Which visual working memory model accounts best for target representation in the attentional blink?

Shuyao Wang[1], Aytaç Karabay[1,2], Elkan G. Akyürek[1]

[1] Department of Psychology, Experimental Psychology, University of Groningen, Groningen, The Netherlands

[2] Science Division, New York University Abu Dhabi, Abu Dhabi, UAE

Running head: Working memory models in the attentional blink

Word count: 14297

Address correspondence to:

Shuyao Wang

Department of Psychology, Experimental Psychology, University of Groningen

Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

Email: shuyao.wang@rug.nl

Abstract

People often fail to detect the second of two targets when there is a short time interval of ~500 msec or less between them. This phenomenon is known as the attentional blink (AB). Accumulating evidence suggests that the AB is a result of a failure to select and consolidate the second target in working memory. The current literature has assumed that the standard mixture model of visual working memory (VWM) explains representation in the AB better than resource-based VWM models. However, no existing study has systematically compared VWM models in the AB domain. Here, we present a comparison of eight widely-used VWM models in four different AB datasets from three separate laboratories. We fitted each model and computed the Bayesian information criterion (BIC) values at an individual level, across different conditions and experiments, based on which we compared the models by their average model ranks. We found that, for most experiments presented here, the standard mixture model, the slot model, and their variants do outperform the others. We nevertheless also observed that certain details, such as the stimuli or spatial arrangement of targets used in the AB task, can result in different model rankings. Our results can help researchers to select the best model for their AB studies in the future, and thereby gain a better understanding of their data.

Introduction

It is challenging for us to consciously perceive all the information in our surroundings. Here attention shines as a useful, selective cognitive function that is crucial to our daily lives. The limited nature of attention has already attracted considerable research. One of the most extensively researched paradigms is the attentional blink (AB), which refers to the phenomenon that people often fail to identify the second (T2) of two briefly presented target stimuli when the time interval (or lag) between them is within 200 – 500 msec (Broadbent & Broadbent, 1987; Raymond et al., 1992). The most commonly used paradigm in AB studies is rapid serial visual presentation (RSVP; Potter & Levy, 1969), in which a quick succession of visual stimuli is shown at the same spot. The observer's task is to identify certain targets while ignoring other distractors. In most AB studies, these targets are categorical items, such as letters, numbers, or images (for a review, see Dux & Marois, 2009; Martens & Wyble, 2010).

Although the RSVP paradigm with categorical items has been demonstrated to be an excellent method for investigating the AB phenomenon, it can only evaluate task performance by measuring discrimination or detection accuracy. Actually evaluating the quality of target representations that are kept in mind during the AB is thereby challenging. However, using a continuous-report task in RSVP has recently enabled researchers to further investigate the target representations in the AB. The continuous-report task has been frequently utilized in the visual working memory (VWM) field (Bays et al., 2009; Fougnie & Alvarez, 2011; Kool et al., 2014; Oberauer et al., 2017; Zhang & Luck, 2008). During the task, participants are required to memorize targeted visual items, which are identified by a basic visual property that varies on a continuous, usually circular, dimension, such as color or orientation. In the reproduction stage, participants are then asked to recall the target on a continuous response scale, such as recreating its orientation by rotating a Gabor grating (Wilken & Ma, 2004). Consequently, the performance of a continuous-report task is not measured in a binary format, but rather as the

divergence between the target value and the corresponding response. The observer's internal response to the target, *s*, follows a Von Mises distribution (the circular analog of the normal distribution), following the previous assumption (Wilken & Ma, 2004):

$$p(x|s,\kappa) = VM(x;s,\kappa) \equiv \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-s)},$$

where I_0 is the modified Bessel function of order zero and κ is a concentration parameter.

Many models have been proposed to explain the representations in VWM in continuous-report tasks (Alvarez & Cavanagh, 2004; Bays et al., 2009; Brady & Alvarez, 2011; Fougnie et al., 2012; Oberauer et al., 2012; van den Berg et al., 2012; van den Berg & Ma, 2018; Zhang & Luck, 2008). Moreover, a debate has unfolded between two groups of models that have different assumptions regarding the nature of VWM. One holds the view that VWM capacity is defined by a limited fixed number of discrete "slots" that are used to store items (Alvarez & Cavanagh, 2004; Bays et al., 2009; Zhang & Luck, 2008). The other group hypothesizes that VWM capacity is a continuous resource that can be distributed freely over all items (Brady & Alvarez, 2011; Fougnie et al., 2012; van den Berg et al., 2012). These models' different theoretical bases lead to different quantitative predictions about the response error distributions in the continuous-report task. Hence, numerous studies have attempted to compare these models and assess their ability to explain the empirical data in the VWM domain (Donkin et al., 2013; Oberauer, 2021; van den Berg et al., 2014).

These VWM models have recently also received attention in the AB field. Several studies have combined continuous reports in the AB with VWM models to investigate the nature of target representations. Asplund et al. (2014) applied a mixture-model analysis (Zhang & Luck, 2008) to data obtained by explicitly asking participants to recall their perception of T2 along a continuous circular dimension. The authors estimated the precision of T2 representations and the proportion of guess responses for T2. They found an increased

percentage of random guess trials at shorter lags, whereas lag had no effect on T2 precision. Their findings indicated that AB affects T2 perception in an all-or-none manner.

Yet, in another study, Karabay et al. (2022) demonstrated that T2 awareness during the AB can be both gradual and discrete, using hierarchical Bayesian estimation to examine how the precision and guess rate parameters of the standard mixture model varied, based on different time intervals between two targets. Interestingly, they found that when the identification of the first target (T1) required the focus of attention on a single spatial location only, awareness of T2 was discrete (guess rate varied); on the contrary, when T1 identification induced a spatial spread of attention, awareness of T2 was gradual (precision varied). Similarly, Sy et al. (2021) evaluated T2 performance using both discrete mixture model and variable resource model. Intriguingly, their findings also confirmed that the temporal loss of T2 information during the AB might be either gradual or discrete. Specifically, when two targets (T1 and T2) shared a visual feature, their results indicated a gradual loss of T2 precision, with the variable resource model providing a better fit to the data. In contrast, they observed a discrete loss of T2 information, and the mixture model performed better than the variable resource model in terms of fitting the data, when participants had to switch their attention to different target features (i.e., color for T1 and orientation for T2).

As mentioned above, the integration of these VWM models into the AB domain has effectively proved these models' ability to account for the AB data. However, findings about which *model* is actually better were mixed. Asplund et al. (2014) gave substantial support for the superiority of the standard mixture model. However, Sy et al. (2021) suggested that both the standard mixture model and variable resource model can be fitted to the AB data well, but under different experimental designs. In addition, previous comparisons of these VWM models in the AB domain have contained limitations regarding their implementations of the models. First, earlier studies have mainly focused on the comparison between the standard mixture

model and the variable resource model. The other commonly employed VWM models, which make distinct predictions about the response distribution, are rarely considered. For example, the *Ensemble Integration model* (Brady & Alvarez, 2011) assumes that other non-target items impact the report of a specific target item. In the context of the AB, this hypothesis is supported by Hommel and Akyürek (2005; see also Akyürek et al., 2012), who discovered that temporal integration occurred with successive targets in RSVP. Second, previous model comparisons in the AB task were built on different experimental conditions and model selection criteria (Asplund et al., 2014; Tang et al., 2020), making them hard to compare. In sum, a comprehensive assessment of the performance of these models in AB domain is lacking to date.

Our study aims to provide the first systematical comparison between the eight commonly used VWM models in the AB domain: the *Standard Mixture model*, *the Slot model*, the *Slots plus Averaging model*, the *Slots plus Resource model*, the *Swap model*, the *Ensemble Integration model*, and two *Variable Precision models*. Previous research has indicated that differences may exist between the models based on "slots", and the models based on variable resources in their measurement of target representations and their relationship to the conscious perception in the AB. Therefore, it is crucial to understand which model(s) provide a more accurate explanation than others, for the empirical data underlying the AB task. Our study uses continuous-report data from three previous studies, each from a different group of researchers (Asplund et al., 2014; Karabay et al., 2022; Tang et al., 2020), and one newly conducted experiment, to ensure the generalizability of our inferences.

Method

Model Details

Standard mixture model (StM). The *StM model* (Zhang & Luck, 2008) proposes a mixture distribution for response errors. When an item has been stored into VWM, the response value tends to center around the actual value, forming a Von Mises distribution, with the standard deviation (*sd*) representing the precision of the memorized representation. When there is no information about the item in VWM, the reported value should be random and fall into a uniform distribution, with the guess rate indicated by this uniform distribution. Formally, the response error distribution in this *StM model* is given by

$$p(x|s,g,sd) = (1-g)VM\big(s-x;0,\kappa(sd)\big) + g\frac{1}{2\pi},$$

where x is the reported value, s is the target value, and g is the guess rate, representing the probability of random guessing. $\kappa(sd)$ denotes the concentration parameter κ of the Von Mises distribution, corresponding to the standard deviation sd in the normal distribution.

Two-component mixture models: Models that assume VWM has a limited number of slots

The family of slot models (the *Slot*, *slots plus averaging*, and *slots plus resource models*) is built upon the core assumption that VWM consists of a limited number of discrete slots for storing visual representations (Cowan, 2001; Luck & Vogel, 1997). If the target object has been stored in one of these slots, its information is maintained in VWM. If the item has not been stored in any of these slots, then no information about it remains in VWM.

Slot model. The classic *Slot model* (Cowan, 2001; Luck & Vogel, 1997) not only assumes the existence of limited slots in VWM, but also makes an additional assumption about the fixed number of these slots. If the number of to-be-remembered items, also called set size, is larger than the maximum number of the available slots, the excess items are ignored. Hence, in the *Slot model*, the guess rate g is correlated with the maximum number of slots. Formally, the response error distribution is given by

$$p(x|s,g,sd) = (1-g)VM\big(s-x;0,\kappa(sd)\big) + g\frac{1}{2\pi},$$

The guess rate g depends on the predicted *capacity K* in the way that

$$g = (1-\max(0,\min(1,\frac{capacity\ K}{set\ size}))).$$

Slots plus averaging model (SA). Zhang and Luck (2008) proposed an evolved version of the *Slot model* – the *Slots plus Averaging model*, which hypothesizes that multiple memory slots can be used to store a single item. When memory is probed, the observer reports the average of these multiple representations. Under this premise, this model further makes specific predictions about the precision variation range. Similar to the *Slot model*, the capacity *K* refers to the available number of memory slots. If the set size is decreased below the capacity *K*, the reported *sd* varies from a worse value (only one slot is allocated to the probed item) to a better value (all the available slots are allocated to the probed item). Specifically, the response error distribution of this model is written by

$$p(x|s,g,sd) = (1-g)\gamma VM\big(s-x;0,\kappa(betterSD)\big)$$
$$+(1-g)(1-\gamma)VM\big(s-x;0,\kappa(worseSD)\big) + g\frac{1}{2\pi},$$

where $worseSD$ equals the *sd* of a single slot, and $betterSD$ is equal to a single slot divided by the square root of the average number of slots, which are used to store the probed item (Palmer, 1990). Likewise, the guess rate is given by

$$g = (1-\max(0,\min(1,\frac{capacity\ K}{set\ size\ N}))).$$

And γ is the proportion of the items that get an extra slot, depending on the *capacity K*:

$$\gamma = \frac{number\ of\ items\ with\ extra\ slot}{number\ of\ items\ represented} = \frac{K\ \%\ N}{number\ of\ items\ represneted}.$$

Slots plus resource model (SR). Although the *SR model* (Awh et al., 2007; Zhang & Luck, 2008) supports the same assumption as the models mentioned above, in which a fixed and limited number of items can be stored in VWM, it further assumes a distinct approach of allocating VWM resources. Specifically, this model allocates most of the VWM resources to a

probed item, leaving only a few resources for other, relatively irrelevant items. Hence, when set size (the number of probed items) do not exceed the VWM limits, the precision (*sd*) here can vary from a very small value to a very large one, as set size decreases. Formally, the response error distribution is given by

$$p(x|s,g,bestSD) = (1-g)VM\big(s-x;0,\kappa(sd)\big) + g\frac{1}{2\pi},$$

$$= (1-g)VM\big(s-x;0,\kappa(bestSD * sqrt(N)\big) + g\frac{1}{2\pi},$$

where $bestSD$ denotes the precision when the majority of VWM resources are used for storing a single item, and N is the number of items that can be stored in VWM, depending on the *capacity K*:

$$N = \min(capacity\ K, set\ size\ n).$$

The guess rate is represented as

$$g = (1 - \max(0, \min(1, \frac{capacity\ K}{set\ size\ N}))).$$

The variable precision models: Models that assume VWM consists of a resource pool

The *VP models* (Fougnie et al., 2012; Van den Berg et al., 2012) propose that VWM resources are allocated to items in a continuous and variable style, which causes the precision of VWM representation to vary across items and trials. It should be noted that there are two key differences between the *VP models* and the *SR model*, although they both propose that VWM is not (entirely) quantized. First, the *VP models* suggest that the limitation of VWM is defined by the representation quality instead of the number of memorized items. Second, the *VP models* imply a variability of VWM resource across trials and items, while the precision is equal across items and trials (given a fixed set size) in the *SR model*. In the original *VP model* proposed by van den Berg et al. (2012), instead of a complete loss of item information, the quality of that item representation may be quite poor (leading to a very low precision). Therefore, the response distribution is given by

$$p(x|s,J) = VM(x; s, \kappa(J)).$$

Here, κ is the concentration parameter that reflects the memory's precision J. The core hypothesis in the *VP model* is that it assumes the precision J varies independently across items and trials. Furthermore, the precision J itself follows a specific distribution as a random variable.

Van den Berg et al. (2012) further assume this specific distribution of precision J to be Gamma with mean precision $\bar{J}$ and scale parameter τ:

$$p(J|\bar{J};\tau) = Gamma(J|\bar{J};\tau).$$

Therefore, the response distribution of the original *VP model* is a mixture of an infinite set of Von Mises distributions:

$$p(x|s;\bar{J},\tau) = \int p(x|s;J)\, p(J|\bar{J};\tau) dJ$$

$$= \int VM((x; s, \kappa(J)) Gamma(J|\bar{J};\tau)\, dJ$$

In the current study, we employed two versions of the *VP models*[1], which have been customized by Suchow et al. (2013; memtoolbox.org), based on the fundamental theory of van den Berg et al. (2012).

