# An Interference Model for Visual and Verbal Working Memory

Klaus Oberauer and Hsuan-Yu Lin

University of Zurich

Author Note

Klaus Oberauer and Hsuan-Yu Lin, Department of Psychology, University of Zurich. This research was supported by grants from the Swiss National Science Foundation (project n° 100014_135002 and 100014_179002) to Klaus Oberauer. We thank Sonja Peteranderl and Justus Spengler for help with collecting the data.

Correspondence should be addressed to Klaus Oberauer, Department of Psychology, University of Zurich, Binzmühlestrasse 14, 8050 Zürich, Switzerland; email: k.oberauer@psychologie.uzh.ch

The data and the modeling code are available on the Open Science Framework at https://osf.io/znqmw

Abstract

Research on working memory has followed two largely independent traditions: One concerned with memory for sequentially presented lists of discrete items, and the other with short-term maintenance of simultaneously presented arrays of objects with simple, continuously varying features. Here we present a formal model of working memory, the Interference Model, that explains benchmark findings from both traditions: The shape of the error distribution from continuous reproduction of visual features, and how it is affected by memory set size; the effects of serial position for sequentially presented items, the effect of output position, and the intrusion of non-targets as a function of their distance from the target in space and in time. We apply the model to two experiments combining features of popular paradigms from both traditions: Lists of colors (Experiment 1) or of nonwords (Experiment 2) are presented sequentially and tested through selection of the target from a set of candidates, ordered by their similarity. The core assumptions of the Interference Model are: Contents are encoded into working memory through temporary bindings to contexts that serve as retrieval cues to access the contents. Bindings have limited precision on the context and the content dimension. A subset of the memory set – usually one item and its context – is maintained in a focus of attention with high precision. Successive events in an episode are encoded with decreasing strength, generating a primacy gradient. With each encoded event, automatic updating of working memory reduces the strength of preceding memories, creating a recency gradient and output interference.

Key words: working memory, serial-order memory, computational model, serial position, set size.

## An Interference Model for Visual and Verbal Working Memory

Working memory is a system for holding information available for our current cognitive activities. It has a limited capacity, and that capacity limit is probably the main cause for many of our cognitive limitations (Oberauer, 2017). There are two traditions of research on working memory (WM). The older tradition emerged from work on verbal learning and on memory systems, and has been kick-started by the publication of the multi-component model of working memory (Baddeley & Hitch, 1974). Its main focus is the short-term retention of brief lists of discrete, usually verbal items – digits, letters, words, sometimes spatial locations – which participants are typically asked to recall in their order of presentation (for a review see Hurlstone, Hitch, & Baddeley, 2014).

A more recent tradition emerged from the psychology of visual perception and attention, which showed that people remember only a small amount of information from a complex scene over the time of an eye blink or a saccade (Simons & Lewin, 1998). This has motivated the concept of a visual WM, which has been intensely studied since the late 1990 using tasks that test people's short-term retention of arrays of visual objects with one or a few simple visual features (e.g., a number of colored squares scattered on the screen). In the change-detection task (Vogel, Woodman, & Luck, 2001) memory is tested through a probe presented in the location of the target object in the original array, and participants decide whether the probe has changed in comparison to the target object. A second paradigm, continuous reproduction (Blake, Cepeda, & Hiris, 1997; Wilken & Ma, 2004), capitalizes on the fact that many visual features (e.g., color hue, line orientation, size) vary on a continuous scale. In this task, a target object from the array is identified at test (usually by marking its location) and participants try to reproduce its feature on a continuous scale (e.g., clicking on it on a color wheel).

### Two Research Traditions

The two research traditions have generated different kinds of data and different models of WM. The aim of the present research is to bring the two traditions together by generating data that

combine the respective strengths of each tradition, and by proposing a formal model that explains these data. We start with a brief review of the two lines of research, with an emphasis on the empirical phenomena that we aim to explain. We have selected these phenomena because they have been the main targets of separate modelling efforts in the two research traditions, and have been identified as benchmark findings that should receive high priority as targets of explanation in the field of working memory (Oberauer et al., 2018).

**Working Memory for Lists of Discrete Items**

The older tradition generated a wealth of knowledge about the error patterns in memory for lists. Among the most robust and important phenomena are the serial-position curve and the transposition gradient. The *serial-position curve* is shown in plots of accuracy as a function of the item's position in the list and shows a primacy and a recency effect: Items at the beginning and at the end of the list are remembered better than in the middle. The balance of primacy and recency depends on the test procedure (Madigan, 1980; Oberauer, 2003), but the U-shaped curve is nearly universally found. The *transposition gradient* is shown by plotting confusions of list items with other list items as a function of the migration distance of the item between its original *input position* in the list and the list position at which it is recalled. The typical observation is that confusions occur predominantly with close neighbors; the confusion probability decreases with larger migration distance (Henson, Norris, Page, & Baddeley, 1996; Murdock & vom Saal, 1967).

Theories of WM emerging from this tradition often take the form of conceptual frameworks with a broad scope (Baddeley, 1986; Cowan, 1995), but detailed mechanistic models have also been designed, which primarily explain memory for serial order of lists (Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Farrell, 2008; Page & Norris, 1998). A number of principles emerged from these efforts that many of these models share, and through a detailed analysis of error patterns and recall latencies Farrell and Lewandowsky (2004) identified those that a successful model of serial recall probably needs: (1) Access to items in lists relies on cue-based retrieval and competitive cueing: Given a retrieval cue, all list items, and often additional extra-list items, receive different

degrees of activation, and the item with the highest activation is selected for retrieval. (2) Serial order is represented by temporary bindings between list items and representations of their temporal or ordinal position in the list; these position representations serve as retrieval cues for the items bound to them. (3) List items are encoded with decreasing strength, resulting in a primacy gradient of memory strength. (4) Response suppression: Once recalled, an item is suppressed in memory to avoid repeating it. These principles together provide qualitative explanations of the serial-position effects and the transposition gradient: Because the similarity between position representations decreases with their distance, and cue-based retrieval engenders confusions as a function of cue similarity, the chance of intra-list confusions decreases with distance in the list. As a side effect, the first and the last list item are confused less often because they have neighbors on only one side of the list; this *edge effect* contributes to the serial-position effects. In addition, the primacy effect is a consequence of the primacy gradient of memory strength, and when the list is recalled in forward order, the recency effect arises from response suppression: Towards the end of the list, most list items have been suppressed, reducing the chance of confusing the final items with other list items.

