visual information, and its concept has been a subject of active research and debate (Ma et al., 2014). While originally theorised as discrete item-storing slots (Luck & Vogel, 1997), recent studies have pointed out that VWM resembles a collection of continuous neural resources (Bays & Husain, 2008; Wilken & Ma, 2004), which is limited in capacity (Gorgoraptis et al., 2011; Bays & Husain, 2008) and can be flexibly allocated across different items (Bays et al., 2009).

Under the idea of continuous, flexible, but limited neural resources, when resource-demanding items are abundant, problem regarding the pattern of their allocation arises. Recent studies have shown that VWM neural resources are not uniformly distributed across different items in a visual scene, as several task-related factors can bias their allocation. Specifically, when competing items are concurrently displayed, human observers tend to allocate more neural resources towards the item that is either more frequently probed for recall (Yoo et al., 2018), highlighted by external cues (Gorgoraptis et al., 2011), or associated with greater rewards (Brissenden et al., 2023), resulting in an enhanced representational accuracy for the prioritised item but reduced memory precisions for the non-prioritised ones. The degree of allocational bias can be subjectively adjusted to accommodate the overall behavioural goal (Yoo et al., 2018; Bays, 2014), and attention, the "gatekeeper" of VWM (Kastner & Ungerleider, 2000; Awh et al., 2006), seems to be the key mechanism that guides this process (Cong & Kerzel, 2021). Beyond task-related factors, however, the influences of explicit item properties on neural resource allocation have seldom been sophisticatedly studied.

In this research project, I investigate the effect of item difficulty, i.e., how easily can an item be perceived and remembered, on the allocation of VWM neural resources among different items. Through a combination of behavioural experiment and computational modelling, I attempt to answer three main questions:

- Does item difficulty bias the allocation of neural resources among items?
- Is this bias, if exists, linked with the competition for attention among items?
- Does the observed bias, if any, match the optimal strategy of neural resource allocation that miminises the expected error in recall?

The behavioural experiment of this project is based on a two-item remember-delay-recall task, using random dot kinematograms (RDKs) as memory stimuli. By manipulating item difficulty via altering perceptual noise within each RDK, I show that the allocation of neural resources is systematically biased towards the easier item. By modifying task configuration, I demonstrate that this allocational bias is linked with the competition for attention among items.

The computational modelling of this project is based on a stochastic population coding model, following the work by Bays (2014). Building on neurobiologically realistic principles, including probabilistic spike generation (Pouget et al., 2000) and divisive normalisation (Louie & Glimcher, 2019), this model can accurately reproduce critical characteristics of human behaviour in VWM recall tasks (Bays, 2014; Schneegans et al., 2020), thus serves as the ideal tool for analysing the pattern of neural resource allocation. Using the model, I show that the idea of continuous, flexible, but limited neural resources can reasonably capture

human recall performance observed in the behavioural experiment. Though model fitting, I quantitatively confirm the presence of bias in neural resource allocation triggered by item difficulty and competition for attention. By deriving the allocational strategy that minimises the expected recall error, I further demonstrate that human behaviour seemingly deviates from the theoretical optimum.

The methods and results are presented in Section 2 and 3 for the behavioural experiment and Section 4 and 5 for the computational modelling, respectively. Discussions regarding main results, model improvements, and future directions are presented in Section 6, with the project plan and ethical reflections provided in the Appendix.

2. Experimental Methods

2.1 Participants

A total of 28 naïve participants (9 males, 19 females) were recruited via the Cambridge SONA system for psychological research, age ranging from 18-44, in which 27 participants completed the experiment[1]. All participants reported to have normal colour vision and normal or corrected-to-normal visual acuity. The written consents of participants were collected before attempting the experiment, and all participants were paid 10 pounds per hour for their participation.

2.2 Experimental Setup

Stimuli were displayed on a 27-inch gamma-corrected LCD monitor with a refresh rate of 60Hz. Participants were seated in a dark room, with their head supported by an optical chin rest and eyes positioned 60cm away from the monitor. A warm-coloured lamp was lighted behind participants to prevent visual fatigue. Gaze position on the monitor was tracked online using a 9-point calibrated eye-tracker (EyeLink 1000 Plus, SR Research Ltd.) with a sampling rate of 250Hz. Trial generation, stimulus presentation, and response registration were powered by Psychtoolbox on MATLAB (MathWorks Inc.).

2.3 Stimuli

Background. The monitor display was set as a mid grey background.

Fixation. A dark circular annulus served as the fixation point. The annulus has an outer radius of 0.25° (in visual angle), an inner radius of 0.15°, and was positioned at the center of the

[1] One female participant failed to complete the experiment due to a technical failure in eye tracking.

monitor. At the start of each trial, participants were instructed to fixate on the circular annulus. Once a stable gaze was registered, the inner radius of the annulus expanded to 0.2° and remained constant for the rest of the trial. During the display of visual stimulus, participants were required to maintain a stable gaze within 2° around the center of the annulus. Once a deviation was detected, the ongoing trial terminated immediately, a warning message appeared on the screen, and the trial restarted with newly randomised stimulus.

Random Dot Kinematograms. A pair of coloured RDKs was used as the memory stimulus. Each RDK was presented in a circular panel with a radius of 2.8°, positioned 5° left/right to the center of fixation point, and contained 40 circular dots moving at a speed of 4° per second. The RDKs were painted in a contrasting colour pair (green vs. blue in one task and yellow vs. magenta in the other task, randomly assigned)[2], with each colour correlated with a specific level of coherence[3] throughout the task. Critically, the RDK painted in one colour, *the harder RDK*, had lower coherence in majority of trials and thus was typically harder to remember, whereas the RDK painted in the other colour, *the easier RDK*, had higher coherence in majority of trials and thus was typically easier to remember. Within a colour pair, the association of colour with RDK difficulty was randomised across participants.

[2] The exact colour in RGB index: green = (0, 199, 128), blue = (0, 187, 241), yellow = (237, 154, 0), magenta = (255, 79, 208), based on 8-bit (255) scale.

[3] Coherence is defined as the proportion of coherent dots in the RDK that move in the same direction. A coherence of 45% indicates 45% of dots are moving in the one coherent direction and 55% of dots are moving in random directions.

2.4 Procedure

Each participant completed two tasks: *the simultaneous task* (see Fig.1), where the RDKs were presented together, and *the sequential task* (see Fig.2), where the RDKs were presented one after another. The structure of each trial is as follows:

1. *Fixation.* A trial started with the presentation of the circular annulus, which participants were required to stably fixate on for 750ms before the trial proceeded.

2. *Remember.* The harder and easier RDK were presented either simultaneously for 750ms or sequentially for 1500ms (750ms each), depending on the task, with randomised spatial (left/right) and temporal order of display. The motion direction of coherent dots in each RDK was randomly drawn from a uniform circular distribution. Participants were instructed to memorise these coherent motion directions while continuously fixating on the annulus. The critical manipulation of the experiment involved varying the coherence of RDKs. In 7 out of 10 trials, *the unequal trials*, the harder RDK has a coherence of 45% and the easier RDK has a coherence of 85%. In 3 out of 10 trials, *the equal trials*, both the harder and easier RDK have an equal coherence of 65%. The order of equal and unequal trials was randomised for each task.

3. *Delay*. The RDK pair was removed, and participants were required to continuously fixate on the annulus for 1000ms.

4. *Probe.* An empty circular panel reappeared at the center of the screen, with its colour indicating the identity of RDK to be recalled. The harder and easier RDK were probed with equal frequency in a randomised order, and only one RDK was probed per trial.

5. *Recall.* Immediately after the probe was displayed, participants were allowed to report the remembered motion direction of coherent dots in the probed RDK. By manipulating a computer mouse with a preferred hand, participants oriented a white arrow to match its direction with the remembered coherent motion direction in the probed RDK, then pressed the left-click button to submit their response. No time limit was imposed on participants during this stage, and successful response submission immediately initiated the next trial.

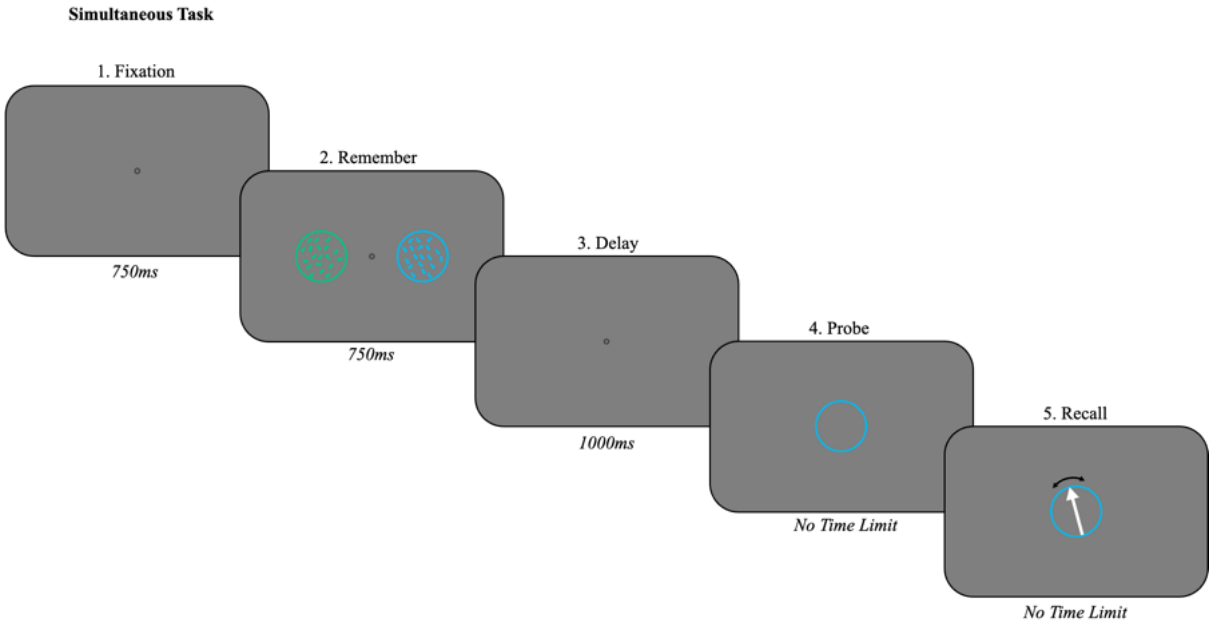


Fig.1: Experimental procedure of the simultaneous task.

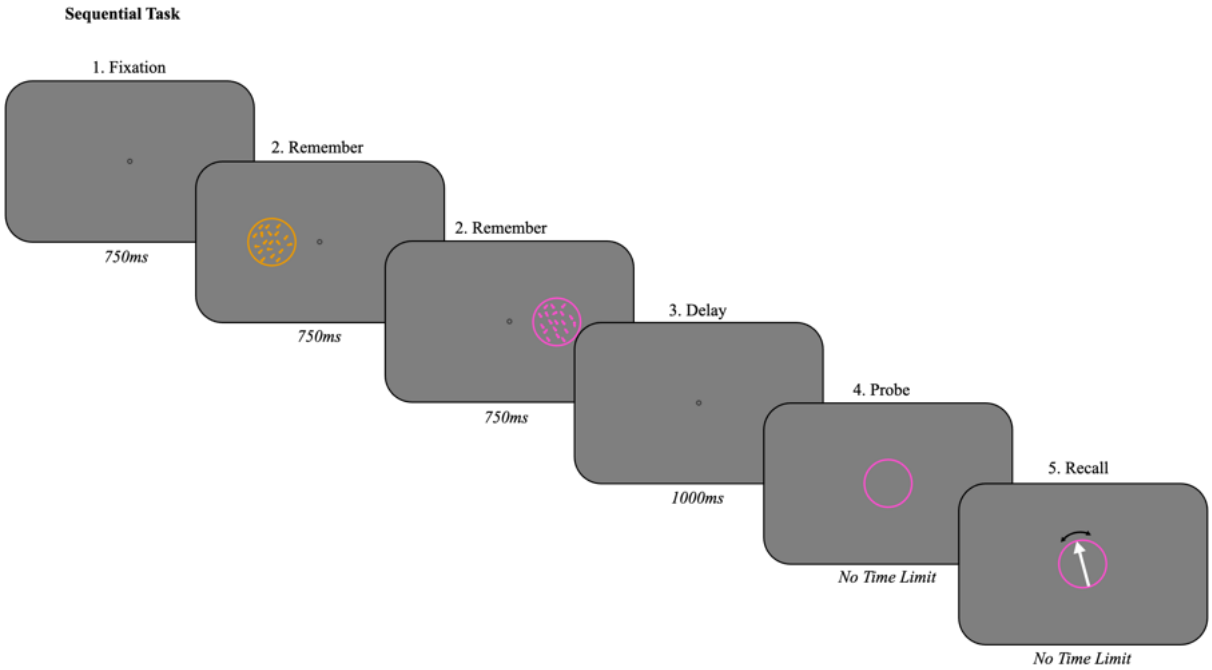


Fig.2: Experimental procedure of the sequential task.