Variable-precision model (VP). In this specific *VP model,* the standard deviation of VWM representation follows a Gaussian distribution:

$$sd = Gaussian(sd|mnSTD; stdSTD),$$

where $mnSTD$ represents the mean of this normal distribution and $stdSTD$ represents the standard deviation of this normal distribution. The response error distribution is then given by

$$p(x|s,g,\kappa) = (1-g)VM\big(s-x; 0, \kappa(sd)\big) + g\frac{1}{2\pi}.$$

[1] In these versions, the guess rate g is included into the variable precision models.

Variable-precision model with Gamma precision (VPG). In the *VPG model*, the precision J of response is assumed to be distributed as a Gamma distribution, as in van den Berg et al. (2012):

$$J = Gamma(J|\bar{J};\tau).$$

And the precision is equal to the inverse variance, $J = \frac{1}{sd^2}$. The response error distribution of the *VPG model* is written by

$$p(x|s,g,\kappa) = (1-g)VM\big(s-x;0,\kappa(sd)\big) + g\frac{1}{2\pi}.$$

Models with interaction between items

Ensemble integration model (EnsInt). Previous models mentioned in this article suggest that each item is stored in the VWM independently, which implies there is no interaction between the representations of these items. However, Brady and Alvarez (2011) demonstrated that the representation in VWM is affected by the item itself, as well as the integration of information of all the stimuli displayed. Their results showed that the reported size of a target item was shifted toward not only the mean size of the items with the same color, but also the mean size of all items displayed. The *EnsInt model* implemented in the current study is a simplified version of the Brady & Alvarez (2011) model, which introduces a parameter that represents the bias towards the mean value of all distractors. Since the first target is the only other relevant item in typical continuous-report AB tasks, in the context of the AB, we used T1's value for this parameter, slightly simplifying the model. Formally, the response error distribution of the *EnsInt model* is given by

$$p(x|s,g,sd,\mu_s) = (1-g)VM\big(s-x;\mu_s,\kappa(sd)\big) + g\frac{1}{2\pi},$$

where μ_s refers to the biased mean value, which is estimated based on the differences between the target (T2) and the other item (T1).

Swap model This model can be considered as an extension of the *StM model*. In addition to the two components of the response error distribution, namely a uniform distribution that indicates the proportion of random-guessing trials and a Von Mises distribution that represents the precision, Bays et al.(2009) introduced a third source of error: The probability of incorrectly reporting non-target items. Consequently, the *Swap model* has a three-component structure: precision, guess rate, and swaps (estimated by the measurement of the distance between reported values and non-target items):

$$p(x|s, g, sd, \mu_s) = (1 - g - \beta)VM\big(s - x; 0, \kappa(sd)\big)$$

$$+\beta\frac{1}{m}\sum_{i}^{m} VM(s - x; s - \theta_i, \kappa(sd)) + g\frac{1}{2\pi}$$

where β denotes the probability of incorrectly reporting the non-target items (distractors; i.e., T1) and $\{\theta_1, \theta_2, \ldots \theta_m\}$ are the values of a set of m distractors.

Datasets

As summarized in Table 1, we obtained three existing datasets from three separate laboratories (Asplund et al., 2014; Karabay et al., 2022; Tang et al., 2020), and added a fourth, new dataset in the present paper (Wang et al.).

Asplund et al. (2014). The first dataset was obtained from Experiment 1 in Asplund et al. (2014), where square-shaped targets were embedded in circle-shaped colored distractors. T1 was filled with either black or white, while T2 color was randomly chosen from 180 equiluminant colors. T2 appeared at the first, second, fourth, or eighth position (labeled as Lag 1, Lag 2, Lag 4, and Lag 8) following T1. The stimulus duration for each participant was set by a staircase method and resulted in an average value of 150 ms (*SD* = 10 ms). At the end of the stimulus stream, participants were asked to reproduce T2 using a color wheel, and to report whether the first target was black or white.

Tang et al. (2020). The second dataset was retrieved from two behavioral experiments in the AB study by Tang et al. (2020), where a RSVP task was again employed. Specifically, there were 20 Gabor patches, and the orientation of the Gabors were integer angles ranging from 0 to 179 degrees, without replacement. Participants were instructed to memorize two targets (high-frequency gratings) among distractors (low-frequency gratings) and reproduce them at the end of each trial. Each target or distractor appeared for 40 ms, and the next item followed after a blank interval of 80 ms. The time interval between two targets was manipulated by setting different numbers of inter-items (lag). In Experiment 1, there were five lag conditions (1, 2, 3, 5, 7), while in Experiment 2, there were two lag conditions (3, 7).

Karabay et al. (2022). Karabay et al. (2022) recently used two different paradigms, dwell time (DT) and RSVP, both with continuous report tasks, to examine the nature of awareness in the AB. We acquired seven sub-datasets from their experiments. Experiment 1A, 1B, 2A, and 2B were run with the DT paradigm. Both targets were orientation gratings (in

Experiment 1A and 1B) or colors (in Experiment 2A and 2B). T1 and a distractor were presented, at the same time, above or below of the fixation dot, at an equal distance. The display of T2 was the same as T1, except the location changed to the left or right side of the fixation dot. On each trial, T1 and T2 target arrays were followed by a mask. The stimulus-onset asynchrony (SOA) between T1 and T2 was set to 250 ms and 800 ms. The RSVP paradigm was used in Experiment 3, 4, and 5, in which the targets were always orientation gratings. The number of distractor items between two targets were 2 (Lag 3) or 7 (Lag 8), and the SOA was set to 300 ms or 800 ms. In Experiment 3, the two targets appeared among a stream of distractors, similar to the design in Tang et al. (2020). In Experiment 4, there were simultaneously two streams of stimuli at the left and right sides of the fixation dot. In Experiment 5, T1 was presented within a single stream of distractors, but split into a dual stream with the presentation of T2, as in Experiment 4.

Table 1

Details of the data sets.

Datasets	Exp.	SOA (ms)	Lag	T1 feature	T2 feature	Original subject number	Subject number for the analysis
Asplund et al., 2014	Exp. 1	NA	1, 2, 4, 8	Categorical	Color (1°-180˚)	27	26
Tang et al., 2012	Exp. 1	50, 240, 360, 600, 720	1, 2, 3, 5, 7	Orient. (1°-180˚)	Orient. (1°-180˚)	22	17
	Exp. 2	360, 720	3, 7	Orient. (1°-180˚)	Orient. (1°-180˚)	22	21
Karabay et al., 2022	Exp. 1A	250, 800	NA	Orient. (1°-180˚)	Orient. (1°-180˚)	23	22
	Exp. 1B	250, 800	NA	Orient. (1°-180˚)	Orient. (1°-180˚)	21	19
	Exp. 2A	250, 800	NA	Color (1°-360˚)	Color (1°-360˚)	26	25
	Exp. 2B	250, 800	NA	Color (1°-360˚)	Color (1°-360˚)	27	25
	Exp. 3	300, 800	3, 8	Orient. (1°-180˚)	Orient. (1°-180˚)	24	22
	Exp. 4	300, 800	3, 8	Orient. (1°-180˚)	Orient. (1°-180˚)	24	23
	Exp. 5	300, 800	3, 8	Orient. (1°-180˚)	Orient. (1°-180˚)	23	22
Wang et al. data	Exp. 1	90, 300, 800	1, 3, 8	Orient. (1°-180˚) or Color (1°-360˚)	Orient. (1°-180˚) or Color (1°-360˚)	28	19

Note. Exp. = Experiment; Orient. = Orientation; Wang et al. data refers to the new data set collected in the present study.

Wang et al. data

In the new experiment, we manipulated feature changes between targets in the AB task (same or different), hypothesizing that matching target features might cause mutual bias in target reports, which might affect the model selection. Like the other studies, this experiment also featured continuous report of targets.

Participants

The minimum sample size of 24 subjects was determined by means of a G-power analysis (Faul et al., 2007), with the following parameters: significance level (α = .05), beta-level (β = .2), and a medium to large effect size (Cohen's d_z = .6). Twenty-eight undergraduate students (seventeen females and eleven males) from the University of Groningen were recruited in exchange for course credits (mean age = 19.6, range = 18-27). All participants reported normal or corrected-to-normal visual acuity and no color blindness. Informed consent forms and instructions were given to the participants prior to participation. Nine participants were excluded from further analysis, based on our exclusion criteria (see Subject Exclusion section). Prior to its execution, ethical approval was obtained from the ethical committee of the Psychology Department of the University of Groningen (Approval code PSY-1819-S-0209). The study was conducted in accordance with the Declaration of Helsinki (2008).

Apparatus and stimuli

Participants were individually seated on a desk chair in sound-attenuated lab chambers, at a viewing distance of approximately 60cm from a 22" CRT monitor (Iiyama MA203DT). The screen was set at a 1280 x 1024-pixel resolution, 16-bit color depth, and a 100Hz refresh rate. The trials were prepared and run in OpenSesame 3.2 (Mathôt et al., 2012) with the PsychoPy back-end (Peirce, 2007, 2009), under the Microsoft Windows 7 operating system. All the stimuli were displayed in the center of the screen on a grey background (RGB: 128,128,128), as shown in Figure 1A. There were two kinds of target stimuli in the RSVP

stream (Figure 1B). The first kind of target consisted of orientation gratings with a spatial frequency of 1.8 cycles/degree of visual angle, presented within a circle of 2.2° of visual angle. The orientation of the gratings was chosen randomly from the range of 0-179°. The other kind of target consisted of circles (2.2° of visual angle) filled with one of 360 colors which were chosen randomly from the HSL color spectrum. A color wheel with these chosen 360 colors was used as the report probe. The distractors were grey-scale mosaic images (2.2° × 2.2° of visual angle), retrieved from Karabay et al. (2022).

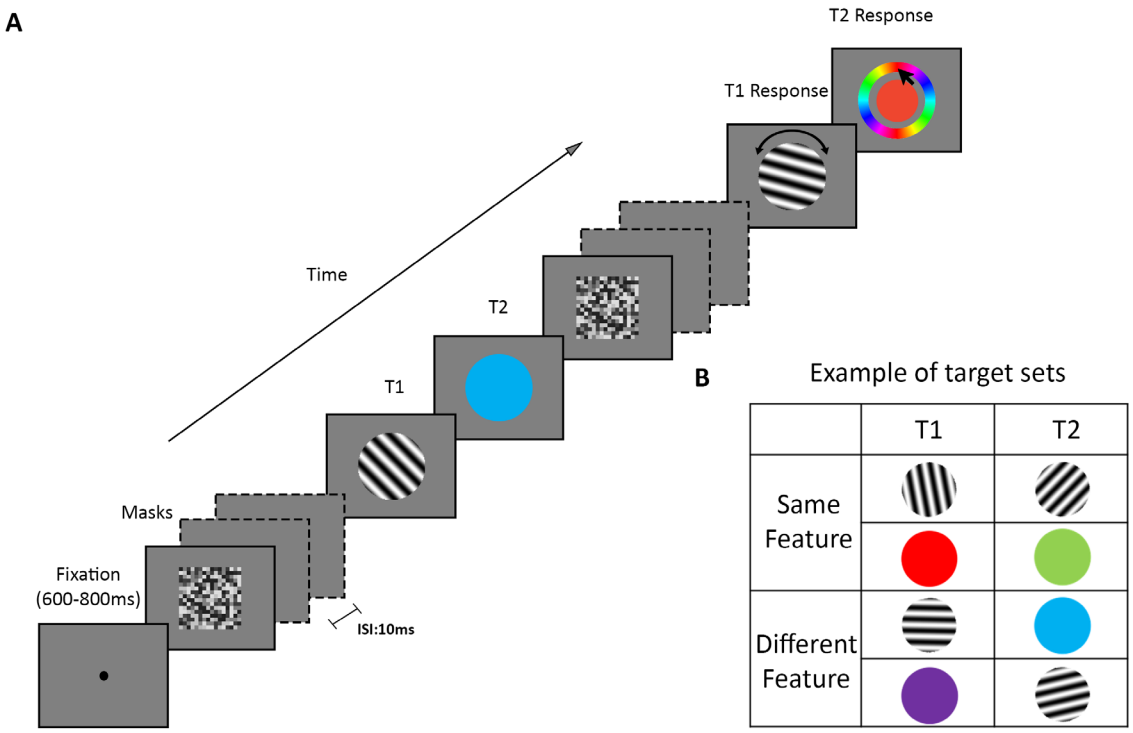
Procedure

A 2 (Target feature: same or different) x 3 (Lag: 1, 3, 8) factorial design was implemented in this experiment. Two practice blocks, each including 16 trials, were administered at the beginning. After that, there were 540 experimental trials evenly distributed into 15 blocks. Practice trials were excluded from the analysis. Each trial started with a fixation dot that randomly lasted for 600-800ms. There were 18 items within the RSVP stream. Each item was shown for 80 ms, separated by a 10 ms inter-stimulus interval (ISI). T1 was randomly presented at the fifth, sixth, or seventh position of the stream. T2 was presented consecutively (Lag 1), after two items (Lag 3), or after seven items (Lag 8), following T1. T1 and T2 could have either the same feature (color or orientation), or different features. At the end of the stream, after a blank interval of 500ms, two response prompts were displayed successively.

Participants were asked to reproduce the two targets in the correct order. They were asked to reproduce the target orientation by rotating a probe grating, or to reproduce the target color by choosing the exact color from a color wheel. Responses were collected with a standard mouse. Trial-based feedback was provided only in practice trials: A happy smiley was shown for correct trials, and an unhappy smiley for incorrect trials. Responses close to target values (below 20° reproduction error) were considered correct responses for feedback. In the experimental blocks, the average performance on T1 and T2 were displayed at the end of each

block. During the experiment, participants could have a self-paced break between blocks. The whole session took about 90 minutes to complete.

Fig. 1 Illustration of the design of the newly conducted experiment. **a.** An example trial of the dual-target RSVP task at Lag 3. In this instance, the two targets have different features. **b.** Examples of possible targets in all combinations and orders of the two targets.



Data Preparation

As indicated, all the data sets we gathered for analysis were obtained from continuous-report AB tasks. We reorganized the datasets to form a uniform structure, as per the following procedure:

1. For each trial, response error for each target (T1 and T2) was calculated by subtracting the response value from the actual one.
2. Trials in which the absolute deviations between T1 and T2 values were less than 22.5 degrees were discarded.
3. Response errors were limited to a range from -90 to 90 for orientations and from -180 to 180 for colors.
4. T2 performance was analyzed under the condition that T1 was correct (T2|T1), as is commonly done in AB studies, which means we only included the trials when the absolute error of T1 was less than 22.5 degrees.