**Working Memory for Arrays of Visual Objects**

Experiments studying visual WM have often investigated the effect of set size (i.e., the number of objects to be remembered) because it reveals the limited capacity of WM. In addition, the continuous-reproduction paradigm yields a continuous measure of recall error, providing an assessment of the precision with which the tested feature is remembered. Distributions of this error have a characteristic shape that deviates from a normal distribution by its heavy tails: At larger set sizes in particular, there are many more very large errors than expected under a normal distribution. Zhang and Luck (2008) have therefore proposed to model the error distributions as a mixture of a normal and a uniform distribution.[1] This mixture model was subsequently extended by a third component reflecting transposition errors (a.k.a. swap errors): Responses closely matching the

[1] Because responses in continuous-reproduction tasks are usually given on a circular scale, the normal distribution is approximated as a von-Mises distribution.

feature of a non-target object in the array (Bays, Catalao, & Husain, 2009). A common finding is that, as set size is increased, the error distributions get wider, and in particular the heavy tails increase, reflecting a growing prevalence of responses far away from the target on the feature similarity scale (Zhang & Luck, 2008).

Theories of visual WM predominantly focus on explaining the capacity limit as manifested in the effect of set size on error rates, and in particular on the shape of the continuous error distribution. Until recently the main contenders were discrete-capacity models (sometimes referred to as "slot models") on the one hand, and continuous-resource models on the other. Discrete-capacity models (Fukuda, Awh, & Vogel, 2010; Luck & Vogel, 1997; Rouder et al., 2008; Zhang & Luck, 2008) build on the assumption that visual WM has a limited number of place holders or slots, each of which can hold one visual object with a certain degree of precision. Any array object exceeding the number of slots is not represented in WM at all, so that when such an object is tested, the person can only guess. By contrast, continuous-resource models (Bays & Husain, 2008; Ma, Husain, & Bays, 2014; Schneegans & Bays, 2017; Smith, Lilburn, Corbett, Sewell, & Kyllingsbæk, 2016; van den Berg & Ma, 2018; van den Berg, Shin, Chou, George, & Ma, 2012) build on the assumption that all array objects are represented in WM, and a limited resource is shared between them. As set size increases, each object is assigned a smaller resource share, which translates into lower precision.

These competing theoretical ideas have been incorporated into formal models that account for the shape of the error distribution in continuous reproduction and how it changes with set size. A comparison of several variants of discrete-capacity and continuous-resource models (van den Berg, Awh, & Ma, 2014) has identified two features that the most successful models share: The precision with which an object's features are remembered varies randomly both within and between trials, and a proportion of responses, which increases with set size, reflect the feature of a non-target rather than the target (i.e., swap errors). Whether all array objects or only a maximum number of them was remembered made comparatively little difference for the model fit.

A few years ago, we broadened the theoretical field by proposing an interference model of visual WM (Oberauer & Lin, 2017). This model borrows some of the principles of models of serial-order memory: Arrays are encoded by binding each object to its spatial location in the same way as list items are bound to their location on the time axis, and they are accessed through cue-based retrieval, using spatial locations as cues. This access mechanism naturally leads to confusions between array objects, explaining the occurrence of swap errors. The model also accounts for the set-size effects on the shape of the error distribution from continuous-reproduction tasks, providing a better fit than slot models and continuous-resource models. Therefore, we argue that the Interference Model (IM) is a good starting point for a model that explains phenomena from both research traditions. We next consider what is needed to integrate these two traditions empirically and theoretically.

**Building the Bridge**

The two research traditions are built on different task paradigms generating different kinds of data, and have generated different theories of WM. The list-memory tradition, using discrete items as materials, has not been concerned with measuring the precision of memory. Although a set-size (or list-length) effect is well documented for list-memory tasks (Oberauer et al., 2018), explaining it has not been a primary goal for theorists in this area (for an exception see Cowan, 2001; Cowan, 2005). Conversely, the visual-WM tradition, using nearly exclusively arrays of simultaneously presented objects, has not been concerned much with effects of serial order (for an exception see Smith et al., 2016). On the theoretical side, the discussion in the visual-WM literature has long been dominated by the controversy between discrete-capacity and continuous-resource theories. These theories have in common that they attribute performance limitations to a (discrete or continuous) commodity for storage; access to what is stored is largely taken for granted (for a dissenting voice see Rajsic, Swan, Wilson, & Pratt, 2017). By contrast, the theoretical landscape in the tradition of list memory is broader. Nelson Cowan has long argued for a discrete-capacity theory (Cowan, 2001); continuous-resource theories have been popular in earlier decades (Anderson, Reder, & Lebiere,

1996; Just & Carpenter, 1992) but have not been advanced recently. Formal models of serial-order memory, however, have not incorporated either of these ideas. Instead, they have attributed performance limitations to difficulties in accessing the right item: Cue-based retrieval combined with competitive cueing engenders the risk of confusing the target item with other potentially recallable items. In addition, models of WM for lists have assumed that time-based decay (Barrouillet, Portrat, & Camos, 2011) or representational interference (Nairne, 1990; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012) contribute to the limit on how much people can retrieve from WM.

The aim of the present research is to build a bridge between the two lines of research on WM. This involves as a first step the development of task paradigms that combine the respective strengths of both traditions: The measurement of serial order effects such as the serial-position curve and transposition errors, and the measurement of memory precision through a continuous response scale. We carried out two experiments that test memory for sequentially presented lists, asking participants to recall every list item on a continuous response scale. Experiment 1 used colors as stimuli, which participants reproduced by selecting the target color on a color wheel. Experiment 2 used verbal stimuli. In an attempt to approximate a continuous response scale, we constructed a set of nonwords with a circular similarity structure: Nonwords were arranged on a closed-loop similarity scale such that each nonword differed from its neighbors on the scale by only one or two phonemic features in one of its phonemes. With this scale we aimed to measure the precision of memory for verbal materials.