Each participant completed a total of 400 trials in 8 blocks, with 4 consecutive blocks being the simultaneous task and the other 4 being the sequential task (50 trials per block, 200 consecutive trials per task). The chronological order of the task taken was randomised across participants. Each block typically required 8-10 minutes to complete, and a short break was provided between blocks. At the start of the experiment, participants conducted 5-15 practice trials to familarised themselves with the experimental setup and task procedures.

2.5 Data Analysis

Recall Performance. The recall error made in each trial was defined as the angular deviation between the recalled and actual coherent motion direction in the probed RDK, which ranges from −180° to 180° or $-\pi$ to π in radians. The recall performance under each condition (unequal vs. equal trials, harder vs. easier RDK) was assessed using the mean absolute error (MAE) in recall, which was computed by averaging the absolute value of recall errors across all the relevant trials.

Data Cleansing. Post-experimental feedbacks and preliminary data inspections indicated that the experiment was challenging, as some participants appeared to complete the task by random guessing. To address this, the collected data were first cleansed by visually inspecting the distribution of recall errors block-by-block for each participant. If the recall errors in a particular block were uniformly distributed across the entire circular space (see Fig.3), indicating random guessing behaviour, data collected from that block would be removed without replacement. Any participant exhibited random guessing in more than 2 blocks per task would be excluded from further analysis. To prevent potential biases, data cleansing was performed prior to any statistical analysis and was independently verified twice to avoid human error.

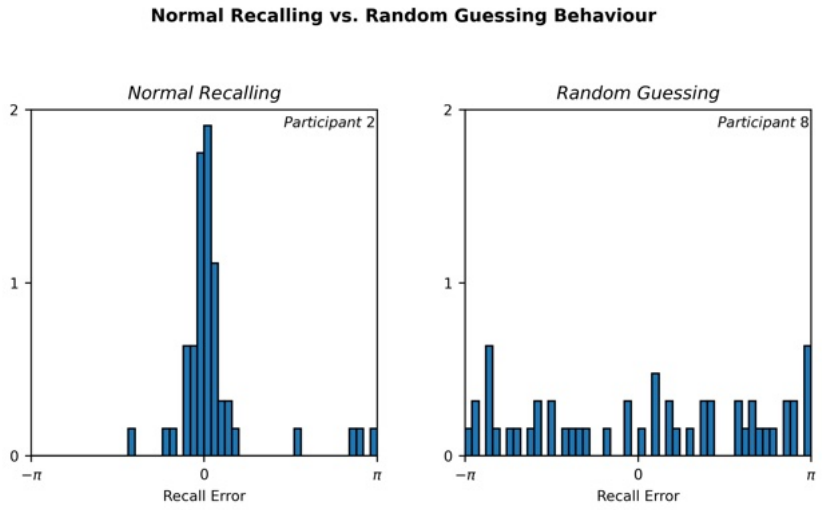


Fig.3: The block-wise recall error distribution produced by normal recalling (left) vs. random guessing behaviour (right), plotted in density-weighted histograms with 50 bins across the entire circular space. Data taken from two representative participants during the same block of the sequential task.

Following the protocol of data cleansing, 15 participants' data were fully retained, 4 participants' data were modified by removing either one or two blocks per task, and 8

participants' data were completely excluded, leaving 19 participants (6 males, 13 females) valid for further analysis.

Statistical analysis. Hypotheses regarding differences in population means and correlations between variables were tested using Bayesian *t*-tests[4] and Bayesian Pearson's correlation analyses, with statistical results reported in Bayes factors (BF_{10}). A larger BF_{10} indicates greater statistical evidence for the alternative hypothesis (H_1), a smaller BF_{10} indicates greater statistical evidence for the null hypothesis (H_0), and $BF_{10} = 1$ indicates equal statistical evidence for both H_1 and H_0. All statistical analyses were conducted using JASP (JASP Team, https://jasp-stats.org/), with the detailed statistical settings provided in the Appendix.

[4] Or its non-parametric equivalent, when presumptions regarding the normality (verified by the Shapiro-Wilk test) or the homoscedasity (verified by the Levene's test) of data for Bayesian *t*-tests were violated.

3. Experimental Results

3.1 Bias in Neural Resource Allocation

Does item difficulty affect the allocation of neural resources among competing items?

To examine the generic effect of item difficulty, i.e., RDK coherence, on recall performance, I compared the MAE for recalling the harder vs. easier RDK during simultaneous unequal trials. As shown by Fig.4, MAE for the harder RDK (MAE = 1.10 ± 0.31) was significantly larger than MAE for the easier RDK (MAE = 0.77 ± 0.32; Bayesian paired samples two-tailed *t*-test: $BF_{10} = 2.7\times10^3$, extremely strong evidence for H_1), indicating that participants systematically made larger errors when recalling the harder RDK compared to the easier RDK. As predicted, the lower the RDK coherence, the harder it is to perceive and remember the motion direction of coherent dots in a RDK, thus the poorer the recall performance.

To analyse the effect of item difficulty on neural resource allocation, I compared the MAE for recalling the harder vs. easier RDK during simultaneous equal trials. Post-experimental feedbacks revealed that most participants were able to correctly identify the colour associated with the harder/easier RDK in the simultaneous task, but none reported noticing the existence of equal trials. Thus, the difference in recall performance between the harder and easier RDK during these trials could act as a reasonable indicator of bias in neural resource allocation triggered by item difficulty. As shown by Fig.4, even under controlled RDK coherences, MAE for the harder RDK (MAE = 0.98 ± 0.36) was still significantly larger than MAE for the easier RDK (MAE = 0.79 ± 0.30; Bayesian paired samples two-tailed *t*-test: $BF_{10} = 5.19$, moderate evidence for H_1), suggesting that participants biasedly allocate their neural resources towards the apparently easier RDK in the simultaneous task.

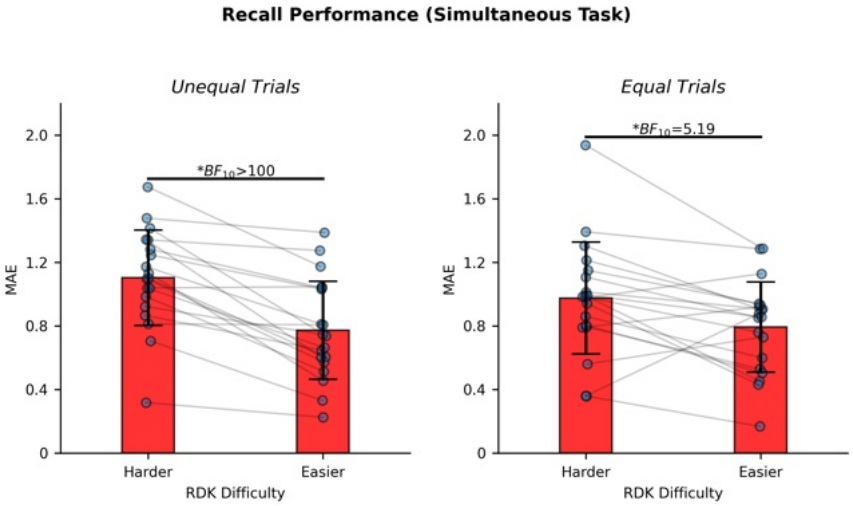


Fig.4: Recall performance of participants during simultaneous unequal (left) and equal trials (right). Red bar heights indicate population mean of MAE, with error bars indicate ± 1*SE*. Blue data points represent individual participants, with gray lines connecting MAE across different conditions derived from the same participant's data. The reported Bayes factors were results from Bayesian paired samples two-tailed *t*-tests. As shown, MAE for the harder RDK was significantly higher than MAE for the easier RDK during both unequal and equal trials, even under controlled RDK coherences, suggesting that neural resource allocation was biased towards the easier RDK.

3.2 Competition Dependent Bias

Is the observed bias in neural resource allocation linked with the competition for attention among items?

To investigate the correlation between attentional competition and bias in neural resource allocation, I compared the MAE for recalling the harder vs. easier RDK in the sequential task, which, due to its sequential nature of item presentation, should eliminate the underlying competition for attention between RDKs. As shown in Fig.5, during sequential unequal trials, participants systematically made larger errors when recalling the harder RDK (MAE = 0.94 ± 0.25) compared to when recalling the easier RDK (MAE = 0.58 ± 0.24; Bayesian paired samples two-tailed *t*-tests: $BF_{10} = 1.9\times10^6$, extremely strong evidence for H_1), consistent with the result from simultaneous unequal trials. However, during sequential equal trials, the

difference in MAE between the harder (MAE = 0.75 ± 0.33) and easier RDK (MAE = 0.66 ± 0.27) was no longer statistically supported (Bayesian paired samples two-tailed *t*-tests: BF_{10} = 0.456, weak evidence for H_0), suggesting the absence of bias in neural resource allocation between RDKs. These results confirm that the observed bias in neural resource allocation during the simultaneous task was indeed caused by attentional competition between RDKs, and removing such competition eliminates the bias.

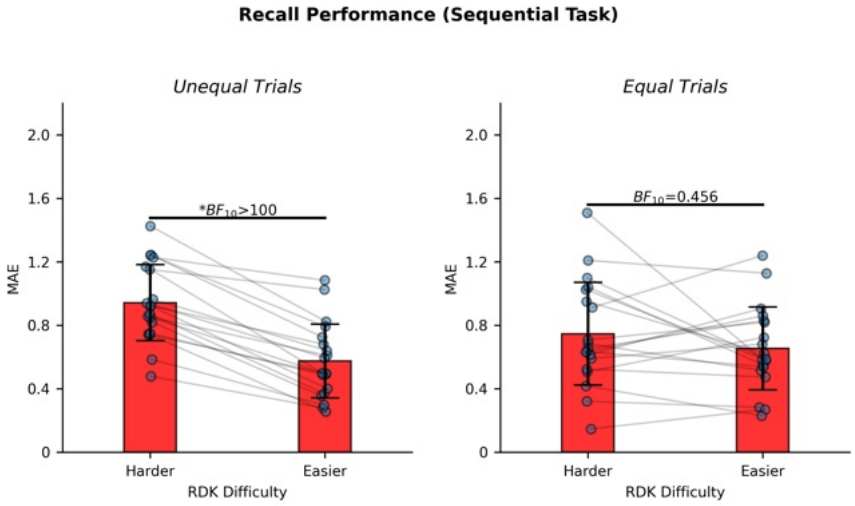


Fig.5: Recall performance of participants during sequential unequal (left) and equal trials (right). Red bar heights indicate population means of MAE, with error bars indicate ± 1*SE*. Blue data points represent individual participants, with gray lines connecting MAE across different conditions derived from the same participant's data. The reported Bayes factors were results from Bayesian paired samples two-tailed *t*-tests. Contrary to the simultaneous task, MAE for the harder RDK was no longer statistically different from MAE for the easier RDK in the sequential equal trials, indicating that the previously observed bias in neural resource allocation is competition dependent.

4. Modelling Methods

4.1 Model Description

Inspired by the concept of continuous, flexible, but limited neural resources, based on the work by Bays (2014), I constructed a population coding model to further analyse the observed bias in VWM neural resource allocation. The model is stochastic, which utilises spikes generated by an idealised neural population to encode and decode motion direction of coherent dots in a RDK, thus replicating the remember and recall behaviour of human participants in the experiment. The general working principle of the model is as follows, with a diagrammatic illustration presented in Fig.6.

Directional Tuning. The model consisted of 100 neurons, where each neuron was tuned to a specific motion direction. The tuning function of the i^{th} neuron in the model, f_i, followed a von Mises function:

$$f_i = \exp(\omega^{-1}\cos(\theta - \varphi_i) - 1) \quad (1)$$

where θ indicates the motion direction of the input stimulus, φ_i indicates the preferred motion direction of that specific neuron, and ω indicates the width of tuning function. The value of φ_i was adjusted to evenly cover the entire circular space. This particular shape of tuning curve was selected to match the measured tuning curve of motion direction selective neurons in primates (Zaharea et al., 2019).

Neuronal activation. Upon receiving moving dots in a RDK as inputs, the firing rate of the i^{th} neuron in the population, R_i, was determined by:

$$R_i = Coh \cdot \frac{\alpha\gamma}{M\bar{f}} f_i \quad (2)$$

where:

- Coh is the coherence of the targeted RDK
- α is a proportionality constant that reflects the degree of bias in neural resource allocation towards the targeted RDK
- γ is the maximum gain of neural activation, which indicates the "responsiveness" of the whole neural population
- M is the total number of neurons in the population
- $\bar{f}$ is the average firing rate of neurons determined by the directional tuning function
- f_i is the neural activity based on the directional tuning function in equation (1)

Specifics regarding critical terms in equation (2):

- Coh linearly influences neuronal firing rate. Past electrophysiological and neuroimaging studies have showed that firing responses of motion selective neurons in human brain are proportional with the coherence of a RDK (Heeger et al., 2000). Therefore, in this model, coherence was approximated as a linear multiplicity gain-controlling factor of the neural population.
- M and $\bar{f}$ scale down the population neural activity. This accounts for the divisive normalisation of neural activities, which is a fundamental feature of the brain when processing sensory information (Louie & Glimcher, 2019).