Model fitting and parameter estimates

To fit all of the models mentioned above to the data sets, we used maximum likelihood estimation (MLE) to find the best estimation of parameters of each model. The general logic of MLE is to maximize the likelihood function to find out the specific parameter value, with which the chosen probabilistic model is most "likely" to generate the observed data. Here, for each subject of each condition within a single experiment, we carried out model fitting with the built-in MLE function in MemToolbox (Suchow et al., 2013; memtoolbox.org). After the parameter value was calculated at an individual level, the medians across all subjects within each condition were listed as the most probable estimate of the specific model. Traditionally, the mean and standard deviation have been used to represent the estimate of the parameters. However, for some subjects' data, there may be some outliers of the parameters, which would cause a high bias of the means. To gain a more robust insight into the parameter estimates and to compare

these parameters between different conditions and models, the current study thus used the medians of parameters instead of means, while keeping the standard deviation to represent the dispersion of the parameters.

Model comparison

Considering the comparison and selection of models, flexible models can fit a wide range of datasets by introducing additional parameters. However, this kind of fitness may lack evidence and also induce overfitting (Hastie et al., 2009). We used the Bayesian Information Criterion (BIC; Schwarz, 1978), a common penalized model comparison measurement, to evaluate the fitness of these models to the data, and to assess which model best describes the data. BIC values were calculated based on the natural logarithm of the likelihood function, often called the log-likelihood:

$$BIC = k * \log(n) - 2LL,$$

where k is the number of parameters, LL is the log-likelihood, and n represents the number of the observed data points. At the individual level, the BIC value was calculated using the model comparison function implemented in MemToolbox (Suchow et al., 2013; memtoolbox.org), with lower BIC scores indicating better model fit. Subtracting the lowest BIC value (the most likely model of each subject) from each model's BIC, we obtained the relative BIC of each model, at the subject level.

Furthermore, to investigate the model performance against group data, we generalized the comparison results by ranking the models for each condition. First, we ranked the models for each participant based on the relative BIC and measured the ranks. Second, we averaged the ranks across all subjects for each condition.

Simulating data from model

Although the BIC indicates how well a model can fit the experimental data compared to other models, it does not detail the absolute performance of each model. To examine which

part of the data each model can fit well and which part less so, we simulated data sets from these models and visualized the comparison between model predictions and the actual T2 errors. To do so, we firstly applied the most probable parameters of each subject and each lag condition, which were obtained from the model fitting mentioned before, to the corresponding model. Then we simulated the same number of data points as we had collected from each subject in the same condition. Finally, we calculated the summary statistics of T2 errors separately for the data generated from the models and the actual data gathered from different studies. This simulation procedure was carried out with a custom-made function in Matlab, based on the SampleFromModel function in Memtoolbox.

Subject Exclusion

The actual subject numbers in the analysis of each study are listed in Table 1. The subject exclusion based on the Z-scores was carried out to allow more reliable results at the summarized group level. For each study, we Z-scored the differences of BIC values between the best model and the worst one for each subject, in each experimental condition (lag or target pair), and subsequently individual BIC values of each model at each condition were calculated. Finally, we removed the subjects whose absolute Z-scores were larger than three.

Results

Asplund et al. (2014)

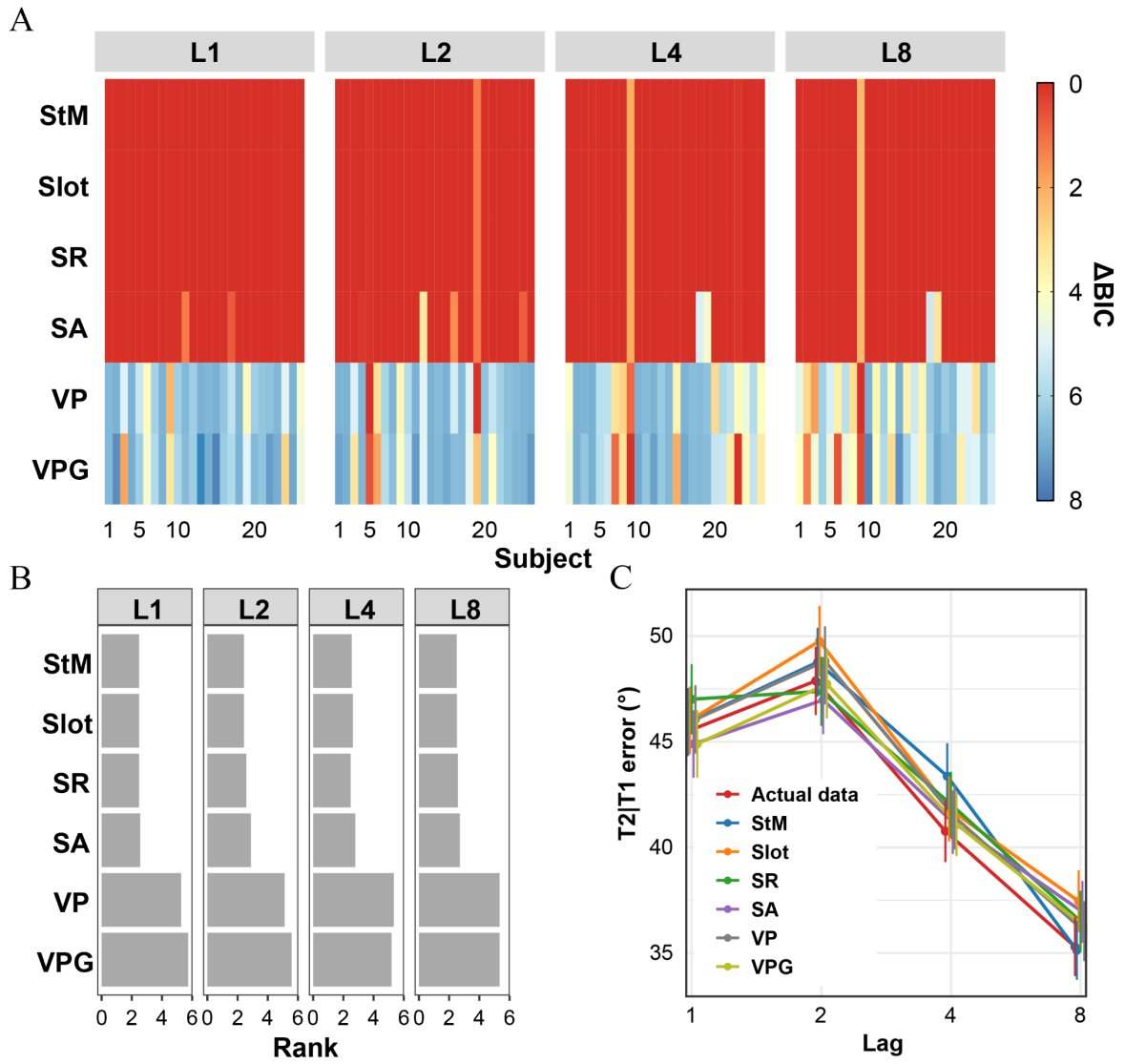
Model Fit and Model Comparison

Figures 2A and 2B present the model comparison analysis for the Asplund et al. (2014) data. Since Asplund et al.'s dataset consisted of entirely different T1 and T2 stimuli, the *EnsInt model* and the *Swap model* were excluded as they both need a third parameter for the similarity of two targets that was absent here.

At the individual level, the *StM*, *Slot*, and *SR models* showed the best fit for most subjects' data in all four lags, while the *SA model* showed a slightly higher BIC value for several subjects (Figure 2A). In contrast, the *VP model* and its variant, the *VPG model*, fitted worse compared to the *StM model* and its variants for most subjects.

To further assess model performance at the group level, we summarized the individual model fitness values. Figure 2B shows the average model ranking for each lag (also see the exact ranks in Table 2). The pattern of the model ranking was similar to the model performance at the subject level. The *StM*, *Slot*, and *SR models* dominated the top ranks at all lags. The *SA model* had slightly larger ranks at Lag 2, 4 and 8. This was consistent with its performance at the subject level, having a higher BIC for a few subjects. Moreover, model performance was consistent across different lags indicating that the AB itself (in particular at Lag 2) did not change the shape of the response error distributions, compared to other lags. Moreover, for each lag condition, we computed the average BIC of each model across all subjects and then calculated the difference from the best model at that lag (Table 2). We found that the BIC of the *VP model* exceeded the best model by 5.95, 5.55, 5.28, and 4.71, respectively, at Lag 1, Lag 2, Lag 4, and Lag 8. The same level of BIC differences from the best model was also found in the *VPG model*.

Fig. 2 Model comparison results for the Asplund et al. data. **A.** Model comparison at the subject level. Each column represents a subject, and each row represents a specific model. Each panel represents each lag condition. L1 = Lag 1; L2 = Lag 2; L4 = Lag 4; L8 = Lag 8. Cell color indicates each model's relative BIC value. Larger relative BIC value means worse performance. **B.** Average model ranking for each lag. Per subject, we ranked these models by their BIC values, after which the average ranks (the bar length shown here) were obtained by averaging each model's ranks within each lag across all subjects. Larger rank means worse performance. **C.** Comparison of the mean T2 error in the actual data sets and in the data simulated from the six selected models.



Model Simulation

Table 3 lists the estimated parameters obtained from MLE, based on which we synthesized the mean T2|T1 error for six models. Figure 2C presents the comparison between the synthetic mean T2 error and the actual mean T2 error, both as a function of lag. Of particular interest was the response error of the second target when the time interval between the first and second target fell between 200 – 500 ms (mostly at Lag 2 here), to get an idea if these models can actually describe the data well during the AB. In general, the pattern of the simulated data was more or less similar to that of the actual data. One surprising point was that the *SR model* estimated a relatively high error at Lag 1, which caused it to fail in predicting the actual increase of response error from Lag 1 to Lag 2.

Table 2

Model comparison results

Lag	Model	BIC	Difference from the best model	Average model rank	Log-likelihood
1	StM	1591.24	0.00	2.48	-788.80
	Slot	1591.24	0.00	2.48	-788.80
	SR	1591.24	0.00	2.48	-788.80
	SA	1591.31	0.08	2.56	-788.84
	VP	1597.18	5.95	5.27	-788.36
	VPG	1597.30	6.06	5.73	-788.42
2	StM	1631.22	0.00	2.42	-808.77
	Slot	1631.22	0.00	2.42	-808.77
	SR	1631.22	0.00	2.58	-808.77
	SA	1631.44	0.22	2.88	-808.89
	VP	1636.77	5.55	5.12	-808.13
	VPG	1636.79	5.57	5.58	-808.14
4	StM	1583.14	0.00	2.56	-784.74
	Slot	1583.14	0.00	2.63	-784.74
	SR	1583.14	0.00	2.48	-784.74
	SA	1583.50	0.36	2.79	-784.92
	VP	1588.42	5.28	5.35	-783.96
	VPG	1588.06	4.92	5.19	-783.78
8	StM	1521.87	0.00	2.50	-754.11
	Slot	1521.87	0.00	2.52	-754.11
	SR	1521.87	0.00	2.58	-754.11
	SA	1522.20	0.33	2.71	-754.28
	VP	1526.58	4.71	5.35	-753.06
	VPG	1526.58	4.71	5.35	-753.05

Table 3

Model parameters

		Lag 1		Lag 2		Lag 4		Lag 8	
Model	**Parameter**	**Median**	***SD***	**Median**	***SD***	**Median**	***SD***	**Median**	***SD***
StM	*g*	0.39	0.09	0.45	0.09	0.34	0.14	0.26	0.14
	sd	21.29	6.15	21.29	4.36	20.86	4.60	19.91	4.04
Slot	*capacity K*	1.21	0.18	1.11	0.18	1.33	0.27	1.47	0.28
	sd	21.29	6.15	21.29	4.36	20.86	4.60	19.91	4.04
SR	*capacity K*	1.21	0.18	1.11	0.18	1.33	0.27	1.47	0.28
	best*SD*	19.42	6.23	19.94	3.36	19.11	3.81	17.37	3.60
SA	*capacity K*	1.21	0.16	1.11	0.16	1.33	0.24	1.47	0.22
	sd	21.29	6.22	21.29	4.18	21.53	4.42	20.19	5.30
VP	*g*	0.38	0.09	0.43	0.11	0.31	0.14	0.23	0.14
	mnSTD	21.88	8.10	22.62	5.34	23.11	8.37	21.21	4.45
	stdSTD	5.15	5.81	7.69	7.35	7.07	10.48	7.34	4.77
VPG	*g*	0.34	0.11	0.39	0.12	0.27	0.15	0.20	0.15
	*mode*Precision	2.55×10^{-3}	8.8×10^{-3}	2.43×10^{-3}	1.4×10^{-3}	2.49×10^{-3}	1.42×10^{-3}	3.21×10^{-3}	1.74×10^{-3}
	*sd*Precision	1.19×10^{-3}	3.92×10^{-3}	1.88×10^{-3}	5.84×10^{-3}	3.02×10^{-3}	4.54×10^{-3}	4.19×10^{-3}	4.77×10^{-3}