The second step of building the bridge is to develop a formal model that explains the data from these experiments. We took the IM for the continuous reproduction of visual features (Oberauer & Lin, 2017) as a starting point and augmented it with mechanisms of serial-recall models so that it can account for the effects of serial position. In the next section we explain the model, followed by presentation of the experiments, and the application of the model to the data from these experiments.

## The Interference Model

We first describe the basic architecture of the IM, and the additional assumptions we incorporated about the focus of attention in WM. Next, we discuss extensions that are necessary to apply the model to sequentially presented memory sets.

### Basic Architecture

The model architecture is illustrated in Figure 1. It combines two feature spaces, one for representing the *contents* of WM and the other for representing their *context*. The contents are the elements of the memory set that participants are asked to remember; they can be visual objects or verbal items such as words or nonwords. The contexts are the representations that the contents are bound to when they are encoded into WM, and which serve as retrieval cues at test. Here we consider the serial position of list elements and their spatial locations as contexts, but in general, every information that serves as unique retrieval cue to a content element is a context for that element.

Both content and context space can have multiple feature dimensions: Visual objects can vary on several feature dimensions such as color, orientation, and size. Verbal items can vary on a large number of phonological and semantic dimensions, but the psychological similarity space of verbal materials is not well known, so that we cannot easily identify its dimensions. Here we consider only a single dimension in content space: Color hue for visual stimuli, and phonological similarity on an ad-hoc constructed similarity scale for verbal stimuli. In context space, the model distinguishes the ordinal serial position of list elements on the temporal axis, and their spatial location on a virtual circle on the screen.

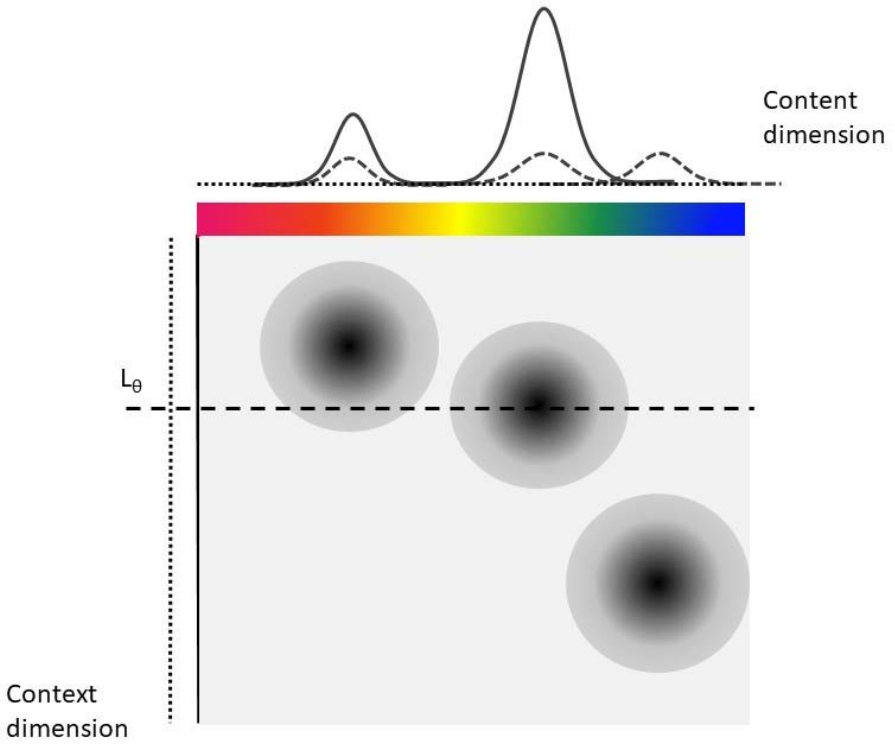


*Figure 1. Schematic of the bindings between a content feature dimension (e.g., colors) and a context dimension (e.g., spatial location) according to the Interference Model. The figure shows the state of the memory system after encoding of three items, one of which is now tested. Each memory feature is bound to its location on the cue-feature dimension through a binding with limited precision in both dimensions, shown here as a two-dimensional distribution in binding space. The target location $L_\vartheta$ acts as a retrieval cue, re-activating all features in proportion to the strength of their binding to $L_\vartheta$, depicted by the darkness level in binding space. The resulting distribution of re-activation is added to the cue-independent activation of the three features, and uniform background noise, to form the distribution of activation A(x) over the content feature dimension.*

We can think of the dimensions in both spaces as populations of neurons with tuning curves centered on different feature values on these dimensions (e.g., neurons with tuning curves centered on a particular color, or a particular spatial location). A content element can be represented by a

distribution of activation over the neurons of the population. Schneegans and Bays (2017) have developed a model with an architecture very similar to the IM, and they specify these hypothetical population codes in detail. The IM is formulated on a more abstract level: Representations of contents and contexts are modeled as distributions over the dimension(s) of their respective spaces. For content dimensions we use a normal distribution because this has been used in all previous models of continuous-reproduction data. In particular, for circular dimensions such as orientations, or colors in a color circle, we use the von-Mises distribution. The precision of content representations is characterized by the dispersion parameter of these distributions. For context dimensions we use a Laplace distribution, which drops exponentially in both directions from its mean. We chose this distribution because it implements the assumption – common in temporal-context models of memory (e.g., Brown, Neath, & Chater, 2007) – that the similarity of context cues drops exponentially with distance.

When an element of a memory list is encoded, each of its content features is bound to each of its context features through a two-dimensional distribution of binding strength, illustrated by the dark blobs in the binding space of Figure 1. The binding strength at a point $(x_i, y_j)$ in binding space is the product of the densities of the context distribution at $x_i$ and of the content distribution at $y_j$. This assumption is taken from the Hebbian learning rule used in connectionist models of serial order (Burgess & Hitch, 1999; Farrell, 2006) that changes the connection weight between context units $x_i$ and content units $y_j$ according to the product of their activation at encoding.