$\bar{f}$ can be expressed as:

$$\bar{f} = I_0(\omega^{-1})e^{\omega^{-1}} \tag{3}$$

where I_0 is the modified Bessel function of order zero. Combining equation (1), (2), and (3), the explicit expression of R_i is therefore:

$$R_i = Coh \cdot \alpha \cdot \gamma \cdot \frac{\exp(\omega^{-1}(\cos(\theta - \varphi_i) - 1))}{100 \cdot I_0(\omega^{-1})e^{\omega^{-1}}} \tag{†}$$

Neural Spiking. In the model, each neuron was constructed as a homogeneous Poisson spike generator. Given the value of R_i, the number of spikes, n_i, generated by the i^{th} neuron in the population was drawn from a Poisson distribution:

$$n_i \sim Poisson(R_i) \tag{4}$$

Maximum Likelihood Decoding. After input-induced spike generation, the model's remembered motion direction, $\hat{\theta}$, was decoded by estimating the parameter θ that maximises the likelihood function:

$$\hat{\theta} = \arg\max_{\theta} \Pr(\boldsymbol{n}|\theta) \tag{5}$$

where $\boldsymbol{n}$ indicates the spike count generated by the neural population during the decoding window. As suggested by Schneegans et al. (2020), for Gaussian-like neural tuning functions and a dense neural population, estimating $\hat{\theta}$ in equation (5) is the equivalent of computing the spike summation:

$$\hat{\theta} = \frac{1}{N}\sum_{j}^{N}\varphi_{(j)} \tag{6}$$

where N is the total number of spikes generated by the neural population, and $\varphi_{(j)}$ indicates the preferred tuning direction of neuron that generates the j^{th} spike. Due to the circular nature of motion direction, the summation in equation (6) was achieved by comptuting the circular mean:

$$\hat{\theta} = \text{atan2}\left(\sum_{j}^{N}\sin\left(\varphi_{(j)}\right), \sum_{j}^{N}\cos\left(\varphi_{(j)}\right)\right) \tag{7}$$

where atan2 is the 2-argument arctangent function.

Recall Error Calculation. Like human observers, the decoded motion direction $\hat{\theta}$ from the population coding model deviated from its original input direction θ. After maximum likelihood decoding, this recall error made by the model, $\Delta\theta$, was estimated by:

$$\Delta\theta = [\hat{\theta} - \theta + \pi] \pmod{2\pi} - \pi \tag{8}$$

which maps the recall error onto the circular space $[-\pi, \pi)$ for further analysis.

Error Distribution Generation. To estimate the recall error distribution generated by the model, I randomly generated 10^6 input stimuli θ from a uniform circular distribution, and I allowed the model to encode, decode, and recall the stimuli following the above procedures,

producing 10^6 estimated $\hat{\theta}$ and recall error $\Delta\theta$. The probability density distribution of recall error made by the model was then estimated from these $\Delta\theta$ via a density-weighted histogram, using 100 bins that evenly covered the circular space.

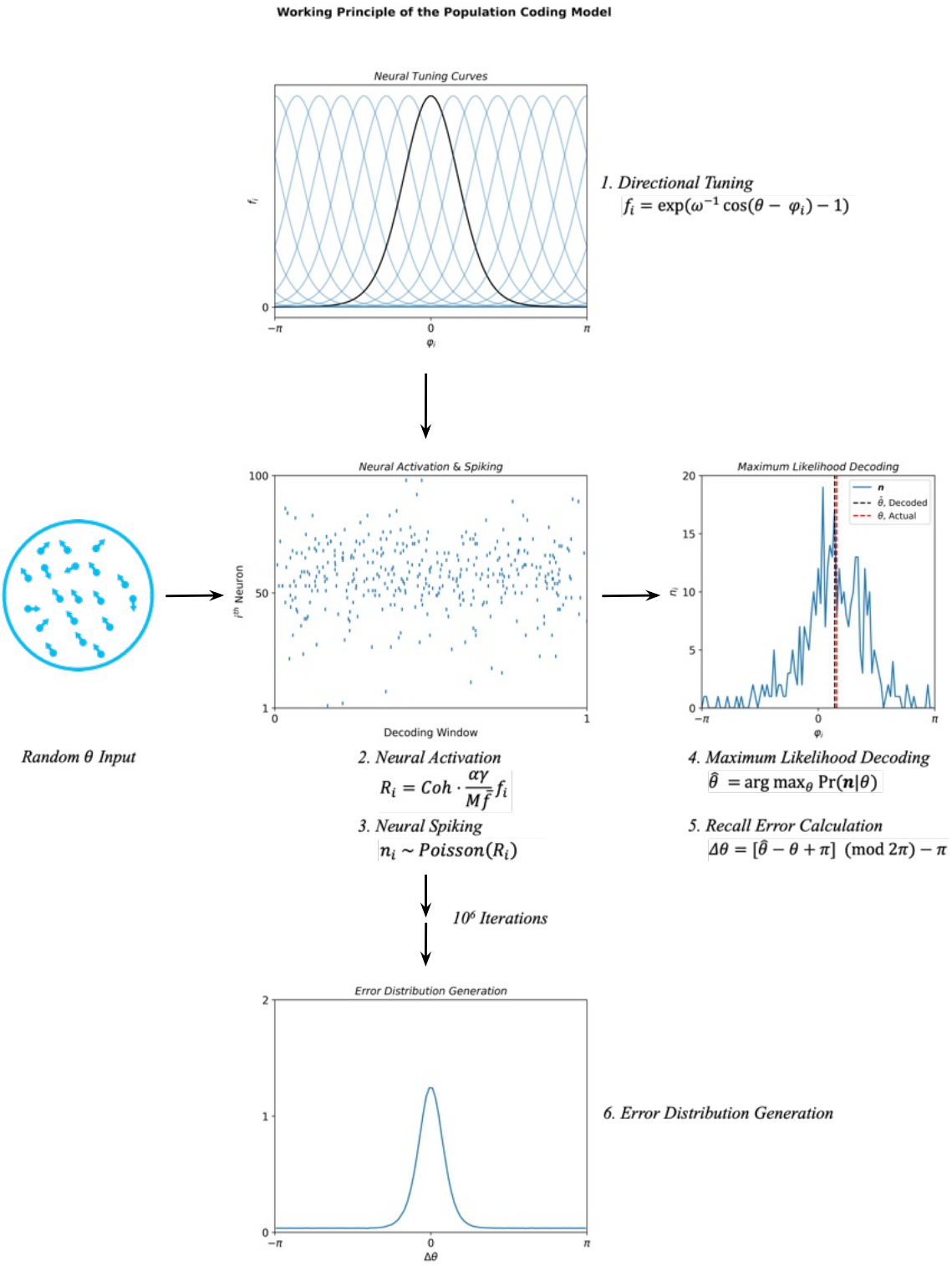


Fig.6: A illustration of the population coding model's working principle. Top, the directional tuning curves of direction-tuned neurons in the model, which densely and evenly covered the entire circular space $[-\pi, \pi)$. Middle, the process of encoding, decoding, and recalling the motion direction of coherent dots from a RDK. First, a RDK with a random coherent motion direction θ was presented to the model, which elicited a specific spiking pattern among the neural population. This spiking pattern was then utilised to compute the decoded motion direction, $\hat{\theta}$, and a corresponding recall error $\Delta\theta$. This cycle was repeated for 10^6 iterations, producing a probability density distribution of $\Delta\theta$, as shown at the bottom.

4.2 Model Fitting

Neural tuning width ω, maximum population gain γ, and proportionality constant α are critical parameters of the model, as they influence the neural spiking pattern and the resultant recall error distribution generated by the model (see Fig.7 and Fig.8).

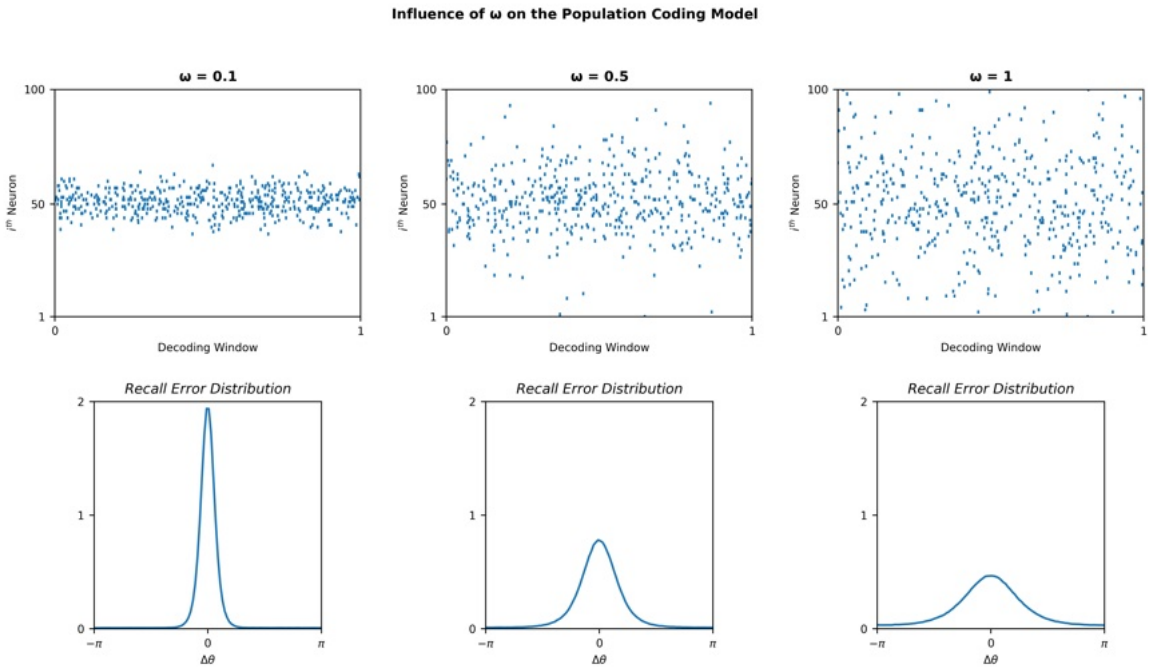


Fig.7: The influence of ω on the neural spiking pattern (top row) and recall error distribution (bottom row) generated by the model, from $\omega = 0.1$ (left) to $\omega = 1.0$ (right), with other parameters kept as constants. As shown by the spike raster plots, ω does not influence the number of spikes generated by the model, but it affects the "concentration" of spikes in the neural population. Thus, increasing ω leads to an increment in recall errors and, generally, a flatter error distribution.

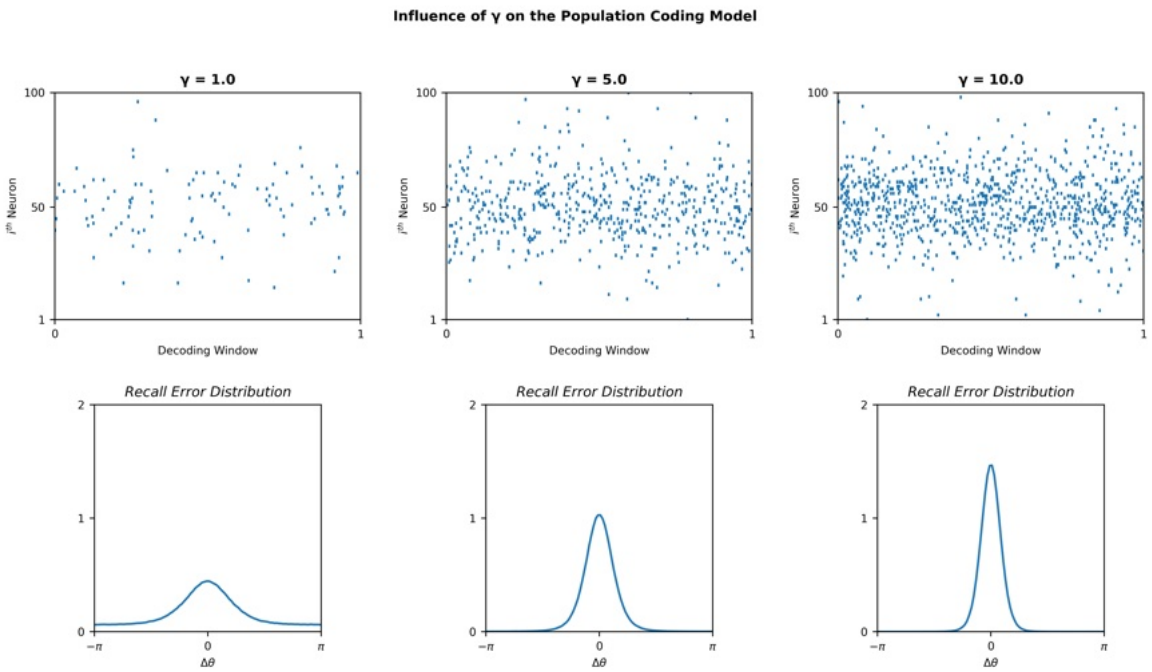


Fig.8: The influence of γ on the neural spiking pattern (top row) and recall error distribution (bottom row) generated by the model, from $\gamma = 1.0$ (left) to $\gamma = 10$ (right), with other parameters kept as constants. In contrast to ω, γ does not alter the "concentration" of spikes in the neural population, but it affects the number of spikes generated by the model. Thus, increasing the numerical value of γ leads to a decrement in average recall errors and a narrower error distribution. Varying other gain-controlling parameters, including Coh and α, essentially yields the same effect as varying γ.