Model Fit and Model Comparison

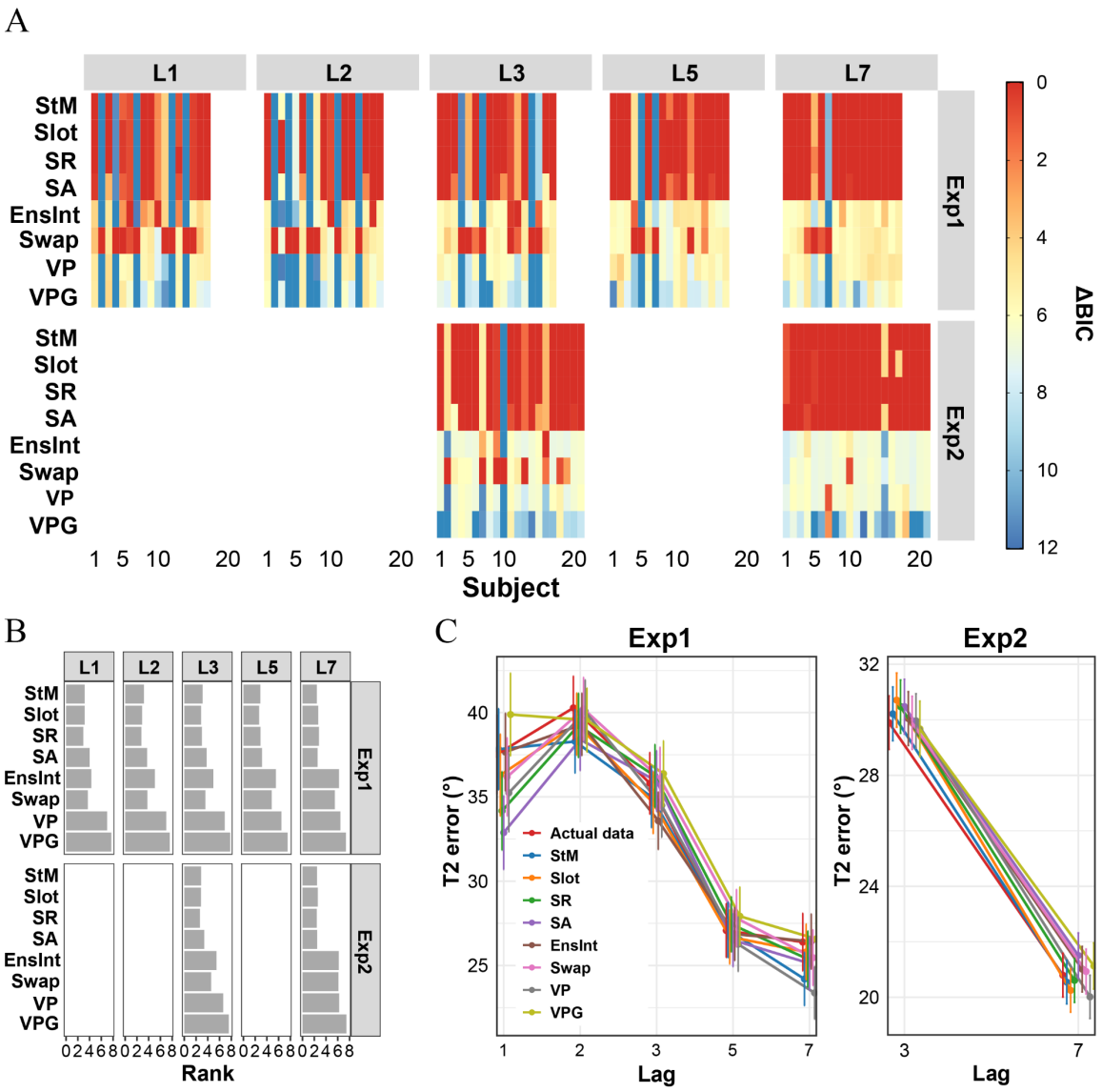
At the individual level, the *StM model* and the slot models (the *Slot*, *SR*, and *SA models*) performed similarly (Figure 3A). For some individuals' data from Experiment 1, the *StM model* and the slot models always showed a better fitness than the other models at all lags. However, it can clearly be seen that the *Swap model* strongly outperformed the other models for several participants at all lags, while it showed a moderate goodness-of-fit for other participants. It is also worth noting that the number of participants for which the *Swap model* was the winning one dropped as a function of lag. At Lag 1, 2, and 3, the *EnsInt model* was the best model for a subset of subjects. The extent of preference to the *EnsInt model* decreased at Lag 5 and 7, where the *EnsInt model* showed a relatively consistent and moderate performance across all subjects. The *VP* and *VPG models* were always the worst fitting ones for all subjects in all lags. In Experiment 2, there were two lag conditions (Lag 3 and Lag 7). The results presented a similar pattern to Experiment 1 with regard to lag. It was noteworthy that the *Swap model* was again the winning model for a subset of subjects at Lag 3, but only for one subject at Lag 7.

The average model ranking shown in Figure 3B illustrates the model comparison results at the group level, and the exact ranks are listed in Table 4. For Experiment 1, at Lag 1, 2 and 3, the *StM*, *Slot*, and *SA models* dominated the top ranks, while the *Swap*, *SA*, and *EnsInt models* were the second-best models. At Lag 5 and 7, there was a clear divergence between the winning models (the *StM*, *Slot*, *SA* and *SR models*) and the losing ones (the *EnsInt*, *Swap*, *VP* and *VPG models*). What stood out in this comparison was that for the *EnsInt model* and the *Swap model* in both Experiment 1 and Experiment 2, the rank differences between lags were quite pronounced. In contrast, other models had similar performance at different lags.

The mean BIC value of each model (Table 4) at each lag showed that the performance of the *Swap model* changed with lags. For Experiment 1, the *Swap model* had the lowest BIC

value at Lag 1, 2, and 3, and it differed by 1.29 and 3.29 points from the best model at Lag 5 and 7. This finding is coherent with existing AB theories such as STST (Bowman & Wyble, 2007; Wyble et al., 2009), and with the notion of temporal integration (Akyürek et al., 2012; Hommel & Akyürek, 2005). These suggest that targets are bound to same episodic event if they are successively shown, possibly when the second target is processed (tokenized) more quickly than the first one. Then, timestamps of the targets might be confused, which cause T2 to be reported as T1 and vice versa. This account also predicts that the amount of swap errors should decrease strongly as a function of lag. For Experiment 2, the *Swap model* had a slight difference of mean BIC from the best model (2.72) at Lag 3, but a more substantial difference (6.34) at Lag 7. Furthermore, the *StM*, *Slot*, *SR*, and *SA models* always had similar mean BIC, which was consistent with the pattern in Figure 3.

Fig. 3 Model comparison results for the Tang et al. data. **A.** Model comparison at the subject level. L1 = Lag 1; L2 = Lag 2; L3 = Lag 3; L5 = Lag 5; L7 = Lag7. **B.** Average model ranking for each lag and each experiment. **C.** Model simulations of T2 error as a function of lag.



Model Simulation

Figure 3C displays the summary statistics for synthetic T2 response errors obtained from model simulation (Table 5 provides the estimated parameters used in the model simulation) and actual T2 response errors obtained from the Tang et al. data (2021). In Experiment 1, the *VPG model* had the highest prediction for T2 response error at Lag 1, over the actual data and other models. The T2 response errors from the *VP*, *SR* and *SA models* were lower than the experimental data. At Lag 2, 3, and 5, the divergences among these models and the actual data were nevertheless pretty narrow. At the longest time interval (Lag 7), the *VP model* and the *StM*

model generated lower T2 response errors. However, for Experiment 2, all models accurately simulated the data.

Table 4

Model comparison results

Experiment	Lag	Model	BIC	Difference from the best model	Average model rank	Log-likelihood
Exp1	**1**	StM	332.16	4.54	3.18	-160.97
		Slot	332.11	4.48	3.18	-160.95
		SR	332.06	4.44	2.94	-160.93
		SA	332.51	4.89	4.00	-161.15
		EnsInt	332.43	4.81	4.29	-158.56
		Swap	327.62	0.00	3.71	-156.15
		VP	337.28	9.65	7.00	-160.98
		VPG	338.00	10.37	7.71	-161.34
	2	StM	530.76	4.91	3.21	-259.81
		Slot	530.41	4.56	2.88	-259.64
		SR	530.41	4.56	2.76	-259.64
		SA	531.10	5.25	3.74	-259.98
		EnsInt	531.92	6.07	5.06	-257.61
		Swap	525.85	0.00	3.76	-254.57
		VP	536.32	10.47	7.00	-259.81
		VPG	536.86	11.01	7.59	-260.08
	3	StM	534.71	3.23	3.12	-261.76
		Slot	534.53	3.05	2.85	-261.67
		SR	534.53	3.05	3.00	-261.67
		SA	535.32	3.83	3.79	-262.06
		EnsInt	536.34	4.85	4.94	-259.77
		Swap	531.48	0.00	3.59	-257.34
		VP	540.31	8.82	6.88	-261.75
		VPG	541.99	10.51	7.82	-262.60
	5	StM	558.18	0.10	2.88	-273.41
		Slot	558.08	0.00	2.68	-273.36
		SR	558.08	0.00	2.91	-273.36
		SA	558.32	0.24	3.18	-273.49
		EnsInt	561.01	2.93	5.53	-271.99
		Swap	558.87	0.79	4.82	-270.92
		VP	563.60	5.52	6.47	-273.29
		VPG	564.70	6.62	7.53	-273.83
	7	StM	515.42	0.08	2.44	-252.10
		Slot	515.42	0.08	2.68	-252.10
		SR	515.34	0.00	2.79	-252.06
		SA	515.43	0.09	2.56	-252.11
		EnsInt	520.72	5.38	6.21	-251.95
		Swap	518.94	3.60	5.50	-251.06
		VP	520.97	5.63	6.44	-252.07
		VPG	522.08	6.74	7.38	-252.63

Experiment	Lag	Model	BIC	Difference from the best model	Average model rank	Log-likelihood
Exp2	**3**	StM	1378.49	0.00	2.86	-682.67
		Slot	1378.49	0.00	2.83	-682.67
		SR	1378.49	0.00	2.67	-682.67
		SA	1378.97	0.48	3.40	-682.91
		EnsInt	1383.55	5.06	5.48	-681.91
		Swap	1381.21	2.72	4.57	-680.74
		VP	1384.84	6.35	6.62	-682.55
		VPG	1386.82	8.33	7.57	-683.54
	7	StM	1321.24	0.19	2.48	-654.03
		Slot	1321.46	0.41	2.60	-654.14
		SR	1321.05	0.00	2.43	-653.93
		SA	1321.23	0.18	2.50	-654.02
		EnsInt	1327.73	6.68	6.17	-653.97
		Swap	1327.39	6.34	6.12	-653.80
		VP	1327.33	6.28	6.24	-653.77
		VPG	1329.83	8.78	7.48	-655.02

Table 5

Model parameters

		Exp1										Exp2			
		Lag 1		Lag 2		Lag 3		Lag 5		Lag 7		Lag 3		Lag 7	
Model	**Parameter**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
StM	*g*	0.58	0.40	0.56	0.45	0.34	0.43	0.26	0.35	0.12	0.37	0.31	0.33	0.16	0.29
	sd	28.88	18.22	24.46	19.26	23.57	18.87	19.51	15.53	22.77	10.12	21.8	17.04	19.81	9.81
Slot	*capacity K*	0.72	1.58	0.88	1.44	1.33	1.61	1.49	0.94	1.76	1.23	1.38	1.51	1.67	1.47
	sd	28.88	18.09	18.63	22.04	28.18	31.26	19.51	29.69	26.24	26.81	22.01	17.26	19.96	10.17
SR	*capacity K*	1.26	1.09	1.28	1.21	1.49	1.08	1.49	1.34	1.76	0.69	1.38	0.91	1.67	0.99
	bestSD	25.00	56.47	35.34	70.69	23.74	37.22	20.06	46.69	20.81	46.00	25.21	41.09	14.96	42.42
SA	*capacity K*	1.60	1.02	2.00	1.55	2.00	1.28	1.49	1.18	1.78	0.39	1.39	0.86	1.74	0.59
	sd	53.73	91.54	58.63	93.62	52.24	108.96	29.71	69.08	27.85	82.17	31.76	48.07	19.96	68.84
EnsInt	*g*	0.29	0.36	2.56×10^{-3}	0.42	0.34	0.38	0.20	0.31	0.14	0.23	0.31	0.3	0.16	0.25
	sd	30.22	19.14	33.77	18.46	28.18	14.89	21.84	14.76	26.24	20.22	24.8	14.91	19.81	13.94
	samples	5.54×10^{-4}	5.23×10^{12}	7.09×10^{-7}	1.39×10^{12}	0.27	6.41×10^{12}	2.47	6.40×10^{12}	419.52	3.90×10^{12}	56.09	6.21×10^{12}	1.02×10^{13}	7.27×10^{12}
Swap	*g*	0.09	0.27	0.07	0.30	2.30×10^{-13}	0.36	0.13	0.26	0.06	0.20	0.08	0.25	0.13	0.23
	B	0.40	0.38	0.33	0.35	0.08	0.36	0.02	0.35	5.36	0.33	0.04	0.25	4.38×10^{-15}	0.22
	sd	24.85	14.20	21.98	22.98	28.18	15.39	23.13	9.28	25.04	11.41	25.41	16.24	19.81	10.85
VP	*g*	0.64	0.35	0.69	0.40	0.67	0.39	0.41	0.36	0.10	0.34	0.35	0.29	0.16	0.27
	mnSTD	39.92	15.50	46.65	16.62	38.95	15.59	24.51	13.22	31.46	13.51	25.42	15.68	21.14	11.23
	stdSTD	0.01	3.48	0.03	3.66	0.01	5.23	0.03	9.88	2.00	8.04	0.02	5.28	2.37	3.88
VPG	*g*	0.92	0.22	0.88	0.17	0.84	0.27	0.60	0.36	0.49	0.33	0.68	0.30	0.26	0.33
	mode-Precision	5.10×10^{-4}	0.79	5.02×10^{-4}	0.47	5.15×10^{-4}	0.03	5.03×10^{-4}	1.32×10^{-4}	5.00×10^{-4}	5.23×10^{-4}	5.00×10^{-4}	7.13×10^{-4}	5.00×10^{-4}	1.66×10^{-4}
	*sd*Precision	5.00×10^{-4}	12.13	5.97×10^{-4}	12.12	5.72×10^{-4}	12.13	5.71×10^{-4}	5.70×10^{-4}	6.55×10^{-4}	3.79×10^{-3}	5.00×10^{-4}	001	5.00×10^{-4}	2.13×10^{-4}