At test, the target element is identified by its context, which could be its spatial location, its serial position in the list, or both. The given context is used as retrieval cue to re-activate the features bound to it. This recreates an approximation to the original distribution of activation in content space corresponding to the target feature. However, in addition to the target feature, other components contribute to the activation distribution generated at retrieval: (1) Because of the limited precision of memory representations on the context dimension, the context cue reactivates not only the target but, to some extent, also those non-targets that are close to the target in context space. (2) The

features of all list elements presented in the current trial are activated to some degree at test, independent of the contextual retrieval cue. This activation could have two sources. It could reflect persistent activation of all to-be-remembered stimuli, as is assumed in many theories of WM (Curtis & D'Esposito, 2003; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Wei, Wang, & Wang, 2012). Alternatively, it could reflect the reactivation of all features that are bound to any location in context space due to unspecific reactivation of contexts, because all elements of the memory list are bound to some context. (3) Because WM is implemented in biological machinery, every representation is contaminated by noise that arises, for instance, from spontaneous firing of neurons involved in a population code. Here we consider two variants of modelling noise, explained below.

Taken together, the activation distribution *A* over the possible feature values *x* that is generated at retrieval in response to a context cue $L_\theta$ can be expressed as a weighted mixture of three components:

$$A(x \mid L_\theta) = cA_c(x \mid L_\theta) + aA_a(x) + bA_b(x) \quad (1)$$

The first component, weighted by parameter *c*, is the component generated through cue-based retrieval. It depends on the location *L* of the target θ in context space that is used as retrieval cue. The second component, weighted by *a*, is the activation of all list elements independent of the cue. The third component, weighted by *b*, is the uniform background noise. The three components are modeled as follows:

The $A_c$ component is a sum of the activation distributions across all *n* elements in the memory set, weighted by how strongly each element is cued through its binding to the context cue:

$$A_c(x \mid L_\theta) = \sum_{i=1}^{n} \exp[-s \cdot D(L_i, L_\theta)] \cdot \mathcal{VM}(x, x_i, \kappa) \quad (2)$$

Each element's weight is an exponentially declining function of the distance *D* in context space between that element's location $L_i$ and the cue location $L_\theta$. The rate *s* of the exponential reflects the precision of the Laplace distribution of context representations in memory. The

element's feature representation is modelled as a von-Mises distribution centered on the element's original feature $x_i$, with precision parameter κ. The von-Mises distribution is a normal distribution on the circle, and has been used in most previous models of feature recall on circular dimensions, starting with Zhang and Luck (2008).

The $A_a$ component is an unweighted distribution of activation of all memory elements:

$$A_a(x) = \sum_{i=1}^{n} \mathcal{VM}(x, x_i, \kappa) \quad (3)$$

We consider two versions of modelling the third component, representing noise. The first is the one we implemented in the original IM. There, noise is represented as uninformative background activity that is added to the representation of each item at encoding. Background noise is is represented by the $A_b$ component of activation, which adds a uniform distribution for every encoded item:

$$A_b(x) = n\mathcal{U} \quad (4)$$

For this variant of implementing noise, we translate the activation distribution generated at retrieval into a response probability through a version of Luce's choice rule that normalizes the activation directly:

$$P(x) = \frac{A(x)}{\sum_{k=1}^{N} A(k)}, \quad (5)$$

with the sum running over all *N* response options. Because of this normalization step – necessary for obtaining a likelihood function for the observed response distributions – the three mixture weights *a*, *b*, and *c* in Equation (1) are not identifiable when freely estimated. Therefore, we fixed *c* = 1 as a scaling parameter, so that *a* and *b* are estimated relative to *c*.

The second variant of modelling noise uses signal-detection theory (SDT), as proposed by Schurgin, Wixted, and Brady (2020). The noise component $A_b$ is modelled as a random variable drawn from a zero-centered noise distribution, added to each possible response *x*. The response option with

the highest activation level after adding noise is chosen as the response. When the noise is assumed to be drawn from a Gumbel distribution with mean = 0 and standard deviation σ, this response-selection process can be modelled by another variant of Luce's choice rule (Yellott, 1977):

$$P(x) = \frac{\exp\left(A(x)/\sigma\right)}{\sum_{k=1}^{N}\exp\left(A(k)/\sigma\right)} \quad (6)$$

Here, the *A(x)* distribution is the weighted sum of only the $A_c$ and the $A_a$ component, because the noise is already incorporated in the normalization function. Whereas in the first variant, the amount of noise (relative to the strength of the memory signal, as represented by $A_c$ and $A_a$) is represented by the weight parameter *b*, in the second variant, it is represented by σ. A recent comparison of measurement models for the continuous-reproduction task (Oberauer, 2021) revealed that models using Luce's choice rule on exponentially transformed memory signals as in Equation (6) fit the data better than models with Luce's choice rule on untransformed signals (Equation (5)). Therefore, we explore whether it results in a better fit also in the context of the extended IM.

To summarize, in the IM the information retrieved from memory at the time of test is a mixture of target-related information, non-target related information, and background noise. Non-target information comes from two sources: One is the activation of all memory contents with equal strength (component $A_a$), and the other is activation generated by the retrieval cue (component $A_c$). The latter source entails a prediction that distinguishes the IM from most other models of visual WM: People's tendency to erroneously report the feature of a non-target should decrease with the non-target's distance from the target in context space. This prediction has been borne out in several studies of visual WM in which spatial location was used as the retrieval cue (Bays, 2016; Emrich & Ferber, 2012; Oberauer & Lin, 2017; Rerko, Oberauer, & Lin, 2014; Tamber-Rosenau, Fintzi, & Marois, 2015), and also when color was used as the retrieval cue, and participants had to reproduce the target's orientation (Oberauer & Lin, 2017). This effect can be seen as a generalization of the transposition gradient typically found in memory of serial order, where erroneous reports of non-

targets (i.e., transposition errors) tend to come from list positions close to the tested position. According to the IM such a gradient is to be expected over any context dimension that is used as retrieval cue; in tests of serial-order memory this happens to be the target's list position.

**The Focus of Attention in Working Memory**

The IM includes the assumption of a focus of attention in WM that can be used to select a subset of elements from the current memory set. This assumption is supported by a substantial body of research on WM for verbal and spatial contents (for a review see Oberauer & Hein, 2012) as well as research on visual WM (for a review see Souza & Oberauer, 2016). We assume that, by default, the focus of attention selects only one content representation and its corresponding context, because superimposing multiple activated contents and contexts entails the risk of confusion (Oberauer & Hein, 2012; Oberauer & Lin, 2017).