To obtain the best-fit values of these parameters, I fit the model to the empirically measured recall errors, both individually for each participant's data and collectively for a group-wise pooled data. Specifically, I set:

- Coh as a trial-specific fixed parameter, that is:
  - $Coh = 0.45$ for the harder RDK during unequal trials
  - $Coh = 0.85$ for the easier RDK during unequal trials
  - $Coh = 0.65$ for both the harder and easier RDK during equal trials
- Neural tuning width ω as a global free parameter
- Maximum population gain γ as a global free parameter

- Allocational proportionality constant as α for recalling the easier RDK and $1-\alpha$ for recalling the harder RDK[5]

Then, I estimated the best-fit values of ω, γ, and α, denoted as $\hat{\omega}$, $\hat{\gamma}$, and $\hat{\alpha}$, via maximising the sum of log-likelihoods combinedly across all targeted conditions:

$$\hat{\omega}, \hat{\gamma}, \hat{\alpha} = \arg\max_{\omega,\gamma,\alpha} \mathcal{L}_{Coh,\omega,\gamma,\alpha}(\boldsymbol{\Delta\theta^{easy}}) + \mathcal{L}_{Coh,\omega,\gamma,1-\alpha}(\boldsymbol{\Delta\theta^{hard}}) \quad (9)$$

where $\boldsymbol{\Delta\theta^{easy}}$ and $\boldsymbol{\Delta\theta^{hard}}$ are the sets of empirical observed recall errors for the easier and harder RDK, respectively, and the log-likelihood function $\mathcal{L}_{Coh,\omega,\gamma,\alpha}(\boldsymbol{\Delta\theta})$ is defined as:

$$\mathcal{L}_{Coh,\omega,\gamma,\alpha}(\boldsymbol{\Delta\theta}) = \sum_{i} \log\big(\Pr(\Delta\theta_i | Coh, \omega, \gamma, \alpha)\big) \quad (10)$$

where $\Pr(\Delta\theta_i | Coh, \omega, \gamma, \alpha)$ indicates the model's probability of generating the i^{th} empirically observed recall error from the dataset $\boldsymbol{\Delta\theta}$, given the model parameters Coh, ω, γ, and α. For practical runtime consideration, this probability was approximated using the neural resource model toolbox provided by Bayslab (Bays, 2014; Schneegans & Bays, 2017; Schneegans et al., 2020; https://bayslab.org), which has a similar mathematical principle with that described previously.

4.3 Model Evaluation

[5] Under this definition, the value of α indicates the degree of bias in neural resource allocation towards the easier RDK, where $\alpha = 0.5$ indicates an unbiased allocation, $\alpha > 0.5$ indicates a biased allocation towards the easier RDK, and $\alpha < 0.5$ indicates a biased allocation towards the harder RDK.

To straightforwardly check the alignment of population coding model with human in recall performance under each condition, I estimated the MAE generated by the best-fit model, denoted as MAE_{model}, using the equation:

$$MAE_{model} = \frac{\pi}{50}\sum_{j}\left|\Delta\theta_j\right| \cdot \Pr\left(\Delta\theta_j \middle| Coh, \omega, \gamma, \alpha\right) \quad (11)$$

where $\Delta\theta_j$ is taken from 100 sample points evenly spaced across the circular space $[-\pi, \pi)$, and $\Pr\left(\Delta\theta_j \middle| Coh, \omega, \gamma, \alpha\right)$ was, again, similar to that in equation (10), approximated using Bayslab's toolbox. The deviation between the model and human in average recall performance was quantified as:

$$\text{Deviation} = \frac{MAE_{model} - MAE_{empirical}}{MAE_{model}} \quad (12)$$

where $MAE_{empirical}$ is the empirically observed MAE corresponded to the best-fit model, identical with that used in previous behavioural analysis. Under this definition, a perfectly matched model would generate a deviation of 0. A positive deviation indicates that the model is performing systematically worse than its human counterpart, and a negative deviation indicates that the model is performing systematically better than its human counterpart.

4.4 Optimal Allocational Strategy

To find the optimal strategy of neural resource allocation during the simultaneous task, after obtaining the best-fit parameters $\hat{\omega}$ and $\hat{\gamma}$ from model fitting, I estimated the mathematically optimal allocational bias, α_{opt}, that minimises the expected recall error across simultaneous

unequal trials. This was achieved by minimising the sum of model-generated MAE for the harder and easier RDK, denoted as MAE_{model}^{hard} and MAE_{model}^{easy}, via:

$$\alpha_{opt} = \arg\min_{\alpha} MAE_{model}^{hard} + MAE_{model}^{easy} \quad (13)$$

where the values of MAE_{model}^{hard} and MAE_{model}^{easy} were estimated using equation (11).

Modelling works were conducted via Python3 on JupyterLab (for simulation) and MATLAB (for model fitting). Parameter optimisation was accomplished using the Nelder-Mead simplex method (Nelder & Mead, 1965).

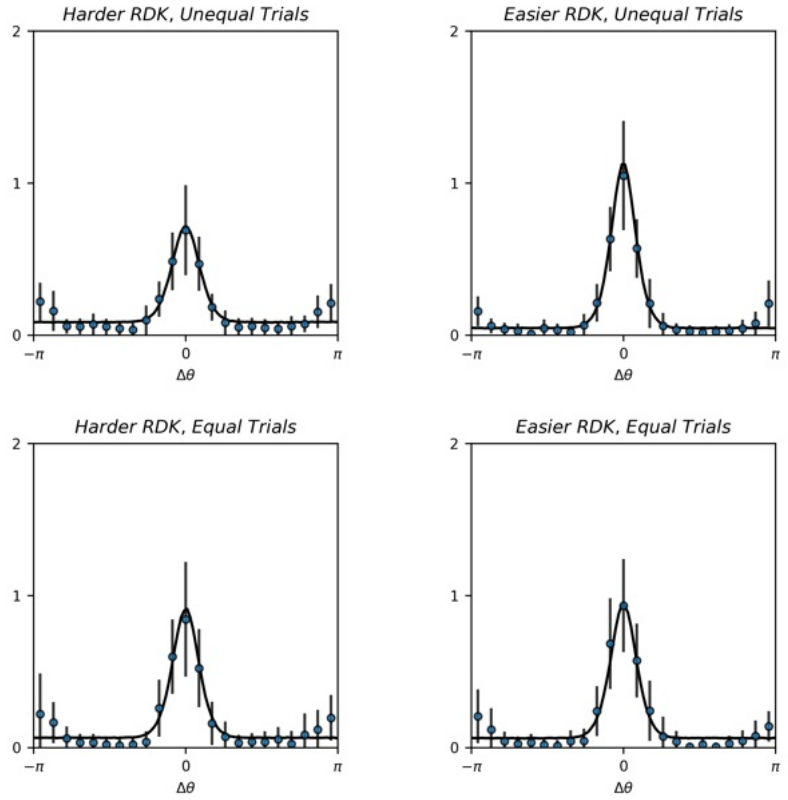
5. Modelling Results

5.1 Model Performance

Does the population coding model accurately reflect human recall performance observed in the simultaneous and sequential task?

To evaluate the model's ability on replicating human recall performances, I fit the model to the empirically measured recall errors from the simultaneous and sequential task, respectively. Following the previously outlined procedures, I globally optimised three model parameters, namely, ω, γ, and α, to fit the model to both individuals' and the pooled data, then I assessed best-fit models' performances. Overall, the population coding model qualitatively captured human recall performances in both simultaneous and sequential task, as the recall error distributions generated by the best-fit models roughly overlapped with the empirical data across all targeted conditions (see Fig.9). In addition, the model-generated MAE closely aligned with its corresponding empirically observed MAE in each condition, implying a satisfactory matching between model and human in average recall performance (see Fig.10), although analyses on deviations suggested that the model significantly outperformed its human counterpart in five of eight conditions (see Fig.11). The numerical values of the best-fit parameters, best-fit model's MAE, and model-empirical deviations were summarised in Table.1, Table.2, and Table.3.

Recall Error Distribution Comparisons (Simultaneous Task)



Recall Error Distribution Comparisons (Sequential Task)

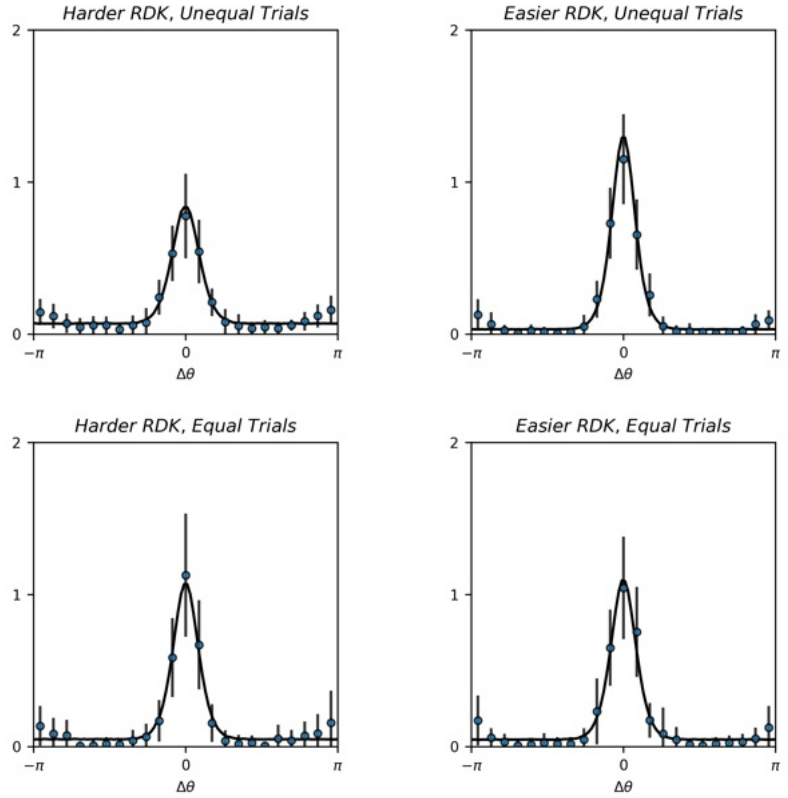


Fig.9: Comparisons between the model-generated and empirically observed recall error distribution. Two subplots respectively demonstrate the results from fitting the simultaneous (top) and sequential task (bottom). Black curves indicate the recall error distributions generated by the best-fit models based on the pooled data, blue circles indicate the mean recall error distributions observed empirically, and error bars indicate ± 1*SE*. Overall, the population coding model qualitatively captured human recall performances across all targeted conditions, except the "180° errors" near $\Delta\theta = \pi$ and $-\pi$ produced by human participants. As demonstrated in later analyses, consequently, these discrepancies made model MAE systematically lower than its corresponding empirical MAE, which lead to negative deviations.