Model Fit and Model Comparison

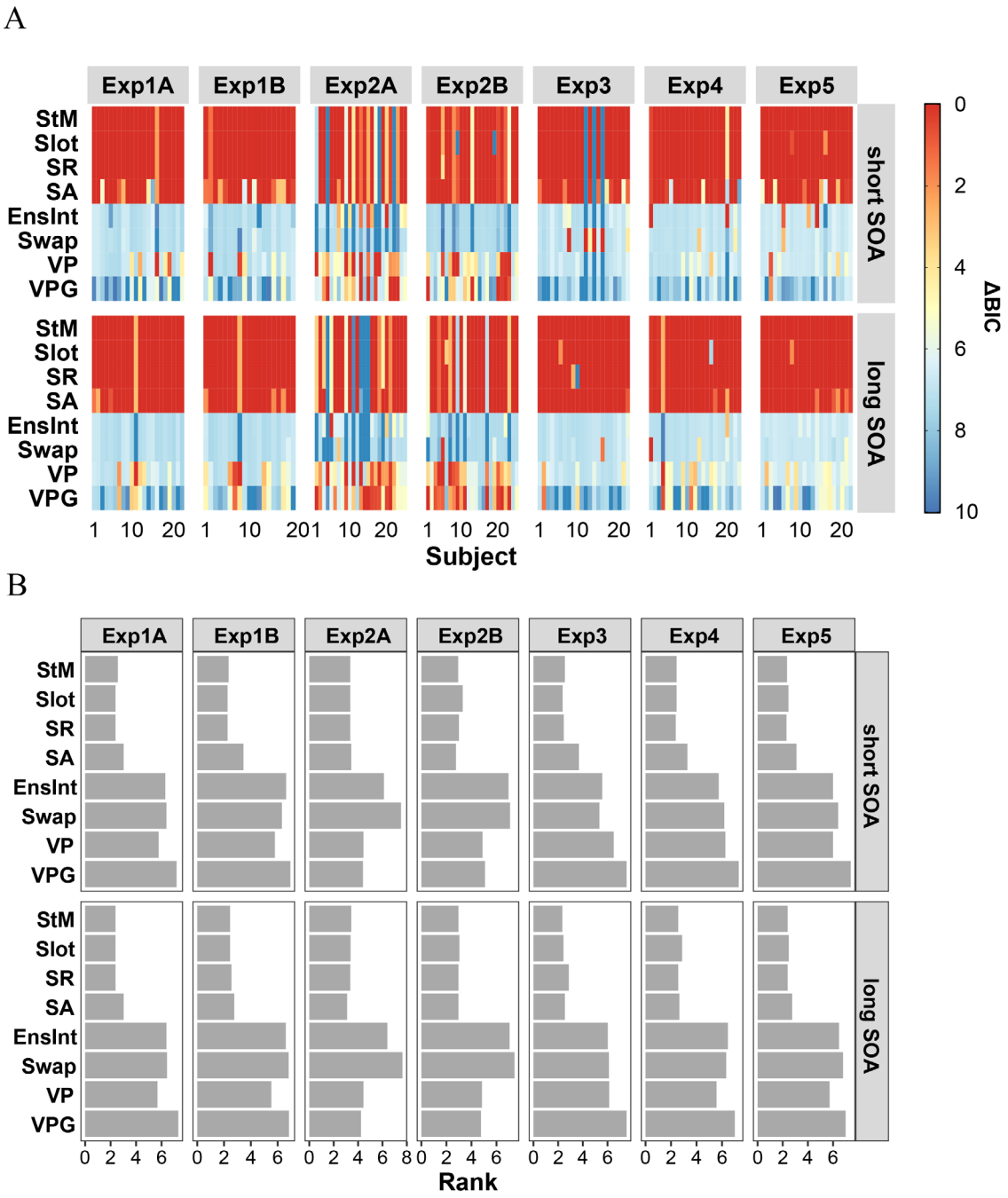
Figure 4A presents the model comparison results at the subject level for the Karabay et al. (2022) data. Comparing these results between different experiments in their study, the most apparent observation was that the models' goodness-of-fit in Experiment 2 was clearly different from those in other experiments. In Experiment 2A and 2B, the targets having only a color feature, the *VP* and *VPG models* showed a good fit for most subjects, at two different SOA's . Especially for some subjects, these two models outperformed all of their competitors. The *EnsInt model* and the *Swap model* produced the worst data fit. However, here the *StM*, *Slot*, *SR* and *SA models* showed two extremes of fitness, either being the best or the worst.

If we turn to the results of other experiments, they shared a similar pattern of model fitness, in which these models were divided into two sides, the *StM*, *Slot*, *SR* and *SA models* being the better ones and the *EnsInt*, *Swap*, *VP* and *VPG models* being the worse. Some further observations can nevertheless be made. First, at the subject level in Experiment 1A and 1B, the *VP model* showed either the best or the worst performance for several subjects, which was rarely the case in Experiment 3 (where higher BIC values were obtained for almost all subjects). At the subject level in Experiment 4 and 5, the *VP model* had moderate fitness for most subjects at long SOA. For most subjects, the BIC value of the *VP model* was lower than the corresponding value of the *EnsInt model* and the *Swap model*. Whereas at short SOA, the *VP model* performed similar to the *EnsInt model* and the *Swap model*. Second, good fitness of the *Swap model* for a few subjects was observed in Experiment 3, at short SOA. Third, for the *SA model*, there was a clear trend of improving fitness scores for some subjects from short SOA to long SOA (except in Experiment 2).

In line with the results at the subject level, the *VP* and *VPG* models had a better fitness at the group level in Experiment 2 than in other experiments (Figure 4B). Moreover, the average

model ranks of the *StM*, *Slot*, *SR* and *SA* models in Experiment 2 were larger than their ranks in other experiments (also see the exact ranks in Table 6). For all the experiments except Experiment 2, the rankings of the *SA model* at long SOA were higher than their counterpart at short SOA, which was consistent with the performance of the *SA model* at the subject level. The mean BIC value of the *SA model* (Table 6) demonstrated the same performance pattern, having a better fit to the data at long SOA. The BIC value differences from the best model were 0.99, 1.29, 1.54, 1.07 and 1.08 at short SOA, respectively for Experiment 1A, 1B, 3, 4 and 5, while at long SOA the corresponding values were 0.22, 0.31, 0.04, 0.15 and 0.27. The *Swap model* obtained its best rank at short SOA in Experiment 3, with a relatively small BIC difference value of 2.09.

Fig. 4 Model comparison results for the Karabay et al. (2022) data. **A.** Model comparison at the subject level. **B.** Average model ranking for each lag and each experiment.



Model Simulation

The results of MLE using the Karabay et al. data are listed in Table 7. We sampled response errors of all eight models with corresponding estimated parameters. Generally, the divergences among these model simulations and the actual data were narrow (Figure 5). Compared to other models, the *VP* and *VPG models* generated somewhat higher T2 errors at

short SOA in Experiment 1A, 1B and 3, whereas in Experiment 2A and 2B, these two models generated simulations that were most similar to the actual data.

Fig. 5 Model simulations for the Karabay et al data. Each panel presents the comparison of the mean T2 error in actual data sets and in data simulated from eight models, separately for each experiment.

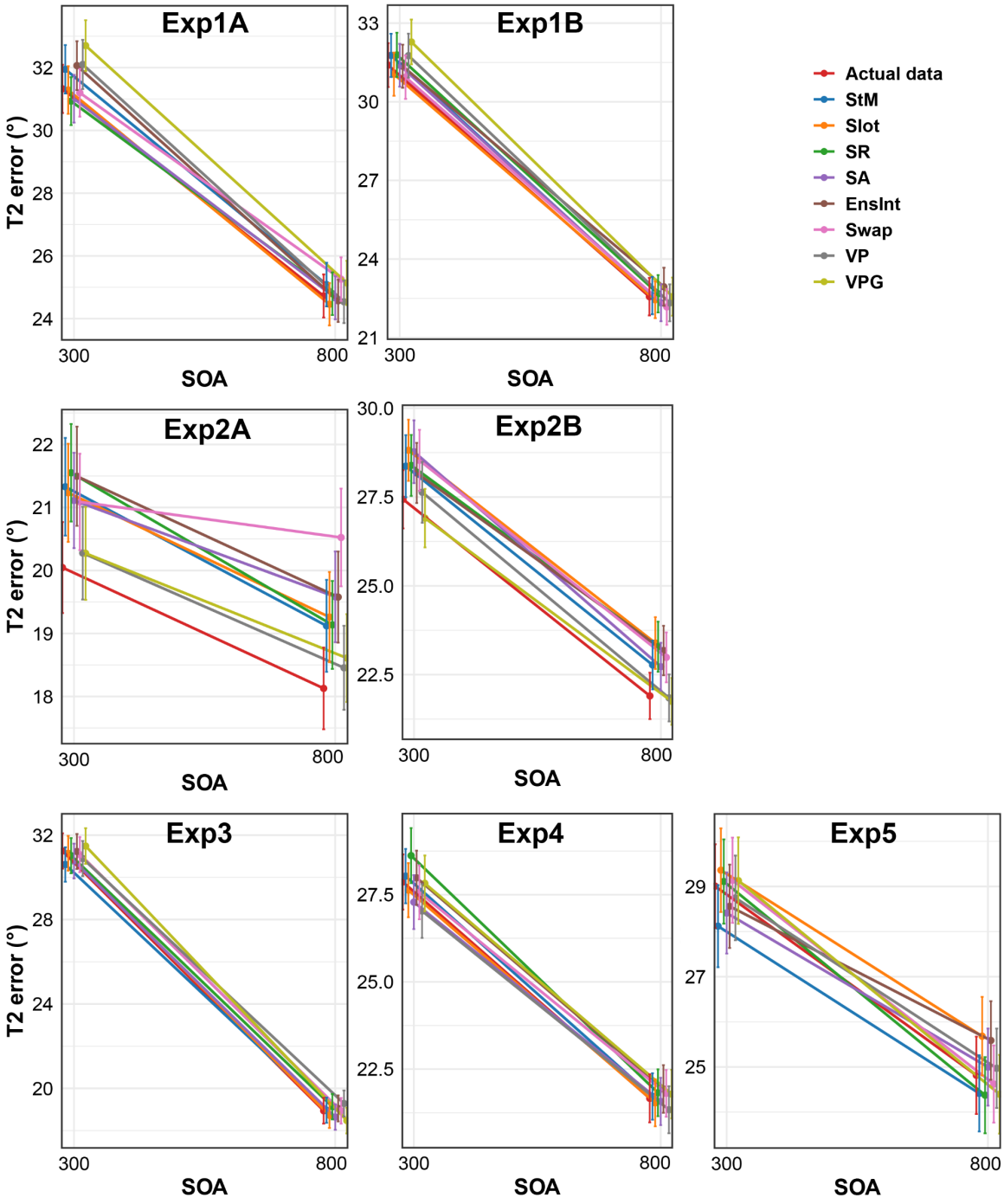


Table 6

Model Comparison Results

		Short SOA				Long SOA			
Experiment	**Model**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**
Exp1A	StM	2175.49	0.00	2.57	-1080.71	2269.85	0.00	2.39	-1127.80
	Slot	2175.49	0.00	2.39	-1080.71	2269.85	0.00	2.39	-1127.80
	SR	2175.49	0.00	2.39	-1080.71	2269.85	0.00	2.39	-1127.80
	SA	2176.48	0.99	3.02	-1081.21	2270.07	0.22	3.02	-1127.91
	EnsInt	2182.62	7.13	6.30	-1080.76	2276.98	7.13	6.39	-1127.80
	Swap	2182.53	7.05	6.39	-1080.71	2276.98	7.13	6.43	-1127.80
	VP	2181.16	5.67	5.77	-1080.03	2275.57	5.72	5.68	-1127.10
	VPG	2184.18	8.69	7.18	-1081.54	2277.58	7.73	7.32	-1128.10
Exp1B	StM	2240.97	0.01	2.34	-1113.41	2273.55	0.00	2.45	-1129.62
	Slot	2240.96	0.00	2.26	-1113.40	2273.55	0.00	2.45	-1129.62
	SR	2240.96	0.00	2.26	-1113.40	2273.55	0.00	2.55	-1129.62
	SA	2242.25	1.29	3.45	-1114.05	2273.86	0.31	2.76	-1129.77
	EnsInt	2248.40	7.44	6.63	-1113.58	2280.70	7.15	6.61	-1129.61
	Swap	2248.12	7.16	6.32	-1113.44	2280.67	7.12	6.82	-1129.60
	VP	2246.86	5.90	5.79	-1112.81	2278.80	5.25	5.53	-1128.66
	VPG	2248.79	7.83	6.95	-1113.78	2280.31	6.76	6.84	-1129.42
Exp2A	StM	1837.75	0.00	3.36	-911.76	1860.37	0.65	3.44	-923.04
	Slot	1837.75	0.00	3.36	-911.76	1860.37	0.65	3.38	-923.04
	SR	1837.75	0.00	3.36	-911.76	1860.37	0.65	3.36	-923.04
	SA	1837.83	0.08	3.44	-911.80	1859.72	0.00	3.10	-922.71
	EnsInt	1842.47	4.72	6.12	-910.56	1865.49	5.78	6.40	-922.03
	Swap	1844.53	6.79	7.52	-911.59	1867.17	7.45	7.64	-922.86
	VP	1838.08	0.33	4.44	-908.37	1860.30	0.58	4.44	-919.43
	VPG	1838.54	0.79	4.40	-908.60	1860.15	0.44	4.24	-919.36
Exp2B	StM	2391.85	0.10	2.94	-1188.60	2353.50	0.00	2.96	-1169.40
	Slot	2393.00	1.26	3.30	-1189.18	2353.63	0.13	3.04	-1169.47
	SR	2391.91	0.16	3.00	-1188.63	2353.50	0.00	2.96	-1169.40
	SA	2391.75	0.00	2.76	-1188.55	2353.50	0.00	2.96	-1169.40
	EnsInt	2399.15	7.40	6.96	-1188.59	2360.40	6.89	7.04	-1169.18
	Swap	2399.16	7.41	7.08	-1188.59	2360.51	7.00	7.44	-1169.23
	VP	2395.52	3.77	4.88	-1186.77	2356.06	2.56	4.84	-1167.01
	VPG	2396.02	4.27	5.08	-1187.03	2355.54	2.04	4.76	-1166.75