In the IM for visual objects we further assume that the focus of attention maintains one memory element with a particularly high precision, $\kappa_f$. If the element in the focus of attention happens to become the target at test, then the content of the focus of attention is added to the activation distribution with weight *c*. Moreover, focusing attention on one element in a memory set increases its relative strength compared to the other elements, either by strengthening the focused element's content-context bindings or by (partially) removing the non-focused elements from WM (Souza & Oberauer, 2016). In the IM the two scenarios – selective strengthening of the focused item or weakening of all others – cannot be distinguished because only relative strength of activation components affects the model's predictions. Therefore, we model the relative strengthening of the target's content-context bindings ($A_c$) when it is in the focus as a proportional reduction of the $A_a$ and $A_b$ components to *A(x)*. To incorporate these effects of having the target in the focus of attention we need to extend Equation (1):

$$
\begin{aligned}
A(x \mid L_\theta, \neg F_\theta) &= cA_c(x \mid L_\theta) + aA_a(x) + bA_b(x). \\
A(x \mid L_\theta, F_\theta) &= cA_c(x \mid L_\theta) + cA_f(x) + rbA_b(x) + raA_a(x). \\
A(x \mid L_\theta) &= \left(1 - P(F_\theta)\right) \cdot A(x \mid L_\theta, \neg F_\theta) + P(F_\theta) \cdot A(x \mid L_\theta, F_\theta).
\end{aligned}
\tag{7}
$$

Here, the first line reflects the case in which the target is not in the focus of attention – the original Equation (1) – and the second line reflects the case in which the target is in the focus. The activation distribution in the focus of attention added in the second line is a normal distribution centered on the target feature with the focus-precision parameter $\kappa_f$.

$$A_f = \mathcal{VM}(x, x_i, \kappa_f) \tag{8}$$

The third line of Equation (7) combines the two cases, weighted with the probability that the target is in the focus of attention. In a typical visual WM task in which the memory set is presented as a simultaneous array, one of its elements is picked at random for being held in the focus, so this probability is 1/*n*. The element selected into the focus of attention can be controlled by a pre-cue (i.e., cueing the most likely target before onset of the memory array) or a retro-cue (presented after array offset). Both pre-cues and retro-cues affect performance in visual WM tasks, improving it when the target is validly cued but decreasing it when the cue is invalid (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Souza & Oberauer, 2016). The IM is able to reproduce these effects in the continuous-reproduction paradigm (Oberauer & Lin, 2017), providing some preliminary support for the assumptions we made about the focus of attention in the model.

A second benefit of holding an item in the focus of attention is to strengthen that item's binding to its context (Rerko & Oberauer, 2013). In the IM we can model this by increasing the *c* parameter relative to *a* and *b* in Equation (1), or equivalently, by reducing *a* and *b*. As we fix *c* to 1 as a scaling parameter, we chose the latter option in the original IM (Oberauer & Lin, 2017), multiplying *a* and *b* with a reduction factor *r* (ranging from 0 to 1) if the target item is in the focus of attention. For the new version of modelling noise, which does not use the *b* parameter, this solution is not suited; therefore, we instead add a free parameter *r* to the representation of the target item in the activation distribution if that item is in the focus of attention.

**Applying the Interference Model to Sequentially Presented Lists**

Presenting a memory set sequentially rather than simultaneously yields richer information because we can look at the effects of serial position of input. Moreover, when we test all elements rather than just one we also obtain information about the effects of output order. To account for these effects, the IM has to be extended by assumptions about serial-position effects, and we next discuss the options we explored. In addition, sequential presentation and testing of all elements has implications for the assumed role of the focus of attention, to which we will turn afterwards.

***Potential Causes of Serial-Position Effects***

In tests of memory for serial order in which the target's input position in the list is used as a retrieval cue, primacy and recency effects emerge to some extent as side effects of the tendency to confuse items with their neighbors (i.e., the transposition gradient): The first list item has no neighbors preceding it with which it could be confused, and the last item has no neighbors following it – this explanation of primacy and recency effects has been referred to as the *edge effect*. However, primacy and recency effects have been observed also when the item's spatial location – uncorrelated with their input position – was used as retrieval cue, and when the item's semantic category – which uniquely identified the target in a list of words – was used as retrieval cue (Jones & Oberauer, 2013). In these experiments the input position was not the retrieval cue given at test, and therefore the edge-effect explanation does not apply in a straightforward manner.

However, it is possible that people still use serial position as retrieval cue indirectly, as follows: At encoding, they bind each list item to its input position, and they bind each item's location (or semantic category) to the same input position. At test, they use the given retrieval cue (location, or semantic category) to retrieve the input position bound to it, and then use that input position as a retrieval cue to access the list item bound to it. Evidence that people occasionally use this mediated retrieval route comes from a study by Cowan, Saults, and Morey (2006), and a model incorporating this form of mediated retrieval accounted well for the data of Jones and Oberauer (2013). The

present experiments provide an opportunity to test this possibility: We presented the memory elements sequentially in randomly chosen spatial locations, and provide only the target's spatial location as retrieval cue at test. To the extent that people use mediated retrieval, we should find a transposition gradient not only over the spatial but also over the temporal context dimension: Targets should tend to be confused with non-targets that are close to them in space, and with non-targets close to them in time (i.e., in their input serial position). Therefore, we included both retrieval paths – the direct path $R_1$ and the mediated path $R_2$ – and modelled the activation strength arising from cue-based retrieval, $A_c$, as a weighted mixture of both components:

$$
\begin{aligned}
A_c(x \mid L_\theta) &= \sum_{i=1}^{n} \left[ w_{direct} R_1 + (1 - w_{direct}) R_2 \right] \cdot \mathcal{VM}(x, x_i, \kappa), \\
R_1 &= \exp[-s_{space} \cdot D(L_i, L_\theta)], \\
R_2 &= \sum_{j=1}^{n} \exp[-s_{space} \cdot D(L_j, L_\theta)] \exp[-s_{time} \cdot D(i, j)].
\end{aligned}
\tag{9}
$$

The direct retrieval path, $R_1$, is as in Equation (2). For the mediated path, $R_2$, we consider every possible input position $j$ that could be retrieved, given spatial position $L_\theta$ as a retrieval cue, and for each input position $j$, we compute the strength with which it cues each item $i$ as a function of the distance between $i$ and $j$ on the input-position dimension. The cueing strength for item $i$ is the weighted sum of the cueing strengths through each input position $j$, weighted by the strength with which position $j$ itself is cued.[2]

We use the weight parameter $w_{direct}$ to estimate the relative weight of the two retrieval routes. If the mediated retrieval route plays an important role in cue-based retrieval, $w_{direct}$ should be estimated to a relatively low value, implying a higher weight for the route mediated through input position. If that is the case, then primacy and recency effects would be predicted to some extent because of the edge effect on the input dimension.