Table.1: Best-fit Parameters

Simultaneous Task

Model Parameter	**Individual**		**Pooled**	
	M	*SE*	*M*	*SE*
ω	0.12	0.05	0.10	/
γ	3.2	2.1	2.8	/
α	0.52	0.07	0.51	/

Sequential Task

Model Parameter	**Individual**		**Pooled**	
	M	*SE*	*M*	*SE*
ω	0.13	0.06	0.11	/
γ	4.2	1.7	3.7	/
α	0.52	0.05	0.51	/

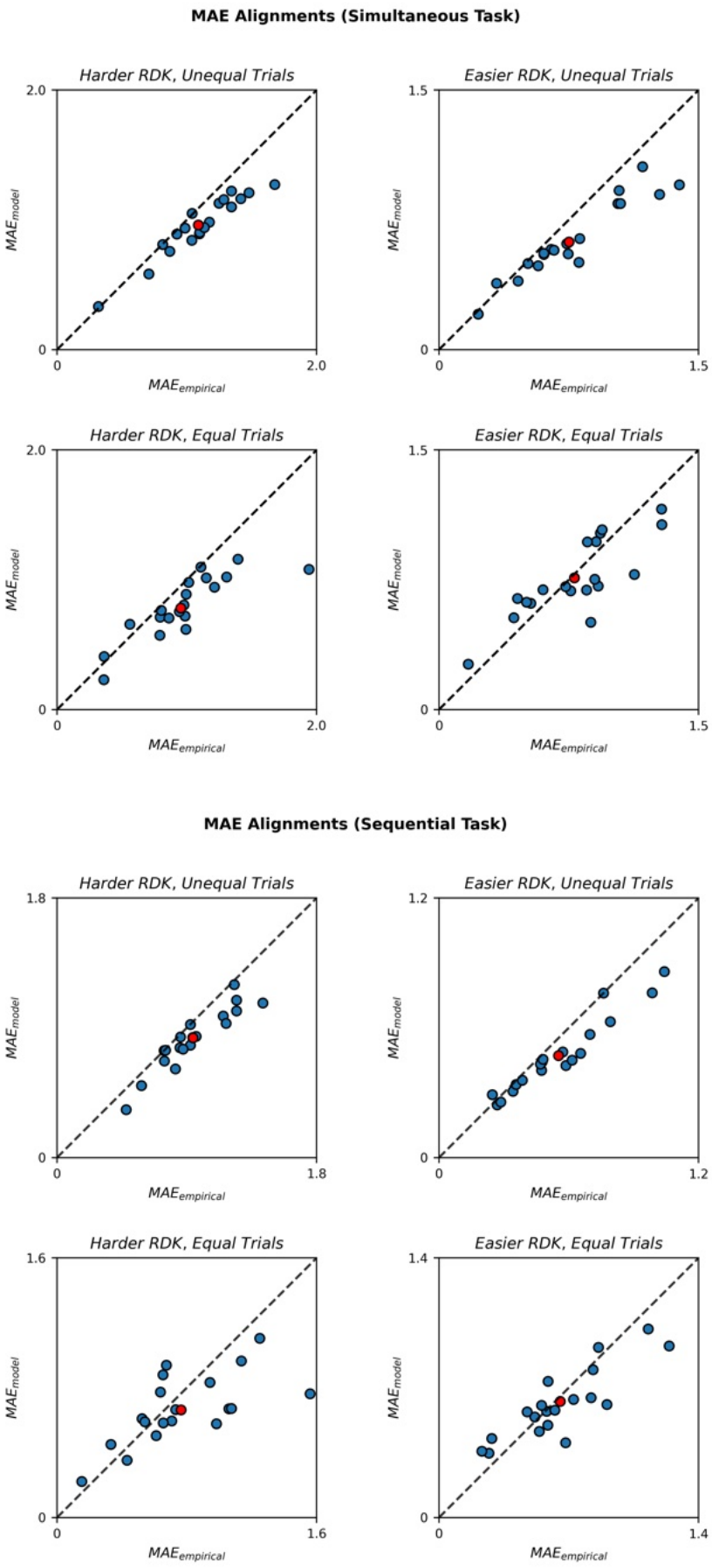

MAE Alignments (Simultaneous Task)
Harder RDK, Unequal Trials
Easier RDK, Unequal Trials
Harder RDK, Equal Trials
Easier RDK, Equal Trials
MAE Alignments (Sequential Task)
Harder RDK, Unequal Trials
Easier RDK, Unequal Trials
Harder RDK, Equal Trials
Easier RDK, Equal Trials
MAE_model
MAE_empirical

Fig.10: Alignments between model and human recall performance. Two subplots respectively demonstrate the results from fitting the simultaneous (top) and sequential task (bottom). Blue data points represent individual participants and red data points represent the pooled data. The horizontal and vertical position of a data point reflects the value of empirically measured and model-generated MAE, respectively. If a best-fit model perfectly aligns with its human counterpart in recall performance, the corresponding data point would be located on the dotted diagonal line. As shown by the figure, across all targeted conditions, most data points clustered diagonally, indicating a satisfactory alignment between model and human recall performance. However, it is noteworthy that a significant proportion of data points, including all the red data points, fell below the diagonal line in every targeted condition, suggesting that the model-generated MAE was slightly, yet systematically, lower than the empirically observed MAE.

Table.2: Individual Model-Empirical MAE

Simultaneous Task

Condition	$\boldsymbol{MAE_{model}}$		$\boldsymbol{MAE_{empirical}}$		**Deviation**		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}*
Harder RDK, Unequal Trials	0.96	0.23	1.10	0.31	-0.14	0.10	3.1×10^{3}
Easier RDK, Unequal Trials	0.64	0.22	0.77	0.32	-0.20	0.17	4.3×10^{2}
Harder RDK, Equal Trials	0.80	0.24	0.98	0.36	-0.23	0.24	57.2
Easier RDK, Equal Trials	0.76	0.23	0.79	0.29	-0.04	0.27	0.275

Sequential Task

Condition	$\boldsymbol{MAE_{model}}$		$\boldsymbol{MAE_{empirical}}$		**Deviation**		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}*
Harder RDK, Unequal Trials	0.82	0.21	0.94	0.25	-0.16	0.12	8.3×10^{2}
Easier RDK, Unequal Trials	0.49	0.18	0.58	0.24	-0.16	0.14	2.6×10^{2}
Harder RDK, Equal Trials	0.65	0.21	0.75	0.33	-0.14	0.37	0.739
Easier RDK, Equal Trials	0.61	0.19	0.66	0.27	-0.05	0.27	0.328

*Results from Bayesian one sample two-tailed *t*-test or its non-parametric equivalent, with the null hypothesis specifying that the population mean ≠ 0.

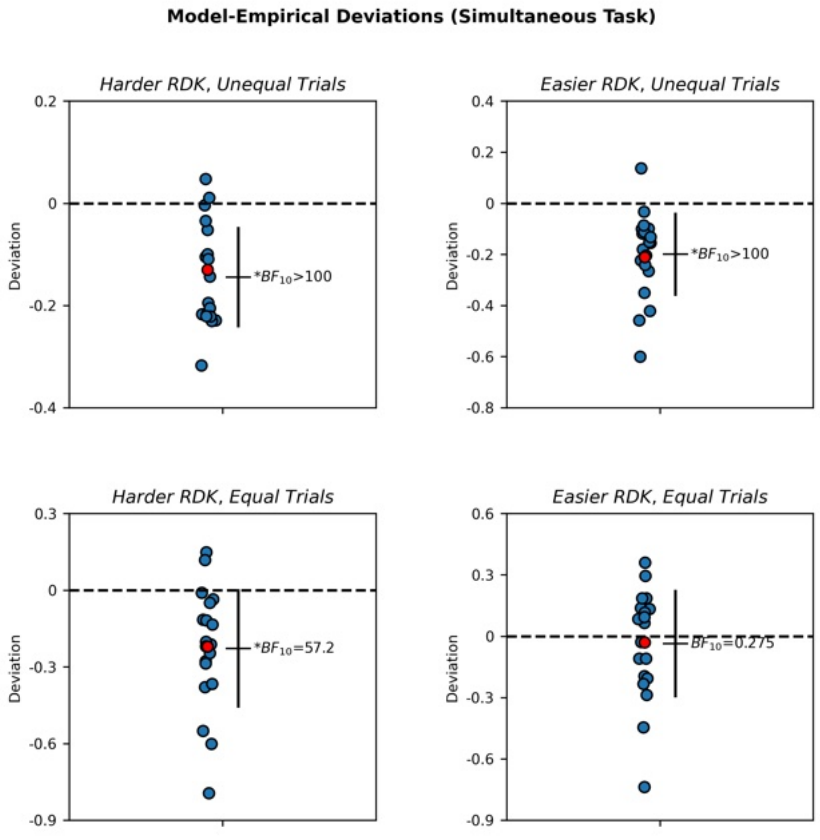
Table.3: Pooled Model-Empirical MAE

Simultaneous Task

Condition	$\boldsymbol{MAE_{model}}$		$\boldsymbol{MAE_{empirical}}$		**Deviation**		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}
Harder RDK, Unequal Trials	0.96	/	1.09	/	-0.13	/	/
Easier RDK, Unequal Trials	0.62	/	0.75	/	-0.21	/	/
Harder RDK, Equal Trials	0.78	/	0.95	/	-0.22	/	/
Easier RDK, Equal Trials	0.76	/	0.78	/	-0.03	/	/

Sequential Task

Condition	$\boldsymbol{MAE_{model}}$		$\boldsymbol{MAE_{empirical}}$		**Deviation**		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}
Harder RDK, Unequal Trials	0.83	/	0.94	/	-0.13	/	/
Easier RDK, Unequal Trials	0.49	/	0.58	/	-0.17	/	/
Harder RDK, Equal Trials	0.64	/	0.74	/	-0.15	/	/
Easier RDK, Equal Trials	0.63	/	0.65	/	-0.04	/	/


Model-Empirical Deviations (Simultaneous Task)
Harder RDK, Unequal Trials
Easier RDK, Unequal Trials
Harder RDK, Equal Trials
Easier RDK, Equal Trials
Deviation
*BF10>100
*BF10=57.2
BF10=0.275

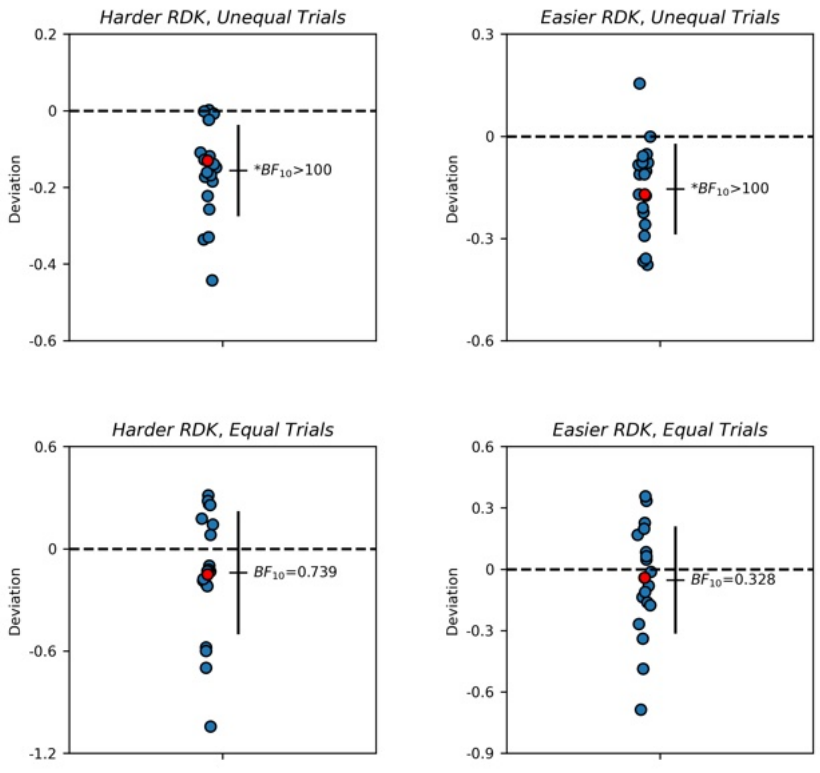

Model-Empirical Deviations (Sequential Task)
Harder RDK, Unequal Trials
Easier RDK, Unequal Trials
Harder RDK, Equal Trials
Easier RDK, Equal Trials
Deviation
*BF10>100
BF10=0.739
BF10=0.328

Fig.11: Deviations between model and human recall performance. Two subplots respectively demonstrate the results from fitting the simultaneous (top) and sequential task (bottom). Blue data points represent individual participants, red data points represent the pooled data, and error bars indicate population mean ± 1*SE*. Dotted horizontal lines indicate baselines for perfect matching (deviation = 0). The reported Bayes factors were the results from Bayesian one sample two-tailed *t*-tests or its non-parametric equivalent. As shown by the figure, the mean deviation was lower than zero across all eight targeted conditions, five of which were statistically significant (Bayesian one sample two-tailed *t*-tests: $BF_{10} > 57.2$, strong evidence for H_1), suggesting the best-fit model systematically outperformed its human counterpart across the majority of simultaneous and sequential conditions.