		Short SOA				Long SOA			
Experiment	**Model**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**
Exp3	StM	1925.87	0.00	2.55	-956.00	1877.01	0.00	2.34	-931.52
	Slot	1925.87	0.00	2.36	-956.00	1877.10	0.09	2.43	-931.57
	SR	1925.87	0.00	2.45	-956.00	1877.61	0.60	2.86	-931.82
	SA	1927.41	1.54	3.68	-956.77	1877.05	0.04	2.55	-931.54
	EnsInt	1931.54	5.67	5.57	-955.37	1883.98	6.97	6.02	-931.51
	Swap	1927.96	2.09	5.34	-953.58	1883.71	6.70	6.11	-931.38
	VP	1932.77	6.90	6.50	-955.99	1883.43	6.42	6.14	-931.24
	VPG	1935.26	9.39	7.55	-957.23	1885.83	8.82	7.55	-932.44
Exp4	StM	1920.94	0.00	2.43	-953.51	1887.85	0.00	2.57	-936.94
	Slot	1920.94	0.00	2.43	-953.51	1888.18	0.33	2.87	-937.10
	SR	1920.94	0.00	2.37	-953.51	1887.85	0.00	2.57	-936.94
	SA	1922.01	1.07	3.28	-954.04	1888.00	0.15	2.65	-937.01
	EnsInt	1927.05	6.11	5.74	-953.08	1894.64	6.80	6.46	-936.84
	Swap	1927.49	6.54	6.17	-953.30	1894.36	6.51	6.33	-936.70
	VP	1927.41	6.47	6.26	-953.26	1893.35	5.50	5.57	-936.19
	VPG	1928.96	8.02	7.30	-954.04	1895.27	7.42	7.00	-937.15
Exp5	StM	1436.58	0.00	2.34	-711.63	1456.38	0.00	2.39	-721.50
	Slot	1436.70	0.12	2.45	-711.69	1456.46	0.09	2.48	-721.54
	SR	1436.58	0.00	2.30	-711.63	1456.38	0.00	2.39	-721.50
	SA	1437.66	1.08	3.09	-712.17	1456.65	0.27	2.75	-721.63
	EnsInt	1442.68	6.09	6.00	-711.35	1463.00	6.62	6.48	-721.46
	Swap	1443.12	6.53	6.41	-711.57	1463.06	6.69	6.80	-721.49
	VP	1442.67	6.09	6.00	-711.35	1462.16	5.79	5.73	-721.04
	VPG	1444.36	7.78	7.41	-712.19	1463.38	7.01	7.00	-721.65

Table 7

Model parameters

		Exp1A				Exp1B				Exp2A				Exp2B			
		Short SOA		Long SOA		Short SOA		Long SOA		Short SOA		Long SOA		Short SOA		Long SOA	
Model	**Parameter**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
StM	*g*	0.35	0.29	0.31	0.18	0.64	0.22	0.35	0.18	0.06	0.11	0.06	0.10	0.13	0.11	0.06	0.06
	sd	23.38	17.02	17.94	7.07	18.01	6.98	15.09	4.53	18.59	4.25	17.74	3.13	23.50	5.05	20.64	2.99
Slot	*capacity K*	1.30	1.45	1.38	0.36	0.73	0.45	1.30	0.35	1.87	0.23	1.89	0.24	1.75	0.43	1.87	0.27
	sd	23.38	17.02	17.94	7.07	18.01	7.55	15.09	4.53	18.59	4.25	17.74	3.13	23.50	5.04	20.64	3.01
SR	*capacity K*	1.30	1.01	1.38	0.36	0.73	0.45	1.30	0.35	1.87	0.23	1.89	0.28	1.75	0.23	1.87	0.11
	bestSD	22.56	11.92	16.23	6.92	18.08	8.09	13.68	4.45	13.64	3.29	13.07	2.41	17.32	4.52	15.51	2.45
SA	*capacity K*	1.30	1.49	1.38	0.28	1.11	0.25	1.30	0.31	1.87	0.19	1.89	0.49	1.75	0.35	1.87	0.11
	sd	31.45	22.38	17.94	8.03	24.48	14.62	15.18	4.43	18.59	4.22	17.84	2.92	23.50	5.14	20.64	2.99
EnsInt	*g*	0.38	0.31	0.31	0.18	0.45	0.27	0.35	0.18	0.07	0.11	0.06	0.10	0.13	0.11	0.06	0.06
	sd	23.38	15.09	17.94	7.07	19.00	17.13	15.09	4.53	18.40	4.19	17.80	3.07	23.50	5.05	20.63	3.04
	samples	8.05×10^{12}	8.14×10^{12}	6.55×10^{12}	5.85×10^{12}	5.96×10^{12}	7.07×10^{12}	6.13×10^{12}	8.47×10^{12}	2.72	1.87×10^{12}	2.99	5.14×10^{11}	2.35×10^{12}	3.40×10^{12}	9.97×10^{11}	2.48×10^{12}
Swap	*g*	0.35	0.29	0.31	0.18	0.53	0.24	0.35	0.18	0.06	0.09	0.05	0.08	0.13	0.11	0.06	0.06
	B	1.67×10^{-15}	6.28×10^{-3}	2.93×10^{-15}	4.81×10^{-15}	2.89×10^{-15}	5.60×10^{-3}	2.31×10^{-15}	5.32×10^{-3}	4.55×10^{-15}	2.26×10^{-2}	9.33×10^{-15}	1.58×10^{-2}	4.21×10^{-15}	1.33×10^{-3}	5.73×10^{-15}	3.51×10^{-3}
	sd	23.38	17.10	17.94	7.07	18.16	12.15	15.09	4.53	18.67	4.16	17.84	3.09	23.50	5.05	20.64	3.00
VP	*g*	0.29	0.29	0.28	0.18	0.52	0.24	0.34	0.20	0.04	0.11	0.03	0.10	0.09	0.10	0.06	0.05
	mnSTD	38.82	17.37	18.79	11.99	18.54	16.29	15.55	5.98	18.75	5.69	17.95	3.13	23.45	6.85	21.07	3.29
	stdSTD	2.44	11.24	5.34	12.62	2.03	14.30	4.64	4.94	7.37	5.76	6.88	3.21	6.91	5.64	6.65	3.21
VPG	*g*	0.67	0.24	0.35	0.23	0.59	0.22	0.30	0.21	0.03	0.11	0.01	0.11	0.06	0.10	0.03	0.05
	mode-Precision	5.00×10^{-4}	3.11×10^{-3}	5.00×10^{-4}	1.25×10^{-4}	5.00×10^{-4}	6.89×10^{-4}	6.18×10^{-4}	3.00×10^{-4}	4.22×10^{-3}	1.69×10^{-3}	4.72×10^{-3}	2.28×10^{-3}	2.29×10^{-3}	1.04×10^{-3}	2.88×10^{-3}	1.02×10^{-3}
	*sd*Precision	6.48×10^{-4}	10.66	5.01×10^{-4}	7.87×10^{-4}	5.00×10^{-4}	6.37×10^{-3}	5.00×10^{-4}	2.96×10^{-3}	4.22×10^{-3}	3.43×10^{-2}	4.77×10^{-3}	4.38×10^{-3}	1.88×10^{-3}	1.79×10^{-3}	2.78×10^{-3}	2.08×10^{-3}

		Exp3				Exp4				Exp5			
		Short SOA		Long SOA		Short SOA		Long SOA		Short SOA		Long SOA	
Model	**Parameter**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
StM	*g*	0.55	0.28	0.15	0.16	0.45	0.21	0.24	0.16	0.33	0.29	0.34	0.23
	sd	17.65	14.04	18.45	6.09	16.10	6.75	15.63	3.98	18.32	13.83	17.01	6.47
Slot	*capacity K*	0.90	0.57	1.70	0.56	1.09	0.43	1.53	0.43	1.34	1.20	1.31	1.47
	sd	17.65	14.04	18.45	6.08	16.10	6.75	15.63	4.06	18.32	13.69	17.22	6.46
SR	*capacity K*	0.90	0.61	1.70	0.86	1.09	0.46	1.53	0.62	1.34	0.78	1.31	0.46
	bestSD	18.66	12.00	15.34	5.77	16.07	5.20	13.48	3.32	16.07	10.28	15.75	5.01
SA	*capacity K*	1.07	1.16	1.70	0.98	1.09	0.28	1.53	0.33	1.35	0.76	1.36	0.33
	sd	26.78	67.46	19.03	9.32	17.46	9.76	15.90	4.42	23.44	46.85	18.03	8.50
EnsInt	*g*	0.41	0.29	0.15	0.16	0.45	0.21	0.24	0.16	0.33	0.29	0.34	0.23
	sd	19.69	15.74	18.45	6.09	15.90	7.04	15.63	3.99	18.08	14.16	17.01	6.47
	samples	2.65×10^{12}	1.06×10^{13}	6.41×10^{12}	8.65×10^{12}	6.17×10^{12}	6.54×10^{12}	6.06×10^{12}	5.86×10^{12}	4.53×10^{12}	6.56×10^{12}	5.06×10^{12}	1.06×10^{13}
Swap	*g*	0.43	0.24	0.15	0.16	0.45	0.20	0.24	0.15	0.32	0.28	0.33	0.18
	B	1.69×10^{-12}	0.13	5.27×10^{-15}	9.68×10^{-3}	5.37×10^{-15}	0.05	4.42×10^{-15}	2.85×10^{-2}	3.25×10^{-15}	7.55×10^{-2}	2.63×10^{-15}	1.43×10^{-2}
	sd	18.32	12.00	18.45	6.07	16.10	5.69	15.90	3.95	19.55	13.18	17.01	7.32
VP	*g*	0.58	0.29	0.13	0.15	0.45	0.22	0.21	0.16	0.32	0.28	0.28	0.21
	mnSTD	19.19	16.04	19.21	10.46	17.37	9.36	16.85	7.71	20.70	16.06	19.26	13.48
	stdSTD	0.02	5.83	0.95	7.40	0.02	6.29	5.58	6.86	2.68	11.66	6.35	13.93
VPG	*g*	0.66	0.28	0.14	0.20	0.44	0.20	0.21	0.16	0.62	0.29	0.36	0.23
	mode-Precision	5.00×10^{-4}	0.37	5.00×10^{-4}	2.63×10^{-4}	5.00×10^{-4}	2.13×10^{-4}	5.00×10^{-4}	1.69×10^{-4}	5.00×10^{-4}	0.84	5.00×10^{-4}	1.58×10^{-4}
	*sd*Precision	5.00×10^{-4}	14.7	5.00×10^{-4}	1.79×10^{-4}	5.03×10^{-4}	1.07×10^{-3}	5.24×10^{-4}	1.59×10^{-3}	5.10×10^{-4}	10.70	5.05×10^{-4}	1.57×10^{-2}

Wang et al. data

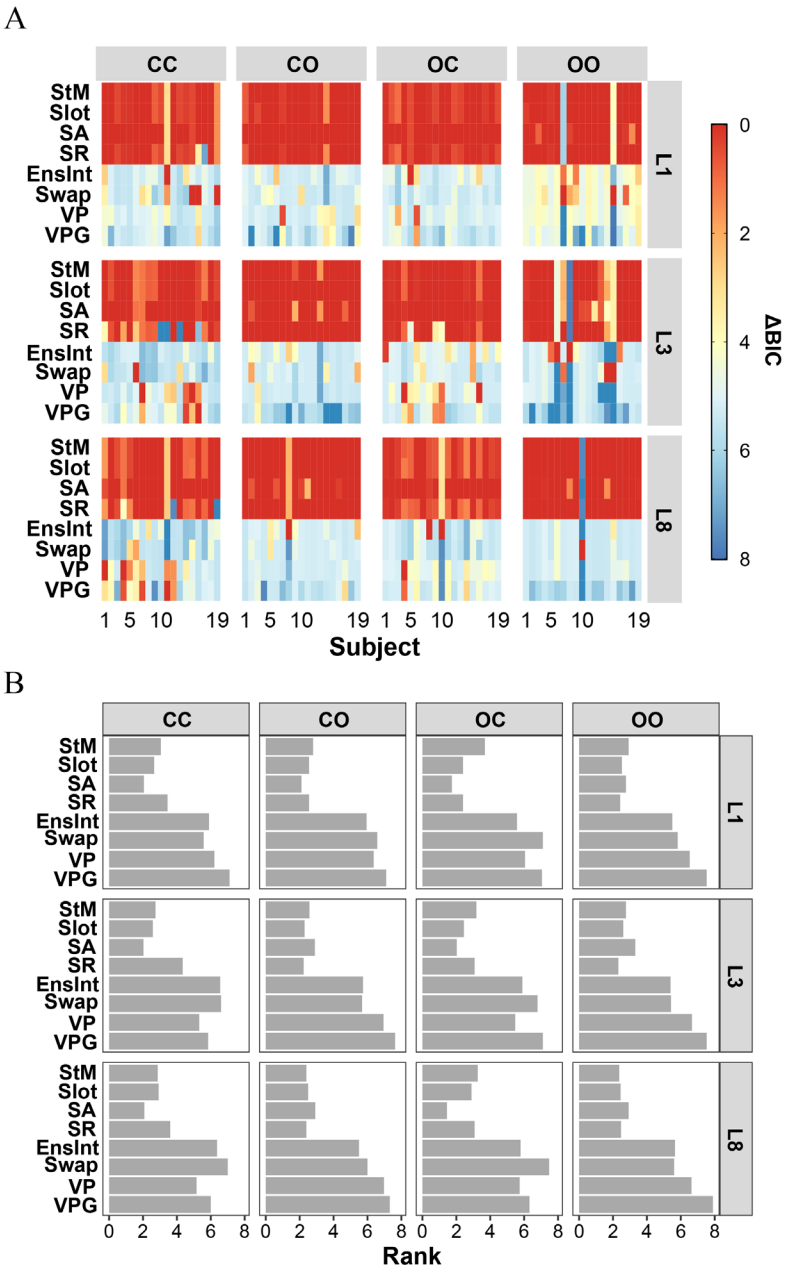
Model Fit and Model Comparison

In general, under different experimental conditions (lags and target pairs), the models with the lowest BIC values were consistent across most subjects, namely the *StM*, *Slot*, *SA* and *SR models* (Figure 6A). Different from the superior and stable performance of the other three top models, the *SR model* had a poor fit for several subjects for the color–color target pair at Lag 3 and 8. When the second target was a color item (color-color or orientation-color target pair), the *VP* and *VPG models* fitted better at Lag 3 than at Lag 1. Especially when the two targets were both colors, the *VP* and *VPG models* had lower BIC values for several subjects, both at Lag 3 and Lag 8, while the *Swap model* best fitted several subjects' data at Lag 1. When the two targets were both orientation Gabors, the *EnsInt* and *Swap models* exhibited their best fitness at Lag 1, obtained lowest BIC values for a few subjects at Lag 3, and the worst fitness consistently at Lag 8. This tendency in the *Swap model*'s performance was similar to that in Tang's two experiments, in which the two targets were orientation Gabors presented in an RSVP stream.