[2] Whereas for the spatial context dimension, the metric distance matters, so that two neighboring items in an array are less confusable when they are spaced further apart (Rerko et al., 2014), this appears not the be the case on the temporal context dimension (Lewandowsky, Brown, Wright, & Nimmo, 2006; Peteranderl & Oberauer, 2018). Therefore, we can set the distance between neighboring input positions to 1, and apply the distance function D(i,j) to the input-position indices directly.

We also considered a second way in which the serial position could be used as a retrieval cue: Participants could scan through the serial positions in forward order and retrieve the location bound to it until the retrieved location matches the target location $L_\theta$ given on the screen. Then they use the current serial position as retrieval cue to access the color bound to it. Again, the model takes a weighted average of the direct and the indirect retrieval path, but the indirect retrieval path $R_2$ now is computed differently:

$$R_2 = \sum_{j=1}^{n} \sum_{i=1}^{n} P(L_\theta \mid j) \exp[-s_{time} \cdot D(i, j)]. \tag{10}$$

The first sum in Equation (10) reflects the scan over the input positions *j*. At each input position, the target location is retrieved with probability $P(L_\theta \mid j)$. If the target location is retrieved, it matches the location given in the test screen, and then input position *j* is used as a retrieval cue for the color. The second sum in Equation (10) describes the strength with which each of the colors *i* of the memory set are cued by the input position *j*. This cueing strength is an exponentially declining function of the positional distance between *j* and *i*. The probability of retrieving the target location, given input position *j* as a cue, is modeled analogous to the probability of retrieving a color: Each location of the array is activated according to the strength with which it is cued by input position *j*, depending on their positional distance. The activation distribution over locations is turned into a probability distribution using whichever version of Luce's choice rule is also used for response selection in the model variant:

$$\begin{aligned} A(L_k \mid j) &= \exp[-s_{time} \cdot D(k, j)]; \\ P(L_\theta \mid j) &= \frac{A(L_\theta \mid j)}{\sum A(L_k \mid j)}, or \\ P(L_\theta \mid j) &= \frac{\exp\left(A(L_\theta \mid j)/\sigma\right)}{\sum \exp(A(L_k \mid j)/\sigma)}. \end{aligned} \tag{11}$$

In initial explorations of the model, we found that the edge effect predicts primacy and recency effects that were much too small to account for those actually observed. Therefore, we

considered additional causes of primacy and recency. One mechanism often included in models of serial-order memory is a primacy gradient of memory strength: As items are presented sequentially, they are encoded with decreasing strength (Grossberg & Stone, 1986; Page & Norris, 1998). This assumption has received empirical support from studies of serial-position effects (Oberauer, 2003) and transposition latencies (Farrell & Lewandowsky, 2004).

The primacy gradient of encoding strength is probably under the person's control, as suggested by a study of Palladino and Jarrold (2008): They tested memory for serial order in two paradigms: Standard serial recall in which the entire list has to be repeated in forward order, and running memory, in which a list of unpredictable length is presented and participants are asked to recall the last *n* items in forward order. Palladino and Jarrold compared serial-position curves from standard serial recall of lists of length *n* with those from running span tests in which the last *n* items had to be recalled, and the presented list happened to be only *n* items long. The only difference between these conditions is the person's foreknowledge about the start of the to-be-recalled series: In the standard task they knew at encoding that they will have to start recall with the first-presented item; in the running-memory task they could not know this. This difference in expectation led to substantially increased primacy effects in the standard task compared to the running-memory task. This finding can be explained by assuming that, when people know the start of the list they will have to recall, they boost the strength of encoding of the first list item, and gradually decrease that boost for subsequent items, thereby generating a primacy gradient of memory strength that helps them to "peel off" the items in the correct order through competitive cueing. Further evidence comes from a study comparing serial-position curves for intentionally and incidentally encoded lists. The primacy effect was substantially reduced when lists were encoded incidentally (Oberauer & Greve, 2021).

As a mechanism for causing the recency effect we considered a process of obligatory WM updating: Whenever a new element is encoded into WM, the strength of all existing WM contents is decreased by a constant proportion $\delta$. This updating mechanism ensures that WM is not overloaded, thereby keeping interference between representations at bay. Its rationale could be described as the

system's default assumption that the most recent information encoded into WM is the most relevant, and older information is increasingly less relevant. Arguably, this is the case in most situations in which we use WM: As our train of thought or our exploration of the environment progresses, earlier information becomes less and less relevant for our current activity. Holding on to too many past events would soon result in a forbidding amount of proactive interference that prevents the system from holding precise and accessible representations of the most recent relevant events. To keep available the most relevant information at any time for an organism that is constantly active – both physically and mentally – requires that WM is seamlessly and rapidly updated. An obligatory updating mechanism as envisioned here would be well suited to ensure that functionality. Its downside is that it jeopardizes access to events that remain relevant even after a number of subsequent events were encoded into WM – as, for instance, the initial items of a memory list. The encoding boost to list-initial items that generates the primacy gradient is a possible mechanism to manage that risk: When the person knows that an event will have to be maintained for longer than usual and survive updating, they give its representation in WM a boost in strength.

The updating mechanism also offers itself as a cause of output interference. When multiple items from WM are tested and the order of testing is not confounded with that of presentation, it is usually found that performance declines monotonically with output position (Cowan, Saults, Elliott, & Moreno, 2002; Oberauer, 2003; Peters et al., 2018). This can be explained by assuming that WM is updated not only when a new event is encoded but also when an event is retrieved: With each output position the remaining WM contents are reduced in strength by factor δ.