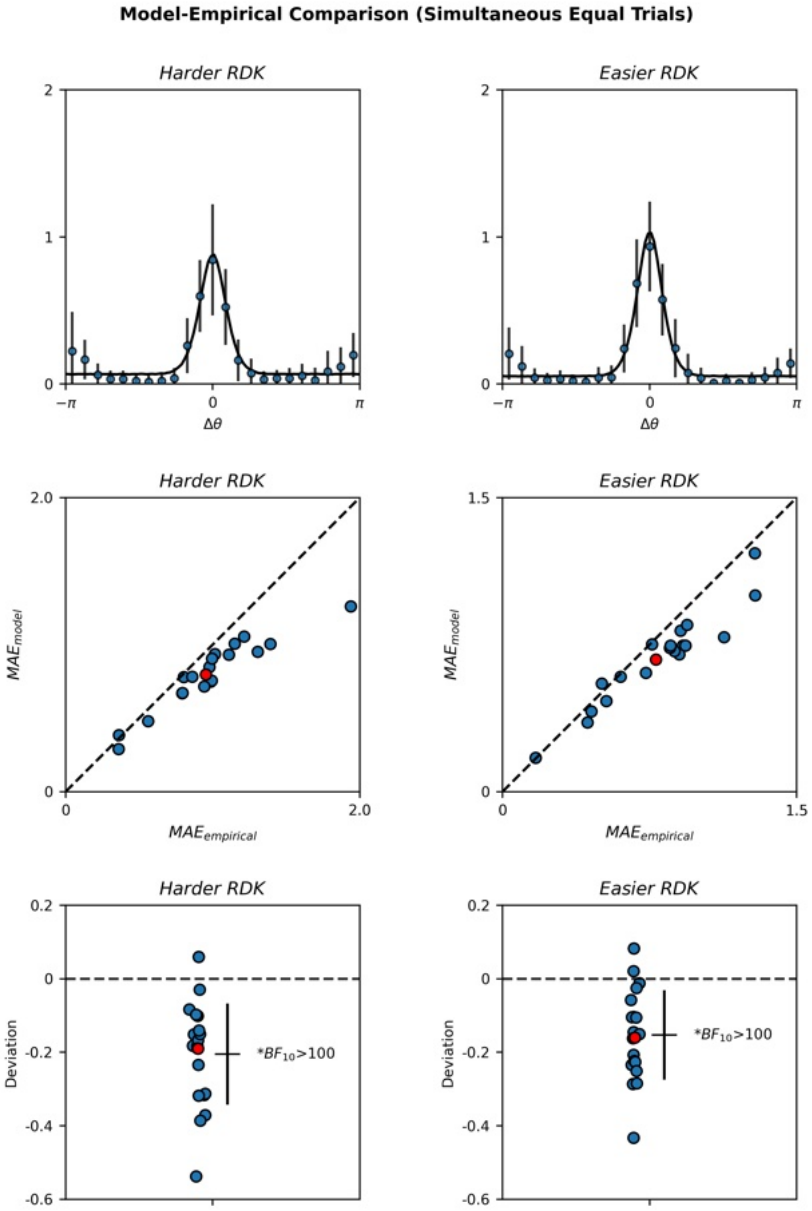
An unexpected qualitative discrepancy between the model and empirical data was the increment in recall errors near $\Delta\theta = \pi$ and $-\pi$ produced by human participants, which the model failed to replicate. Consequently, these errors made the model-generated MAE systematically lower than its corresponding empirically observed MAE, potentially explaining the negative deviations between model and human in recall performance. These "180° errors," I speculate, might be related to the inaccurate perception of motion direction at peripheral visual field, which reflect a potential flaw in either the experimental design or the population coding model that could be further improved. See Section 6 for more discussions regarding this issue.

5.2 Bias in Resource Allocation

Does the population coding model support the existence of bias in neural resource allocation between competing RDKs, as evident by the behavioural analysis?

Having verified that the model could reasonably, albeit with imperfections, capture human recall performance observed in the experiment, I quantitatively estimated the degree of bias in neural resource allocation in the simultaneous and sequential task. To minimise the individual variability in response to RDK coherences and to maintain a parallel comparison

with the previous behavioural analysis, I focused solely on equal trials and fit the model to both individuals' and the pooled data. As detailed by Fig.15, consistent with previous results on model fitting, despite issues regarding 180° errors and negative deviations persisted, the population coding model showed satisfactory qualitative and quantitative alignment with human recall performance across all targeted conditions. The numerical values of best-fit parameters and model-generated MAE were summarised in Table.4 and Table.5.



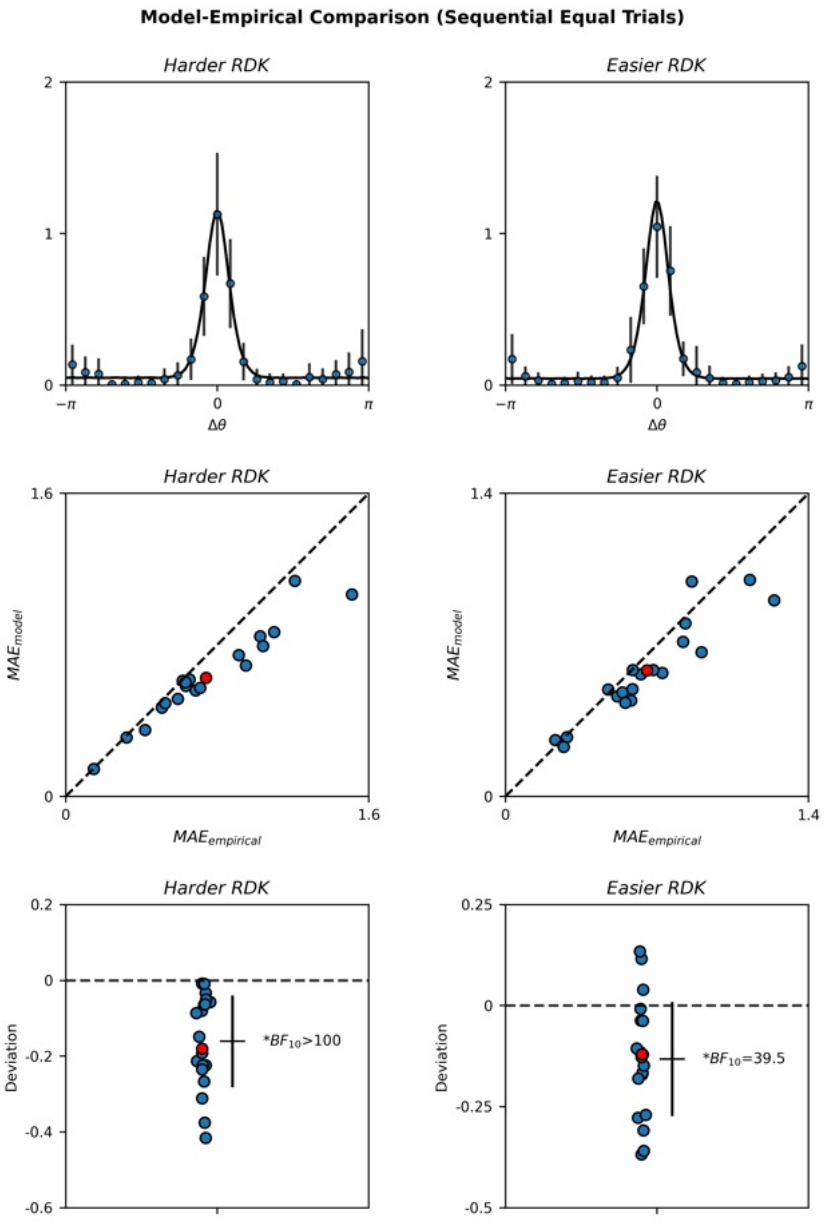


Fig.15: Comparisons between model and human recall performance. Two subplots respectively demonstrate the results from fitting the simultaneous (top) and sequential equal trials (bottom). Top row, comparisons between model-generated and empirically observed recall error distribution. Black curves indicate the recall error distributions generated by the best-fit models based on the pooled data, blue circles indicate the mean recall error distributions observed empirically, and error bars indicate ± 1*SE*. Similar with the previous results on model fitting, the population coding model showed good qualitative fitness with empirical data across all four targeted conditions, despite the 180° errors near $\Delta\theta = \pi$ and $-\pi$ persisted. Middle row, alignments between model-generated and empirically observed MAE. Blue data points represent individual participants, red data points represent the pooled data, and diagonal lines represent perfect match. Similar with the previous results on model fitting, most data points clustered diagonally in all four targeted conditions, indicating a good alignment between model and human recall performance. Nevertheless, still, a significant proportion of data points fell below the diagonal lines, suggesting that the model-generated MAE was systematically lower than the

empirically observed MAE. Bottom row, deviations between model and human recall performance. Blue data points represent individual participants, red data points represent the pooled data, and error bars indicate population mean ± 1*SE*. Dotted horizontal lines indicate baselines for perfect matching (deviation = 0). The reported Bayes factors were the results from Bayesian one sample two-tailed *t*-tests or its non-parametric equivalent. The population mean deviation was significantly lower than zero across all four targeted conditions (Bayesian one sample two-tailed *t*-tests: $BF_{10} > 39.5$, strong evidence for H_1), suggesting that the best-fit model still outperformed human participants in recall accuracy. Overall, despite the issues regarding 180° errors and negative deviations persisted, the population coding model demonstrated good qualitative fitness and satisfactory MAE alignment with its human counterpart during both simultaneous and sequential equal trials.

Table.4: Best-fit Parameters

Simultaneous Equal Trials

Model Parameter	Individual		Pooled	
	M	*SE*	*M*	*SE*
ω	0.12	0.06	0.11	/
γ	3.6	2.2	3.0	/
α	0.56	0.10	0.56	/

Sequential Equal Trials

Model Parameter	Individual		Pooled	
	M	*SE*	*M*	*SE*
ω	0.12	0.05	0.10	/
γ	4.7	3.0	3.8	/
α	0.52	0.12	0.52	/

Table.5: Model-Empirical MAE

Simultaneous Equal Trials

Condition	MAE_{model}		$MAE_{empirical}$		Deviation		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}*
Harder RDK, Individual	0.80	0.24	0.98	0.36	-0.21	0.14	3.5×10^3
Easier RDK, Individual	0.68	0.24	0.79	0.30	-0.15	0.13	5.7×10^2
Harder RDK, Pooled	0.80	/	0.95	/	-0.19	/	/
Easier RDK, Pooled	0.67	/	0.78	/	-0.16	/	/

Sequential Equal Trials

Condition	MAE_{model}		$MAE_{empirical}$		Deviation		
	M	*SE*	*M*	*SE*	*M*	*SE*	BF_{10}*
Harder RDK, Individual	0.75	0.33	0.63	0.25	-0.16	0.12	9.9×10^2
Easier RDK, Individual	0.66	0.27	0.58	0.23	-0.13	0.15	39.5
Harder RDK, Pooled	0.74	/	0.63	/	-0.18	/	/
Easier RDK, Pooled	0.65	/	0.58	/	-0.12	/	/

*Results from Bayesian one sample two-tailed *t*-test or its non-parametric equivalent, with the null hypothesis specifying that the population mean ≠ 0.

To examine the bias in neural resource allocation in each task, I analysed the best-fit values of α derived from model fitting, which quantitatively reflect the bias in neural resource allocation towards the easier RDK. As shown by Fig.16, during simultaneous equal trials, the population mean of α was significantly higher than 0.5, the unbiased baseline ($\alpha = 0.56 \pm 0.10$; Bayesian one sample two-tailed *t*-test: $BF_{10} = 3.27$, moderate evidence for H_1), indicating that neural resources were biasedly allocated towards the easier RDK. Conversely, during sequential equal trials, the population mean of α did not statistically deviate from the unbiased baseline ($\alpha = 0.52 \pm 0.12$; Bayesian one sample two-tailed *t*-tests: $BF_{10} = 0.341$, weak evidence for H_0), suggesting a roughly equalised allocation of neural resources between the harder and easier RDK. These findings, consistent with previous behavioural analysis, support the conclusion that the allocation of neural resources is biased towards the easier RDK, and removing the attentional competition between RDKs eliminates the bias.

A subtle yet interesting result from the model fitting is the huge variation in strategies of neural resource allocation among participants, as evident by the widely distributed α for both simultaneous ($\alpha = 0.32$ to 0.67) and sequential task ($\alpha = 0.31$ to 0.76). In addition, as shown by the figure, for each participant, the value of α fluctuated greatly across tasks, although pairwise comparisons indicated that these fluctuations did not lead to systematic deviation in α between different tasks (Bayesian paired samples two-tailed *t*-tests: $BF_{10} = 0.377$, weak evidence for H_0). These results potentially indicate the individual- and task-variability in strategies of neural resource allocation, however, considering limitations regarding sample sizes and model fitness, they can also be interpreted as the by-products of large random errors in model fitting.