The *SA model* ranked first when the second target was a colored circle, regardless of the lags (see the exact ranks in Table 8; Figure 6B). For the other two target pair conditions, the *SA model* had good performance with only small differences from the best models (Table 8).

Looking at the results for the different target pairs, in specific lag conditions it was obvious to see that the *VP* and *VPG models* reached higher rankings (smaller rank numbers) for the color-color and orientation-color target pairs than for the other target pair conditions, at Lag 3 and Lag 8. At Lag 1, the *EnsInt* and *Swap models* had larger ranks when the two targets shared the same feature (color-color or orientation-orientation), while at Lag 3 and Lag 8, these two models fitted better (having smaller ranks) when the second target was an orientation Gabor.

Fig. 6 Model comparison results for the Wang et al. data. **A.** Model comparison at the subject level. The panels in each row present the results with the same lag but different target pairs; the panels in each column present the results with the same target pair but different lags. 'CC' means color-color target pair; 'CO' means color-orientation target pair; 'OO' means orientation-orientation target pair; 'OC' means orientation-color target pair. L1 = Lag 1; L3 = Lag 3; L8 = Lag 8. **B.** Average model ranking for each lag and each target pair.



Model Simulation

The estimated parameters (Table 9) were used to simulate data from eight models. In the simulated data, we did not observe decreased performance at Lag 3 when both targets were colors (Figure 7; left two panels). When the second target was an orientation Gabor, the actual T2 response errors revealed the same pattern as the other data sets analyzed before. The eight models explained the actual data well for the color-orientation and orientation-orientation target pairs. Nevertheless, at Lag 1 of the color-orientation target pair, the *EnsInt*, *Swap*, *VP* and *VPG models* simulated lower T2 errors. Moreover, the *Swap model* had problems simulating the errors at Lag 3 and Lag 8. Most models simulated T2 error well when both targets were orientations, except for the *Slot model*'s lower T2 errors at Lag 3 and the *StM model*'s lower T2 errors at Lag 8.

Fig. 7 Model simulations for the Wang et al. data. Each panel presents the comparison of the mean T2 error in actual data sets and in data simulated from eight models, separately for each lag and each target pair condition.

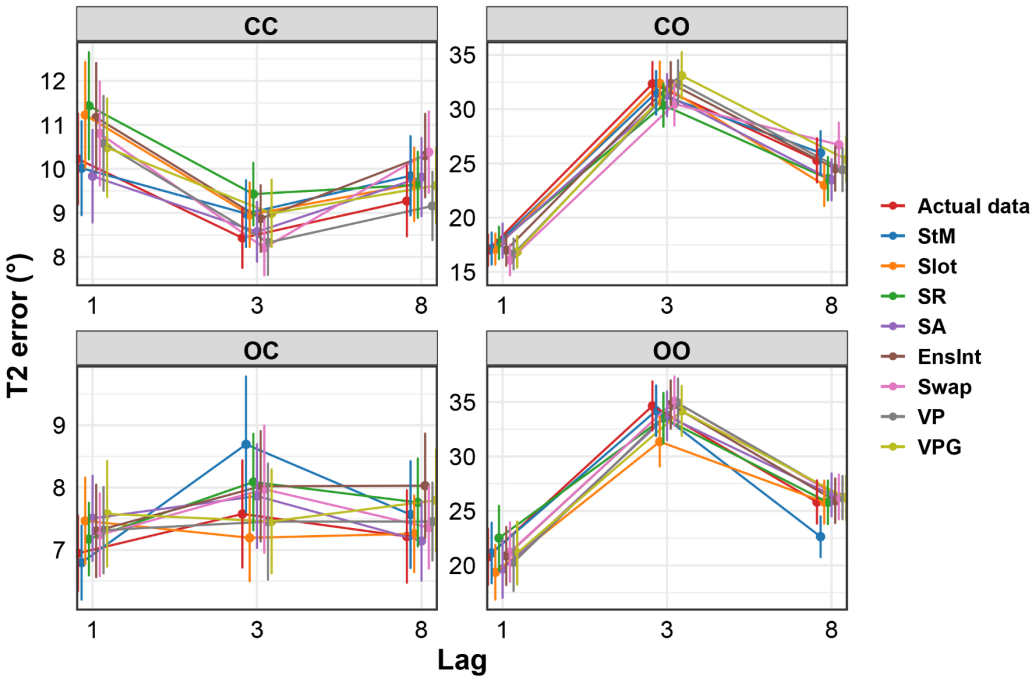


Table 8

Model comparison results

		CC				CO			
Lag	**Model**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**
1	StM	292.13	0.13	3.05	-140.82	348.75	0.14	2.79	-169.04
	Slot	292.13	0.13	2.66	-140.82	348.77	0.16	2.55	-169.05
	SR	292.69	0.69	3.45	-141.10	348.75	0.14	2.55	-169.04
	SA	292.00	0.00	2.05	-140.76	348.62	0.00	2.11	-168.97
	EnsInt	296.48	4.48	5.89	-140.38	353.70	5.08	5.95	-168.84
	Swap	295.85	3.86	5.58	-140.06	353.71	5.10	6.58	-168.85
	VP	296.71	4.71	6.21	-140.49	353.24	4.63	6.37	-168.62
	VPG	297.14	5.14	7.11	-140.71	354.18	5.57	7.11	-169.09
3	StM	379.12	0.21	2.74	-184.01	382.00	0.12	2.58	-185.68
	Slot	379.12	0.21	2.58	-184.01	381.87	0.00	2.29	-185.62
	SR	382.18	3.27	4.34	-185.54	381.87	0.00	2.24	-185.62
	SA	378.91	0.00	2.03	-183.90	382.16	0.28	2.89	-185.76
	EnsInt	384.31	5.41	6.55	-183.83	386.88	5.01	5.74	-185.46
	Swap	383.97	5.06	6.61	-183.66	386.80	4.92	5.68	-185.42
	VP	382.78	3.87	5.32	-183.07	387.22	5.35	6.95	-185.63
	VPG	383.05	4.14	5.84	-183.20	388.23	6.36	7.63	-186.14
8	StM	383.86	0.25	2.87	-186.38	346.27	0.00	2.39	-167.87
	Slot	383.87	0.26	2.92	-186.39	346.27	0.00	2.50	-167.87
	SR	385.21	1.61	3.61	-187.06	346.27	0.00	2.39	-167.87
	SA	383.61	0.00	2.08	-186.25	346.49	0.22	2.92	-167.98
	EnsInt	389.06	5.46	6.37	-186.21	350.75	4.48	5.50	-167.48
	Swap	388.87	5.26	7.00	-186.11	351.33	5.06	6.00	-167.77
	VP	387.14	3.54	5.16	-185.25	351.42	5.15	6.97	-167.81
	VPG	387.55	3.94	6.00	-185.45	351.65	5.38	7.32	-167.92

		OC				OO			
Lag	**Model**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**	**BIC**	**Difference from the best model**	**Average model rank**	**Log-likelihood**
1	StM	246.32	0.19	3.68	-118.01	139.90	0.00	2.92	-65.63
	Slot	246.32	0.19	2.39	-118.01	139.91	0.01	2.53	-65.64
	SR	246.32	0.19	2.39	-118.01	139.90	0.00	2.42	-65.63
	SA	246.13	0.00	1.74	-117.92	140.05	0.15	2.76	-65.71
	EnsInt	250.71	4.58	5.58	-117.63	143.37	3.47	5.50	-65.21
	Swap	251.43	5.30	7.11	-117.99	142.79	2.89	5.82	-64.92
	VP	250.75	4.62	6.05	-117.65	144.17	4.27	6.53	-65.61
	VPG	251.29	5.16	7.05	-117.92	145.04	5.13	7.53	-66.04
3	StM	227.58	0.11	3.18	-108.76	326.45	0.20	2.76	-158.10
	Slot	227.58	0.11	2.45	-108.76	326.40	0.14	2.61	-158.07
	SR	228.22	0.75	3.08	-109.09	326.25	0.00	2.32	-158.00
	SA	227.47	0.00	2.03	-108.71	326.74	0.48	3.32	-158.24
	EnsInt	231.75	4.28	5.89	-108.34	329.97	3.72	5.39	-157.30
	Swap	232.39	4.92	6.79	-108.66	330.30	4.05	5.42	-157.46
	VP	231.40	3.93	5.47	-108.16	331.67	5.41	6.66	-158.14
	VPG	232.06	4.59	7.11	-108.49	332.36	6.11	7.53	-158.49
8	StM	227.24	0.37	3.26	-108.60	341.08	0.00	2.37	-165.34
	Slot	227.27	0.39	2.89	-108.61	341.08	0.00	2.45	-165.34
	SR	227.34	0.47	3.08	-108.65	341.08	0.00	2.47	-165.34
	SA	226.88	0.00	1.45	-108.41	341.26	0.18	2.92	-165.42
	EnsInt	231.47	4.59	5.79	-108.19	346.20	5.12	5.66	-165.29
	Swap	232.27	5.39	7.47	-108.60	345.62	4.54	5.61	-165.00
	VP	231.17	4.30	5.74	-108.05	346.20	5.12	6.63	-165.29
	VPG	231.43	4.55	6.32	-108.17	346.92	5.83	7.89	-165.65

Table 9

Model parameters

		CC						CO					
		Lag 1		Lag 3		Lag 8		Lag 1		Lag 3		Lag 8	
Model	**Parameter**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
StM	*g*	0.08	0.11	0.04	0.04	0.04	0.08	0.07	0.16	0.26	0.29	0.33	0.28
	sd	9.21	2.39	9.11	1.81	9.15	1.63	14.47	8.16	25.94	17.43	12.99	5.35
Slot	capacity	1.84	2.50	1.91	0.90	1.92	1.06	1.86	1.37	1.35	1.26	1.34	0.55
	sd	9.21	2.39	9.11	1.81	9.15	1.68	14.75	8.15	19.58	17.03	12.99	9.12
SR	capacity	1.89	0.86	2.59	1.00	2.62	0.91	1.86	0.72	1.35	0.93	1.40	0.53
	bestSD	6.87	1.91	7.64	1.30	7.12	1.14	10.73	5.75	22.31	10.67	11.37	38.06
SA	capacity	1.84	0.62	1.93	1.43	1.92	1.75	1.86	1.11	1.53	1.45	1.40	0.40
	sd	9.66	2.46	9.57	2.62	9.91	3.73	14.47	8.65	33.24	26.43	14.61	72.65
EnsInt	*g*	0.07	0.11	0.04	0.04	0.04	0.08	0.09	0.20	0.24	0.29	0.31	0.26
	sd	9.04	2.52	9.11	1.85	9.03	1.62	13.73	5.14	25.94	18.60	12.77	7.91
	samples	5.82	2.10×10^{12}	4.22×10^{11}	3.73×10^{12}	14.94	2.15×10^{12}	7.70	2.05×10^{12}	1.07×10^{12}	5.25×10^{12}	10.80	5.82
Swap	*g*	1.27×10^{-7}	7.50×10^{-2}	3.52×10^{-2}	4.11×10^{-2}	4.28×10^{-4}	0.07	0.04	0.13	0.16	0.25	0.29	1.27×10^{-7}
	B	8.50×10^{-10}	0.05	6.06×10^{-14}	1.72×10^{-2}	1.50×10^{-13}	0.02	5.61×10^{-8}	5.67×10^{-2}	9.83×10^{-11}	0.14	3.60×10^{-14}	8.50×10^{-10}
	sd	9.46	2.16	9.06	1.71	9.15	1.80	14.47	7.50	30.39	16.31	12.47	9.46
VP	*g*	0.05	0.11	1.73×10^{-14}	2.86×10^{-2}	7.96×10^{-15}	6.32×10^{-2}	5.58×10^{-2}	0.14	0.32	0.27	0.33	0.05
	*mn*STD	9.21	2.60	8.18	2.62	8.78	2.35	14.76	12.03	25.93	15.98	13.22	9.21
	*std*STD	2.63	3.16	3.36	3.34	4.43	2.76	3.29	11.98	0.02	8.38	0.02	2.63
VPG	*g*	2.84×10^{-3}	0.11	3.0×10^{-7}	2.43×10^{-2}	3.36×10^{-4}	4.89×10^{-2}	7.48×10^{-2}	0.20	0.63	0.25	0.32	2.84×10^{-3}
	*mode*Precision	1.67×10^{-3}	1.04×10^{-3}	2.03×10^{-3}	1.7×10^{-3}	1.86×10^{-3}	4.49×10^{-3}	7.09×10^{-4}	5.02×10^{-4}	5.00×10^{-4}	0.20	7.94×10^{-4}	1.67×10^{-3}
	*sd*Precision	6.98×10^{-4}	1.53×10^{-2}	1.51×10^{-3}	3.63×10^{-2}	1.81×10^{-3}	3.17×10^{-2}	5.00×10^{-4}	7.59E-03	5.00×10^{-4}	15.80	5.00×10^{-4}	6.98×10^{-4}