Put together, the primacy boost and the updating mechanism yield the following expression for the strength of content-context bindings as a function of input position *i* and output position *o*.

$$C_{i,o} = (1 + \beta\pi^{i-1})\delta^{n-i+o-1}. \quad (12)$$

Here, β is the size of the primacy boost for the first-encoded item, added to a baseline strength of 1, and $\pi$ is the proportional decline of the primacy boost with each successive input

position; δ is the proportional decrement of WM contents at each updating step. Figure 2 illustrates the joint effects of the primacy gradient and the proportional decrement of strength on the resulting binding strength $C_{i,o}$ across input positions, for several values of the parameters π and δ.

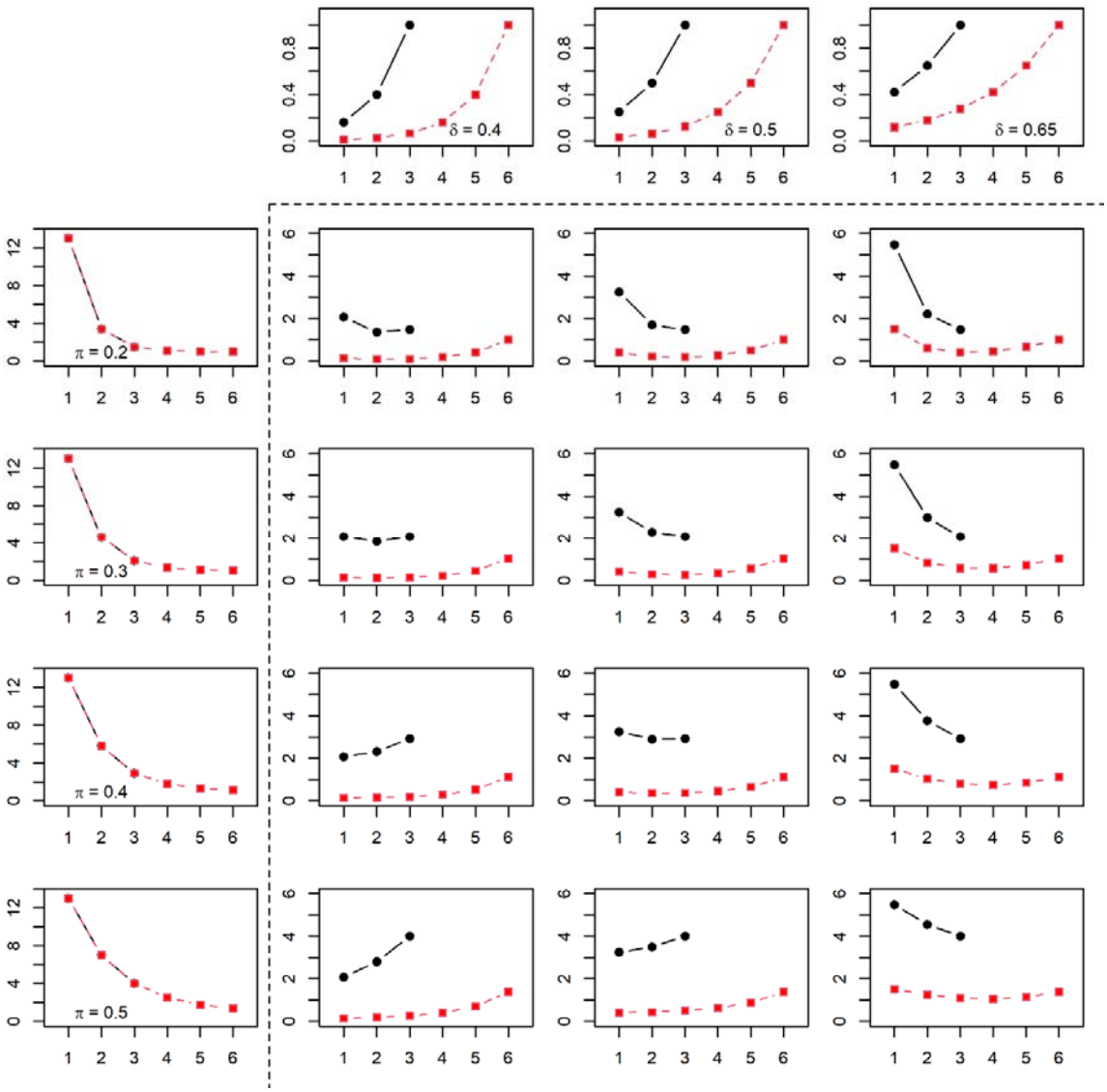


*Figure 2: Values of binding strength $C_{i,o}$ resulting from three values each for parameters π and δ, for set size 3 (black) and set size 6 (red), with β=12, evaluated at the first output position (o=1), and input position on the x-axis. The marginal panels show the separate effect of the primacy gradient governed by parameters π and β, and of the recency gradient governed by δ, respectively. The interior panels show the resulting $C_{i,1}$ values for each combination that arise from multiplying the primacy and the recency gradients.*

We assume that the two mechanisms causing serial-position effects act only on the strength of content-context bindings – that is, on component $A_c$ of the activation distribution – because previous research on serial-position effects has shown that these effects can be explained well with a model in which only binding strength, not the strength of individual content elements, changes with serial position (Oberauer, 2008). Hence, to incorporate serial-position effects into the IM we expand the first line of Equation (9) as:

$$A_c(x \mid L_{\theta,o}) = \sum_{i=1}^{n} C_{i,o} \left( w_{space} R_1 + (1 - w_{space}) R_2 \right) \mathcal{VM}(x, x_i, \kappa) \tag{13}$$

with $R_1$ and $R_2$ for the cueing strengths along the two retrieval routes as defined in Equation (9).