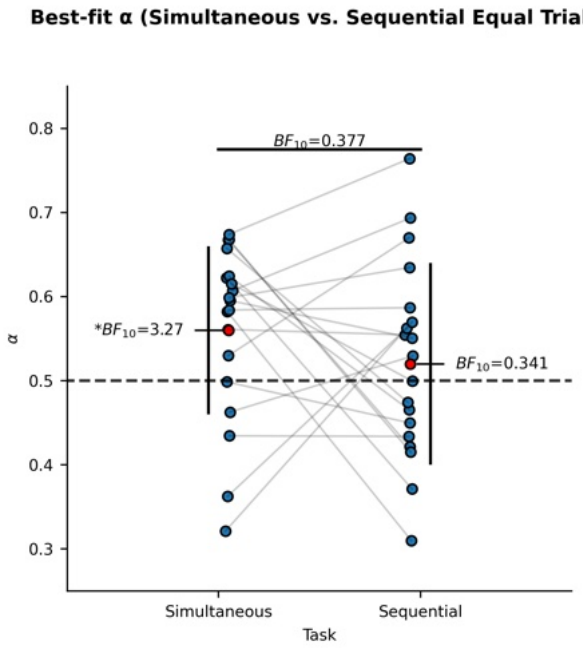


Fig.16: The bias in neural resource allocation during simultaneous and sequential equal trials. Blue data points represent individual participants and red data points represent the pooled data, with gray lines connecting best-fit α across different tasks derived from the same participant's data. Error bars indicate population mean ± 1*SE*, and the dotted horizontal line indicates the unbiased baseline (α = 0.5). The reported Bayes factors were results from Bayesian one sample (for comparisons within a task) or paired samples (for comparisons between tasks) two-tailed *t*-tests. During simultaneous probe trials, most data points situated above the unbiased baseline, with population mean significantly higher than 0.5, indicate a biased resource allocation towards the easier RDK. This bias was not systematically observed during sequential probe trials, confirming the conclusions from the previous behavioural analysis. However, an interesting observation to note from these results is the huge variation in strategies of neural resource allocation among participants, as evident by the widely distributed α for both simultaneous (α = 0.32 to 0.67) and sequential task (α = 0.31 to 0.76). Additionally, for each participant, α fluctuate greatly across tasks, although statistical analysis showed that these fluctuations were random and non-systematic at the population level.

5.3 Optimal Resource Allocation

Does the observed bias in neural resource allocation match the optimal strategy that minimises the expected error in recall?

After obtaining the best-fit models from fitting simultaneous equal trials, based on the derived best-fit parameters ω and γ, I computed the optimal bias in neural resource

allocation, α_{opt}, that minimises the expected model-generated MAE during simultaneous unequal trials, then I compared the best-fit value of α with its corresponding α_{opt}. As shown by Fig.17, at the population level, α ($\alpha = 0.56 \pm 0.10$) did not significantly deviate from α_{opt} ($\alpha_{opt} = 0.54 \pm 0.10$; Bayesian paired samples two-tailed *t*-test: $BF_{10} = 0.277$, moderate evidence for H_0), which seemingly suggests that the observed allocational strategies did not systematically differ from the optimal. However, while the population mean of α was significantly higher than the unbiased baseline, the population mean of α_{opt} did not statistically differ from 0.5 (Bayesian one sample two-tailed *t*-tests: $BF_{10} = 1.01$, weak evidence for H_1). In other words, while the observed resource allocation was biased towards the easier RDK, the optimal strategy suggests that neural resources should be allocated roughly equally among RDKs. Furthermore, as illustrated by Fig.18, at the individual level, α showed no correlation with α_{opt} (Bayesian Pearson's correlation analysis: $r = -0.03$, $BF_{10} = 0.285$, moderate evidence for H_0), indicating that participants did not follow the optimal strategy of neural resource allocation, but instead, they took an arbitrary one. Thus, at both population and individual level, the disagreements between α and α_{opt} indicate that participants did not allocate their neural resources optimally during the simultaneous task, at least not according to the strategy outlined by the presented population coding model, which aimed at minimising the expected MAE in recall.

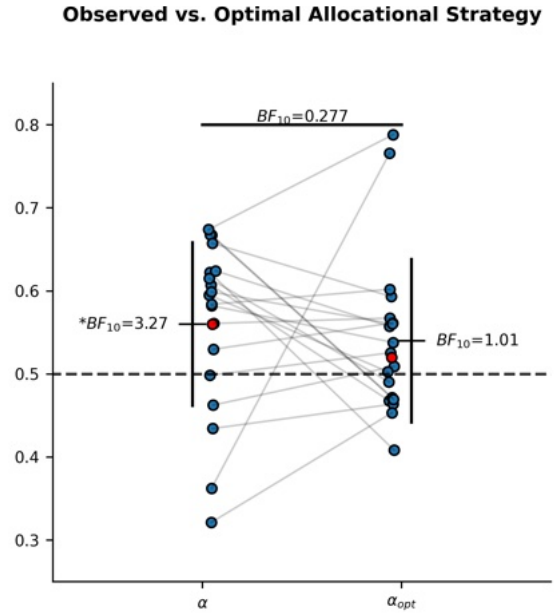


Fig.17: Comparisons between the observed and optimal strategy of neural resource allocation during simultaneous unequal trials. Blue data points represent individual participants and red data points represent the pooled data, with gray lines connecting best-fit α across different tasks derived from the same participant's data. Error bars indicate population mean ± 1*SE*, and the dotted horizontal line indicates the unbiased baseline. The reported Bayes factors were results from Bayesian one sample (for comparisons within a task) or paired samples (for comparisons between tasks) two-tailed *t*-tests. At the population level, despite α did not deviate statistically from α_{opt}, unlike α, the population mean of α_{opt} showed no significant deviation from the unbiased baseline, implying that the optimal strategy should be roughly equalised, instead of biased, allocation. This suggests a qualitative difference between the observed and optimal allocational strategy during the simultaneous task.

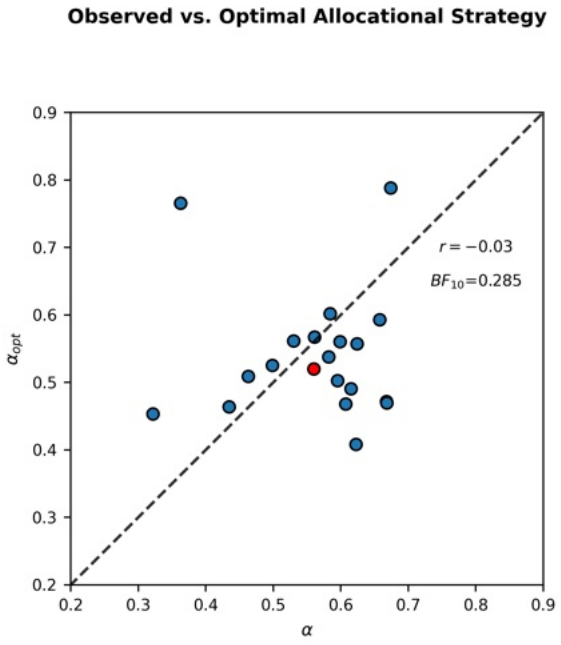


Fig.17: Correlations between the observed and optimal allocational strategy during the simultaneous task. Blue data points represent individual participants and red data points represent the pooled data. The reported Pearson's *r* and BF_{10} are results from Bayesian Pearson's correlation analyses. The horizontal and vertical position of a data point reflects the observed degree of bias, α, and optimal degree of bias, α_{opt}, in neural resource allocation, respectively. If a participant strictly followed the optimal strategy to allocate neural resources, his/her corresponding data point would be located on the dotted diagonal line. As shown by the figure, α showed no correlation with α_{opt}, indicating that the optimal allocational strategy was not obeyed. Interestingly, more than half of data points, including the red data point, fell below the diagonal line, seemingly suggests that human participants tend to sub-optimally allocate more resources towards the easier RDK compared to the optimal strategy.

6. Discussion

6.1 180° Errors

In the behavioural experiment, empirically measured recall distributions abnormally peaked near 180° and −180°, particularly in the simultaneous task, which qualitatively deviate from the expected human recall performance (see Fig.9). These 180° errors are rather unexpected, but they have been reported in perceptual tasks involving RDKs, particularly under low coherences (Bae & Luck, 2022; Keown et al., 2023).

The peculiar nature of human motion perception could provide a plausible explanation for these 180° errors. Past electrophysiological studies have identified two subpopulations of neurons in visual areas that are sensitive to motion direction: the unidirectional neurons, which are tuned to motion directions, and the bidirectional neurons, which are tuned to orientations of motion axis (Movshon et al., 1985; Okamoto et al., 1999). When a motion stimulus is displayed at peripheral visual field, due to the eccentricity effect (Hirano et al., 2023), response acuity of unidirectional neurons drops (Orban et al., 1986; see Bower et al., 2012 and Mckee & Nakayama, 1984 for psychophysical evidence), yet the residual activities of bidirectional neurons could still provide some information regarding motion axis orientation. Thus, on occasional trials, participants may struggle to perceive or remember the motion direction of coherent dots in the probed RDK, but they might still capture the orientation of its motion axis. Under these scenarios, instead of submitting the actual recalled motion direction, participants may choose to report an arbitrary direction alone the remembered motion axis, which, on half of these trials, leads to 180° errors. This strategy of relying on motion axis in recall has been reported by several participants, including the author and another experimenter, when attempting the task. Reducing the eccentricity of

visual stimuli by presenting RDKs closer to the fixation point is perhaps a plausible way to mitigate these 180° errors.

6.2 Model Improvements

The results of model fitting revealed several qualitative and quantitative mismatches between the population coding model and human behaviour, particularly regarding 180° errors. To improve the model, following the previous idea of unidirectional and bidirectional neurons, I modified the original model by introducing a second neural population, which specifically responds to motion axis orientations. The orientation-tuned neurons in the new neural population followed the tuning function:

$$f_i = \exp(\omega^{-1}\cos(2(\theta - \varphi_i)) - 1) \tag{14}$$

which has a similar shape, but contains two maxima, compared with the tuning curve of direction-tuned neurons. Other procedures regarding neural activation, spike generation, and orientation decoding of the new neural population were identical with the original population, under a new set of model parameters.

After acquiring the decoded motion direction and decoded motion axis orientation, the recalled motion direction produced by the modified system, $\hat{\theta}$, was generated via:

$$\hat{\theta} = \begin{cases} \hat{\theta}_o, & with\ probability\ of\ \beta \\ \hat{\theta}_n, & with\ probability\ of\ 1-\beta \end{cases} \tag{15}$$

where $\hat{\theta}_o$ is the motion direction decoded by the original neural population, $\hat{\theta}_n$ is a randomly selected direction along the orientation decoded by the new neural population, and β is a weight constant, ranging from 0 to 1, which reflects the lapse rate in response.

To test whether the modifications regarding motion axis orientation enhance model performance, I fit both the original and modified model to the empirically measured recall errors during simultaneous unequal trials for the harder RDK, where 180° errors were most prominent. To test the fitness of each model with empirical data, following the previously outlined procedures, I estimated the MAE generated by the best-fit model and analysed the deviation between human and model in average recall performance. Additionally, to compare the performance between the original and modified model, I compared the Akaike information criteria[6] (AIC; Akaike, 1974) for both models from fitting the pooled data. As shown by Fig.18, the modified model showed exceptional qualitative fitness with the empirical data, as it nicely accounted for the 180° errors generated by human participants. In addition, MAE generated by the modified model closely aligned with empirically observed MAE, indicating a nearly perfect quantitative matching between model and human in recall performance, although systematic negative deviations still persisted. These results, as well as the AIC comparison (original model's AIC = 3771; modified model's AIC = 3597), indicate that modifications regarding motion axis orientation can significantly improve the performance of the population coding model. Although not elaborated further in this report, it would be interesting to see how the results of model fitting would change under this modified model.

[6] AIC = $2k - 2\ln(\hat{L})$, where k is the number of free parameters and $\hat{L}$ is the maximised likelihood function for the model.