		OC						OO					
		Lag 1		Lag 3		Lag 8		Lag 1		Lag 3		Lag 8	
Model	**Parameter**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
StM	*g*	2.27×10^{-14}	0.03	4.87×10^{-14}	5.41×10^{-2}	2.67×10^{-14}	3.83×10^{-2}	8.32×10^{-14}	0.29	0.48	0.38	0.29	0.27
	sd	7.79	2.92	7.78	2.66	7.51	2.40	15.56	12.02	16.34	16.66	18.94	12.25
Slot	capacity	2.47	2.62	2.19	1.84	2.62	2.58	2.14	1.20	1.04	1.29	1.42	1.93
	sd	7.79	2.92	7.78	2.66	7.67	2.50	15.56	11.94	21.61	18.32	18.94	12.25
SR	capacity	2.73	0.74	2.78	1.03	2.77	0.86	2.06	1.30	1.26	1.12	1.42	0.73
	bestSD	5.51	2.05	5.80	1.78	6.29	1.78	13.26	8.07	19.29	44.53	16.06	9.42
SA	capacity	2.99	1.98	2.38	1.33	2.90	1.48	2.00	2.23	1.40	1.16	1.42	1.50
	sd	11.30	3.43	8.89	3.04	9.97	4.28	19.40	22.45	39.64	90.98	20.07	18.06
EnsInt	*g*	0.00	0.06	0.00	0.05	0.00	0.04	0.00	0.25	0.52	0.38	0.30	0.27
	sd	7.79	3.08	7.58	2.60	7.51	2.56	18.75	11.01	19.87	23.04	18.69	7.91
	samples	1.55	1.80×10^{12}	1.56	2.51×10^{12}	15.24	1.52×10^{12}	1.08×10^{12}	7.33×10^{12}	0.56	6.64×10^{12}	6.57×10^{12}	9.01×10^{12}
Swap	*g*	4.74×10^{-9}	2.88×10^{-2}	5.19×10^{-9}	4.76×10^{-2}	1.59×10^{-8}	3.83×10^{-2}	2.99×10^{-5}	0.24	6.38×10^{-2}	0.30	0.29	0.24
	B	1.13×10^{-13}	1.36×10^{-2}	1.23×10^{-12}	2.29×10^{-2}	1.49×10^{-14}	8.68×10^{-6}	6.57×10^{-6}	0.18	2.11×10^{-10}	0.32	9.26×10^{-15}	0.11
	sd	7.79	2.77	7.78	2.64	7.51	2.40	14.49	10.60	27.28	18.24	17.94	11.64
VP	*g*	5.80×10^{-15}	2.54×10^{-2}	5.01×10^{-14}	0.04	4.49×10^{-15}	0.03	1.15×10^{-14}	0.28	0.49	0.35	0.29	0.27
	*mn*STD	7.69	3.30	7.49	3.82	7.36	2.71	16.34	12.36	35.49	16.36	19.66	14.08
	*std*STD	0.88	2.16	1.69	3.20	2.88	2.01	0.02	4.54	0.02	2.99	0.02	9.76
VPG	*g*	3.55×10^{-8}	1.64×10^{-2}	1.97×10^{-4}	3.79×10^{-2}	4.30×10^{-8}	0.03	0.24	0.31	0.67	0.24	0.36	0.29
	*mode*Precision	2.03×10^{-3}	3.67×10^{-3}	2.73×10^{-3}	8.64×10^{-3}	2.75×10^{-3}	1.86×10^{-3}	5.97×10^{-4}	3.73×10^{-2}	5.50×10^{-4}	6.73×10^{-3}	5.00×10^{-4}	1.85×10^{-4}
	*sd*Precision	5.05×10^{-4}	0.01	5.00×10^{-4}	3.82×10^{-2}	1.48×10^{-3}	4.40×10^{-3}	5.34×10^{-4}	0.12	5.00×10^{-4}	11.50	5.00×10^{-4}	7.98×10^{-4}

Discussion

Our study's fundamental motivation was to systematically compare eight commonly used VWM models in the AB domain: the *StM*, the *Slot*, the *SA*, the *SR*, the *EnsInt*, the *Swap*, the *VP*, and the *VPG models*. Previous research has successfully used the *StM model* to account for the representations in the AB task (Asplund et al., 2014; Sy et al., 2021; Karabay et al., 2022), but some studies have suggested that VWM representations could also be explained well with variable precision across and within trials (the *VP* and *VPG models*; Fougnie et al., 2012; Van den Berg et al., 2012). Meanwhile, the *EnsInt* and *Swap models*, in which potentially important nontarget reports are considered, have gained little attention in the model choices in the AB domain to date. Since no existing studies have systematically compared these VWM models, we did so in the present study. We applied MLE for the parameter estimation, which allowed us to assess these models' fitness to four data sets of the AB task, from three different laboratories, and then compare them based on the BIC. Moreover, we utilized the parameter values of each model to generate simulated data sets, in order to further investigate if these models produce a good description of the experimental data in all circumstances.

Our findings showed a clear separation between the models. Both at the group and subject level, there was substantial evidence for the superiority of the *StM*, *Slot*, *SA* and *SR models*, across different experiments, whereas the *EnsInt*, *Swap*, *VP*, and *VPG models* were not among the best models. These results align with that of a previous study by Asplund et al. (2014), who also found that the *StM model* better explains the data from the AB task over the variable precision model families. The four 'top' models (the *StM*, *Slot*, *SA* and *SR models*) share a crucial characteristic, that is, they are based on the hypothesis that not all the items can be encoded into VWM, but only a fixed and limited number of items can be stored in VWM. In the error distributions of these models, once the set sizes exceed the number of available 'storage slots' in VWM, the errors obey a uniform distribution representing these random responses.

In our results, there were minor differences among these four models. A possible explanation for this might be that the set size of all the data sets we analyzed here can be considered as two (the first and second targets), which is most probably smaller than the number of available VWM slots (cf. Cowan, 2001). The hypothesis regarding the set sizes is the major theoretical divergency of these four models. For the *Slot*, *SA* and *SR models*, the proportion of remembered items is based on predictions about the available memory slots (the *capacity K*), while the *StM model* has no predictions about the exact number of slots. Therefore, our model comparison results between these four models may be limited by the constant and same set size across different data sets. In this case, future research, in which the different set sizes are applied in the AB domain, will need to be undertaken (e.g., Akyürek, Hommel, & Jolicœur, 2007).

One of our interests in this article was to investigate if the time interval (lag or SOA) between the two targets can be a factor in comparing models' fitness. The results were mixed when comparing a single model's performance across different lags. In the two experiments of Tang et al. (2020) and several experiments (except Experiment 2A and 2B) of Karabay et al. (2022), the *SA model* provided a better explanation of the data at longer time intervals. A possible account for this pattern may be that when the set size is smaller than the available 'memory slots', one item may be stored in multiple slots, according to the assumptions of the *SA model*, which leads to an averaging representation precision. Hence, it is possible that, when the time interval is short, there is not enough time for resource allocation and averaging (Bonnel & Miller, 1994; Luck et al., 1996). Nevertheless, the lag between targets did not always affect model performance in the color AB task. For example, Experiment 2A and 2B of the Karabay et al. (2022), the Asplund et al. (2014) data, and the Wang et al. data set with the color-color target pair, there was no evidence showing that the time interval influenced the model fitness.

Regarding the models that consider nontarget reports (the *EnsInt* and *Swap models*), it should be noted that these models performed better at shorter lags, and were highly favored at

Lag 1 (See Fig 3). In the present datasets, this tendency was only observed under the single-stream RSVP paradigm, with the two targets both being orientations. Our results about the performance of the *Swap model* accorded with previous research, which has demonstrated that temporal target integration and target report order reversals occurred frequently at Lag 1 (Akyürek et al., 2007, 2012; Hommel & Akyürek, 2005). This, in turn, is in line with the account offered by the eSTST model of the AB (Wyble et al., 2009) for the phenomenon of Lag 1 sparing (i.e., the lack of an AB at that lag), which holds that both targets are consolidated together when they successively follow each other without intervening distractors. Although the inner mechanism of the *EnsInt model* and the *Swap models* is different, the key divergence between these two models and the others is that they suggest the representations of each item in the VWM are not independent but can be influenced by other items. Especially for the AB task, the T2 error distribution predicted by these two models can be affected by the first target. It might thus be reasonable to conclude that the failure of these two models at longer lags could be attributed to the weaker influence of T1 on T2.

For the *VP* and *VPG models*, our findings indicate that they have better performance at longer lag or SOA, in each of the data sets. First, looking at the subject level, more subjects favored these two models at longer lags. Then at the group level, the rankings for these two models were slightly higher at longer lags than those at shorter lags. In previous studies, the assessments for the models with variable precision were mixed. Asplund and his colleagues (Asplund et al., 2014), who suggested that the perception in the AB is a discrete process in their study, showed that the variable precision model did not fit the data well, but the standard mixture model did. However, Sy et al. (2021) demonstrated that the variable precision model could describe the graded loss of T2 information in the AB in their data. Although our analysis cannot address whether the nature of the awareness in the AB is gradual or discrete, the findings

that the *VP* and *VPG models*' performance is affected by the lag may imply that lag may also be a factor in the process of visual information gaining access to awareness.

Given the results we discussed above, it should be noted that there was some evidence that target features also affect model performance to an extent. For the Wang et al. data set, the model comparison results of distinct experimental settings within the same study showed strong evidence that the *VP* and *VPG models* performed better for colored targets over orientations. This finding broadly supports the work of Bae et al. (2014), who reported increased variability in color WM. Furthermore, the performance of the *EnsInt* and the *Swap models* was also associated with the target features. These models explained the data better when the two targets shared a common feature, such as the orientation-orientation and the color-color target pairs in the Wang et al. data set. Based on the core principles behind the *EnsInt* and the *Swap models*, it is plausible that when the two targets in the AB task have the same characteristic, the T2 response is more easily influenced by T1. Indeed such interactions including attraction and repulsion are also observed in VWM tasks when two targets share the same feature (Chunharas et al., 2022), highlighting an interesting commonality between the two domains. In future research in the AB domain, it might be advisable to select and assess model-based performance, dependent on the specific target feature applied in the task.

Conclusion

In summary, this study set out to evaluate the fitness of eight commonly used VWM models in four different data sets obtained from the AB task. We found that the models that hypothesize there are limited 'storage slots' in VWM, such as the *StM model*, the *Slot model* and its variants, performed best in accounting for the data in the majority of experiments we analyzed. The second significant finding was that both the lag and the commonality in, and the kind of, target feature (color or orientation) played an important role in model performance. In

our simulation analysis, the divergences among the synthetic data sets from all the eight models were too small to draw strong inferences, but our simulation results at least showed the applicability of these VWM models in the AB domain. Overall, the systematical model comparisons in our study, not only across different lag conditions, but also across different kinds of targets, may help to guide the future selection of models to assess representation and task performance in the AB field.

Declarations

Acknowledgements

Shuyao Wang was supported by the China Scholarship Council (CSC), grant 201906020179. We would like to thank Christopher L. Asplund for providing the original data sets from Asplund et al.(2014).

Conflicts of interest

Authors declare that they have no conflict of interest.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the Declaration of Helsinki (2008). Approval was obtained from the ethical committee of the Psychology Department of the University of Groningen (Approval code PSY-1819-S-0209).

Consent to participate

Informed consent was obtained from all individual subjects included in the study.

Open Practices Statements

Data generated during the study (Wang et al. data sets), and all the analysis scripts are available on the Open Science Framework at: https://osf.io/zak5f/. The study was not preregistered.

The data from Tang et al. (2020) and Karabay et al. (2022) are publicly available through their published articles, which are cited in the reference section. The Tang et al. (202) data can be

accessed at: https://osf.io/f9g6h. The Karabay et al. (2022) data can be accessed at: https://osf.io/x5dru.

References

Akyürek, E. G., Eshuis, S. A. H., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at Lag 1 in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1448–1464. https://doi.org/10.1037/a0027610

Akyürek, E. G., & Hommel, B. (2005). Target integration and the attentional blink. *Acta Psychologica*, *119*(3), 305–314. https://doi.org/10.1016/j.actpsy.2005.02.006

Akyürek, E. G., Hommel, B., & Jolicœur, P. (2007). Direct evidence for a role of working memory in the attentional blink. *Memory & Cognition 2007 35:4*, *35*(4), 621–627. https://doi.org/10.3758/BF03193300

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/J.0963-7214.2004.01502006.X

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W., & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological Science*, *25*(3), 824–831. https://doi.org/10.1177/0956797613513810

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7–7. https://doi.org/10.1167/14.4.7

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), 1–11. https://doi.org/10.1167/9.10.7

Bonnel, A. M., & Miller, J. (1994). Attentional effects on concurrent psychophysical discriminations: investigations of a sample-size model. *Perception & Psychophysics*, *55*(2), 162–179. https://doi.org/10.3758/BF03211664

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*(1), 38–70. https://doi.org/10.1037/0033-295X.114.1.38

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. https://doi.org/10.1177/0956797610397956

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics 1987 42:2*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

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*. https://doi.org/10.1037/XGE0001191

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. https://doi.org/10.1017/S0140525X01003922

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, *120*(4), 873–902. https://doi.org/10.1037/a0034247

Dux, P. E., & and Rene Marois. (2009). The attentional blink: A review of data and theory. *Attention, Perception, and Psychophysics*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical

power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), 3–3. https://doi.org/10.1167/11.12.3

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*(1), 1–8. https://doi.org/10.1038/ncomms2237

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer New York, NY. https://doi.org/10.1007/978-0-387-84858-7

Karabay, A., Wilhelm, S. A., de Jong, J., Wang, J., Martens, S., & Akyürek, E. G. (2022). Two faces of perceptual awareness during the attentional blink: gradual and discrete. *Journal of Experimental Psychology: General*, *151*, 1520–1541. https://doi.org/10.1037/xge0001156

Kool, W., Conway, A. R. A., & Turk-Browne, N. B. (2014). Sequential dynamics in visual short-term memory. *Attention, Perception, and Psychophysics*, *76*(7), 1885–1901. https://doi.org/10.3758/s13414-014-0755-7

Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? *Journal of Experimental Psychology. Human Perception and Performance*, *22*(3), 725–737. https://doi.org/10.1037//0096-1523.22.3.725

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–284. https://doi.org/10.1038/36846

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind

spot in perceptual awareness. *Neuroscience and Biobehavioral Reviews*, *34*(6), 947–957. https://doi.org/10.1016/j.neubiorev.2009.12.005

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/S13428-011-0168-7

Oberauer, K. (2021). Measurement models for visual working memory—A factorial model comparison. *Psychological Review*. https://doi.org/10.1037/rev0000328

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin and Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

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), 1–27. https://doi.org/10.1167/17.5.11

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332–350. https://doi.org/10.1037/0096-1523.16.2.332

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*(JAN), 10. https://doi.org/10.3389/neuro.11.010.2008

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*(1), 10–15. https://doi.org/10.1037/h0027470

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology. Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037//0096-

1523.18.3.849

Schwarz, G. (2007). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9–9. https://doi.org/10.1167/13.10.9

Sy, J. L., Miao, H.-Y., Marois, R., & Tong, F. (2021). Conscious perception can be both graded and discrete. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001009

Tang, M. F., Ford, L., Arabzadeh, E., Enns, J. T., Visser, T. A. W., & Mattingley, J. B. (2020). Neural dynamics of the attentional blink revealed by encoding orientation selectivity during rapid visual presentation. *Nature Communications*, *11*(1), 1–14. https://doi.org/10.1038/s41467-019-14107-z

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, *7*. https://doi.org/10.7554/eLife.34963

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 of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

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

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: sparing at a cost. *Journal of Experimental Psychology: Human*

Perception and Performance, *35*(3), 787–807. https://doi.org/10.1037/A0013902

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860