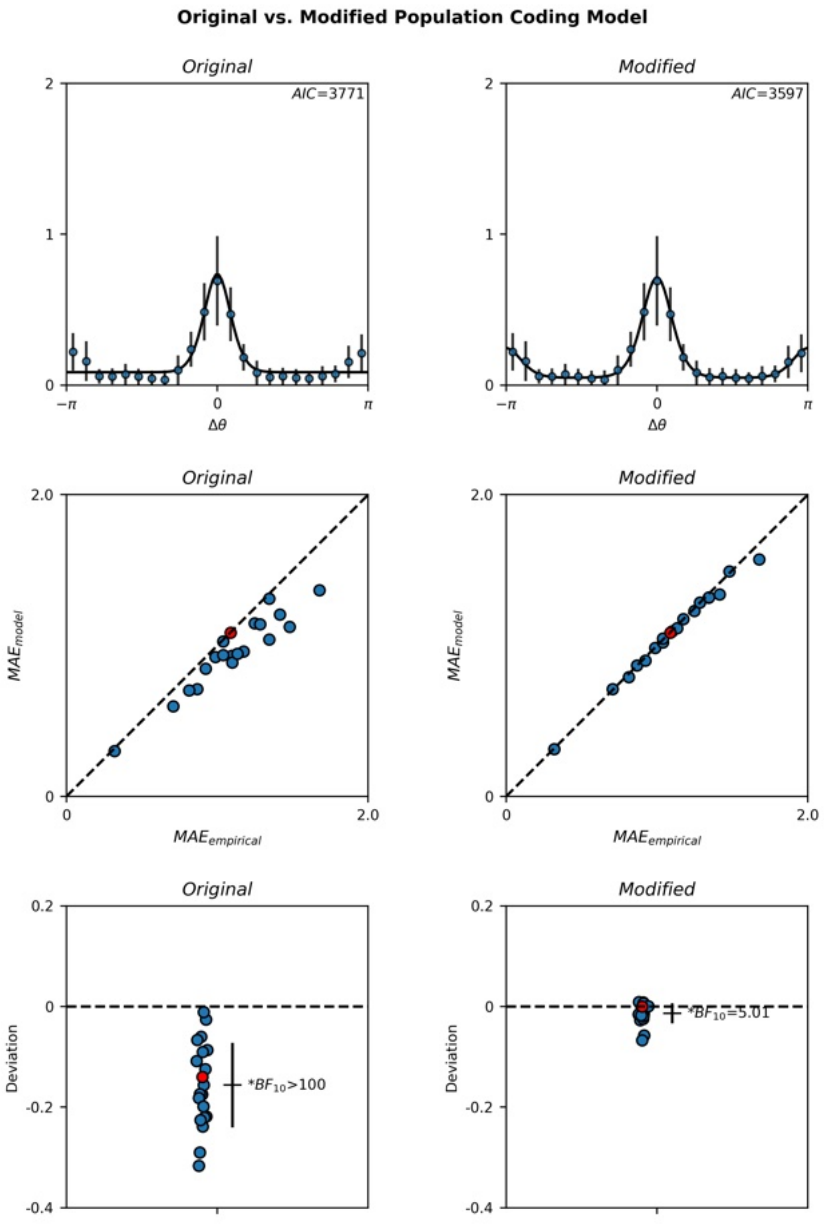


Fig.18: Comparisons between the original (left) and modified population coding model (right). Top row, comparisons between model-generated and empirically observed recall error distribution. Black curves indicate the recall error distributions generated by the best-fit models based on the pooled data, and blue circles indicate the mean recall error distributions observed empirically (with error bars indicate ± 1*SE*). The modified model showed better qualitative fitness compared to the original model, as it nicely captured the 180° errors produced by human participants. Middle row, alignments between model-generated and empirically observed MAE. Blue data points represent individual participants, red data points represent the pooled data, and the diagonal lines represent perfect match. Unlike the original model, MAE generated by the modified model aligned nearly perfectly with the empirically measured MAE for each participant, as all data points were situated on the diagonal line. Bottom row, deviations between model and human recall performance. Blue data points represent individual participants, red data points represent the pooled data, and error bars indicate population mean ± 1*SE*.

Dotted horizontal lines indicate baselines for perfect matching (deviation = 0). The reported Bayes factors were the results from Bayesian one sample two-tailed *t*-tests or its non-parametric equivalent. The deviation between the modified model and human (deviation = –0.01 ± 0.02) was significantly smaller than the deviation between the original model and human in recall performance (deviation = –0.16 ± 0.09; Bayesian paired samples two-tailed *t*-tests: $BF_{10} = 1.4\times10^4$, extremely strong evidence for H_1). However, statistically, the population mean of the modified model's deviations was still statistically below zero (Bayesian one sample two-tailed *t*-tests: $BF_{10} = 5.01$, moderate evidence for H_1). Overall, the modified population coding model showed exceptional qualitative and quantitative fitness to the empirical data and outperformed the original model.

6.3 Imperfect Observers

As the end of Section 5, I concluded that human observers tend to biasedly allocate more neural resources towards the easier item, which seemingly deviates from the theoretical optimum. This conclusion is rather surprising, as previous studies have shown that human observers do tend to allocate VWM neural resources in a mathematically optimal strategy that minimises the expected recall errors during behavioural tasks (Yoo et al., 2018; Bays, 2014).

Why do human observers sub-optimally allocate more neural resources towards the easier item? Aside from the methodological issues regarding flaws in the population coding model or random errors in parameter estimation, several alternative explanations are worth considering.

One possibility is that humans are "picky" observers. Different participants could be sensitive to recall errors of different magnitudes, i.e., possess non-linear loss functions[7], and this could lead to sub-optimal allocational strategies. Most participants preferentially attend to small recall errors but ignore the larger ones, thus they sub-optimally allocate more neural

[7] Loss functions are transformations that map the physical value of a variable, i.e., recall error, onto a psychological loss or "penalty." For a non-linear loss function, the loss increases disproportionally with the magnitude of recall error.

resources towards the easier item to perfect their responses. Conversely, few participants preferentially attend to large recall errors yet overlook the smaller ones, thus they sub-optimally allocate more neural resources towards the harder item to avoid extreme errors. The use of non-linear loss function in visual estimation has been previously reported (Körding & Wolpert, 2004), and the individual variability in loss functions could explain not only the deviation between human and optimum but also the large variation in resource allocational strategies among participants (see Fig.16). To straightforwardly verify this possibility, one could recalculate the optimal biases in neural resource allocation under varying loss functions, then reassess their correlations with the observed behaviour.

Another possibility is that humans are "confident" observers. Previous studies have shown that humans can accurately evaluate the representational precision of items in their VWM via metacognitive confidence ratings (Honig et al., 2020), which can occur without external feedbacks (Guggenmos et al., 2016) and play critical roles in maintaining item representations within VWM (Rademaker et al., 2012). Thus, the observed sub-optimal strategy in neural resource allocation could be originated from the variations in confidence ratings for different items. As a reasonably speculation, when observing a pair of items, the harder one elicits a nosier spiking pattern among the VWM neural population, resulting in a vague item representation with a lower signal-to-noise ratio, and, consequently, a lower internal confidence rating. In response to this, few participants choose to compensate by sub-optimally allocating more resources towards the harder item, while most participants decide to relinquish by sub-optimally allocating less resources towards the harder item. This, again, accounts for not only the mismatch between human and optimum but also the observed individual variability in resource allocational strategies. To verify this possibility, one could modify the behavioural task by asking participants to provide a subjective confidence score

regarding their response after each trial. Subsequently, one could analyse whether the degree of allocational bias quantitatively correlates with this confidence rating.

The final possibility is that humans are "conservational" observers. Prominent cognitive theories on decision-making suggest that humans are not necessarily ideal inference machines that minimise the absolute errors in decisions or actions, but occasionally cognitive "misers" that attempt to optimise performances within neurobiological constrains (Kahneman, 2011; Lieder & Griffiths, 2019). When strategising the allocation of neural resources, human observers consider the cognitive efforts, i.e., neural costs, associated with their actions. Under this criterion, the optimal degree of bias would be the one that not only maximises the expected recall error but also minimises the total spike count generated by the VWM neural population. Since the harder item requires a greater cognitive effort to perceive and remember, participants, for the sake of price-performance ratio, may choose to sub-optimally allocate less neural resources towards it. To verify this possibility, one would need to quantitatively estimate the total neural cost of the VWM neural population under different degrees of allocational bias, recalculate the optimal allocational strategy that maximises the price-performance ratio, then reassess its correlation with the observed behaviour.

6.4 Future Directions

Expanding on this research project, several topics are worth for further investigation. These include:

- Changing the experimental settings to reduce 180° errors and reanalysing the empirical data using the modified population coding model.
- Studying the bias in neural resource allocation under other combinations of RDK coherences with three or more RDKs in a visual scene.

- Investigating the strategy of neural resource allocation using more ecologically valid paradigm with naturalistic visual stimuli.

Conclusion

In this research project, I show that item difficulty can influence the allocation of VWM neural resources among competing items in a visual scene. Specifically, human observers tend to allocate more neural resources towards the item that appears to be easier to perceive and remember. This allocational bias is linked with the competition for attention among items, and the degree of the observed bias seemingly deviates from the theoretical optimum.

Reference

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. https://doi.org/10.1016/j.neuroscience.2005.08.023

Bae, G.-Y., & Luck, S. J. (2022). Perception of opposite-direction motion in random dot kinematograms. *Visual Cognition*, *30*(4), 289–303. https://doi.org/10.1080/13506285.2022.2052216

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Bower, J. D., Bian, Z., & Andersen, G. J. (2012). Effects of retinal eccentricity and acuity on global motion processing. *Attention, Perception & Psychophysics*, *74*(5), 942. https://doi.org/10.3758/s13414-012-0283-2

Brissenden, J. A., Adkins, T. J., Hsu, Y. T., & Lee, T. G. (2023). Reward influences the allocation but not the availability of resources in visual working memory. *Journal of Experimental Psychology. General*, *152*(7), 1825–1839. https://doi.org/10.1037/xge0001370

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic Updating of Working Memory Resources for Visual Objects. *Journal of Neuroscience*, *31*(23), 8502–8511. https://doi.org/10.1523/JNEUROSCI.0208-11.2011

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife*, *5*, e13388. https://doi.org/10.7554/eLife.13388

Heeger, D. J., Huk, A. C., Geisler, W. S., & Albrecht, D. G. (2000). Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? *Nature Neuroscience*, *3*(7), 631–633. https://doi.org/10.1038/76572

Hirano, R., Numasawa, K., Yoshimura, Y., Miyamoto, T., Kizuka, T., & Ono, S. (2023). The effect of eccentricity on visual motion prediction in peripheral vision. *Physiological Reports*, *11*(22), e15877. https://doi.org/10.14814/phy2.15877

Honig, M., Ma, W. J., & Fougnie, D. (2020). Humans incorporate trial-to-trial working memory uncertainty into rewarded decisions. *Proceedings of the National Academy of Sciences*, *117*(15), 8391–8397. https://doi.org/10.1073/pnas.1918143117

Huynh Cong, S., & Kerzel, D. (2021). Allocation of resources in working memory: Theoretical and empirical implications for visual search. *Psychonomic Bulletin & Review*, *28*(4), 1093–1111. https://doi.org/10.3758/s13423-021-01881-5

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience*, *23*(1), 315–341. https://doi.org/10.1146/annurev.neuro.23.1.315

Körding, K. P., & Wolpert, D. M. (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences*, *101*(26), 9839–9842. https://doi.org/10.1073/pnas.0308394101

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Louie, K., & Glimcher, P. W. (2019). Normalization Principles in Computational Neuroscience. In *Oxford Research Encyclopedia of Neuroscience*. https://doi.org/10.1093/acrefore/9780190264086.013.43

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Mc Keown, P., Corbett, E., & O'Connell, R. G. (2023). Investigating opposite direction motion reports in random dot kinematograms. *Visual Cognition*, *31*(8), 559–570. https://doi.org/10.1080/13506285.2024.2315779

Mckee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research*, *24*(1), 25–32. https://doi.org/10.1016/0042-6989(84)90140-8

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The Analysis of Moving Visual Patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern*

Recognition Mechanisms (Vol. 11, pp. 117–151). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-09224-8_7

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Okamoto, H., Kawakami, S., Saito, H., Hida, E., Odajima, K., Tamanoi, D., & Ohno, H. (1999). MT neurons in the macaque exhibited two types of bimodal direction tuning as predicted by a model for visual motion detection. *Vision Research*, *39*(20), 3465–3479. https://doi.org/10.1016/S0042-6989(99)00073-5

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. *Journal of Neurophysiology*, *56*(2), 462–480. https://doi.org/10.1152/jn.1986.56.2.462

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*(2), 125–132. https://doi.org/10.1038/35039062

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), 21. https://doi.org/10.1167/12.13.21

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schneegans, S., Taylor, R., & Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences*, *117*(34), 20959–20968. https://doi.org/10.1073/pnas.2004306117

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports*, *8*(1), 16162. https://doi.org/10.1038/s41598-018-34282-1

Zaharia, A. D., Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2019). Compound Stimuli Reveal the Structure of Visual Motion Selectivity in Macaque MT Neurons. *eNeuro*, *6*(6). https://doi.org/10.1523/ENEURO.0258-19.2019

Appendix

I. Ethics

The behavioural experiment of this research project was reviewed and approved by the Department of Psychology Research Ethics Committee at the University of Cambridge. Before the experiment, participants were clearly informed about the basic experimental procedure, their freedom to withdraw, and the confidentiality and anonymity of data collection. The tasks involved were non-invasive and associated with no physical risk, although few participants reported feeling perceptually tired when attempting the experiment. After the experiment, participants were briefed on the key manipulations that were concealed from they during the experiment, and they were provided a written debrief sheet explaining the scientific aim of the study.

II. Ethical Paperwork

(Omitted)

III. Project Plan

(Omitted)

IV. Statistical Setting in JASP

Bayesian one sample and paired samples t-tests. Prior distribution for the null hypothesis, H_0: $\delta = 0$, was set as a Cauchy distribution with $r = 1/\sqrt{2}$ (the default setting in JASP).

Bayesian Wilcoxon one sample and paired samples sign-ranked tests. Prior distribution for the null hypothesis, H_0: $\delta = 0$, was set as a Cauchy distribution with $r = 1/\sqrt{2}$. Data augmentation algorithm in JASP was set to the maximal number of iterations (10^4 iterations).

Bayesian Pearson's correlation analysis. Prior distribution was set as a stretched beta distribution, with width set as 1 (the default setting in JASP).