Whereas we motivated the δ parameter that generates the recency gradient as reflecting automatic updating of WM, we note that there are two other potential mechanisms causing a recency gradient of cueing strength that lead to the same formalization. One is the assumption that each item in a list is cued by two context cues – the cue given at retrieval (e.g., its spatial location), and the current temporal context at the time of test. With each event, the temporal context changes at a constant rate δ, so that with increasing temporal distance between encoding and retrieval of an item (i.e., with more intervening events), the overlap between the item's temporal context and the temporal context at retrieval decreases. As a consequence, the strength of the temporal-context cues declines exponentially with distance, generating a recency gradient. The second alternative interpretation of the δ parameter is that, again, each item is cued by two context cues: The cue given at retrieval, and an end marker that represents the end of the current episode (an idea inspired by the Start-End Model of Henson, 1998). As the episode – for instance the presentation of a list of items – unfolds, the person does not know when it will end, so each new event (i.e., each item) is bound to the end marker with constant strength. With each new event, the binding of all previous events to the end marker is proportionally decreased by δ. In this way, whenever the episode ends,

there is a recency gradient of binding strength of each event in the episode to the end marker. In Appendix A we show that both ideas, when expressed formally, are mathematically equivalent to Equation (12).

***Implications of Sequential Presentation and Multiple Tests for the Focus of Attention***

When list elements are presented sequentially and then tested in a random order, it is still conceivable that one element is selected at random for being held in the focus of attention, and maintained until it is tested. That assumption, however, would entail that the focus of attention in WM is neither involved in encoding further items nor in retrieving them. We find this very implausible, and therefore consider a last-in, first-out (LiFo) constraint on the use of the focus of attention: After list presentation the focus of attention maintains the last-presented item (McElree, 2006). Once an item is cued for retrieval, the focus of attention selects that item and drops any previously-held item. It follows that if, and only if, the last-presented item is tested first, the high-precision representation in the focus of attention can be used for improving performance. This LiFo constraint implies that $F_\theta$ in Equation (6) is true for $i$=$n$ and $o$=1, and false otherwise.

## Overview of Experiments

We carried out two experiments with sequentially presented memory lists, testing memory with the continuous-reproduction paradigm. Experiment 1 used colors as memoranda and Experiment 2 used nonwords. Set size was varied from $n$ = 1 to 6. The testing method and experimental design of the two experiments was chosen to be as similar as possible for investigating WM for simple visual features and WM for verbal stimuli. The sample sizes were determined as a compromise between desirability (i.e., more data points yield more precise estimates, providing better information for model selection and parameter estimation) and economic constraints.

Both experiments were carried out in accordance with the ethical guidelines of the Ethics Committee of the Faculty of Arts and Sciences at the University of Zurich. The experiments were

exempt from formal ethics approval because they did not involve any risk as described on the checklist of the Ethics Committee.

## Experiment 1: Sequential Presentation of Colors

### Methods

*Participants.* Participants were 43 students of the University of Zurich, who took part in exchange for partial course credit or 15 Swiss Francs (~15 USD) per hour as reimbursement. Two of them completed only one session; the others completed two one-hour sessions.

*Design.* Set size *n* was varied from 1 to 6, with 70 trials per set size, administered in a random order.

*Materials.* Memory stimuli were colored squares (40 x 40 pixels). For each trial *n* colors were drawn at random without replacement from 360 colors equally spaced on a circle in the CIE L*a*b* color space, with a radius of 60 centered at ($a = 20, b = 38$), and luminance set to 70. For each trial, *n* locations were drawn at random from 13 equidistant locations on a virtual circle centered on the screen center, with a radius equal to ¼ of the screen height.

*Procedure.* Each trial started with a central fixation cross, followed 0.5 s later by the first colored square. Each stimulus was presented for 0.5 s, immediately followed by the next one until all *n* colored square have been displayed. We chose this presentation rate because short-term consolidation of simple visual stimuli is largely complete after this time (Jolicoeur & Dell'Acqua, 1998; Ricker & Hardman, 2017). Each stimulus was presented in a different location, chosen at random from 13 equidistant locations on a virtual circle centered on the screen center. To prevent verbal labeling of the colors, participants repeatedly spoke the syllable "bi" aloud during presentation.

Testing commenced after a 1 s retention interval. To measure the effects of input serial position (i.e., the order of presentation) independently of the effects of output position (i.e., the order of recall), we tested the items in a new random order on every trial, so that input and output position are uncorrelated over trials (Oberauer, 2003). At the beginning of the test episode, empty frames were presented in the locations of all array elements, and a color wheel was displayed,

centered on the screen center, with radius of ¼ of the screen height. For each tested stimulus, the frame around its location was made thicker than the others. Participants were asked to select the color of the probed location by left-clicking one of the colors on the color wheel. The next target location was highlighted 0.3 s after the preceding response. After the last response the screen went blank for 2.5 s before the next trial started. For each trial the color wheel was rotated into a new random orientation; its orientation remained constant throughout the testing episode of that trial.

**Results**

*Descriptive Results*

Performance was measured by the reproduction error, defined as the mean absolute deviation of the response from the target color in degrees. Figure 3 presents the mean errors for all combinations of input and output positions at the six set sizes. The data show the familiar increase of error with set size. Errors increased over successive output positions, in particular from the first to the second output position – this trend is clearer in Figure 5 below. Similar output-position effects in continuous reproduction of visual features have been observed before (Adam, Vogel, & Awh, 2017; Peters et al., 2018).

There was a primacy effect and, at the larger set sizes, a recency effect over input position. The largely symmetric serial-position curves at the larger set sizes agrees well with the serial-position curve observed with verbal materials when memory for all items is tested in a random order (Jones & Oberauer, 2013; Oberauer, 2003). It does not agree, however, with previous studies of serial-position effects with simple visual stimuli, which showed strong recency effects with little, if any primacy (Gorgoraptis, Catalao, Bays, & Husain, 2011; Kool, Conway, & Turk-Browne, 2014; Pertzov & Husain, 2014). In those studies, only a single randomly chosen item was tested. In our data, the serial-position curves for the first output position – most equivalent to a single test – showed somewhat stronger recency effects than those for later output positions, but still a marked primacy effect. We have no convincing explanation for this discrepancy between studies. In the context of our model we

can speculate that the expectation of multiple tests has motivated participants to give initial list items a stronger primacy boost than in other experiments, because they knew that these items might have to survive the updating decrement not only of several subsequent list items but also several retrieval events.

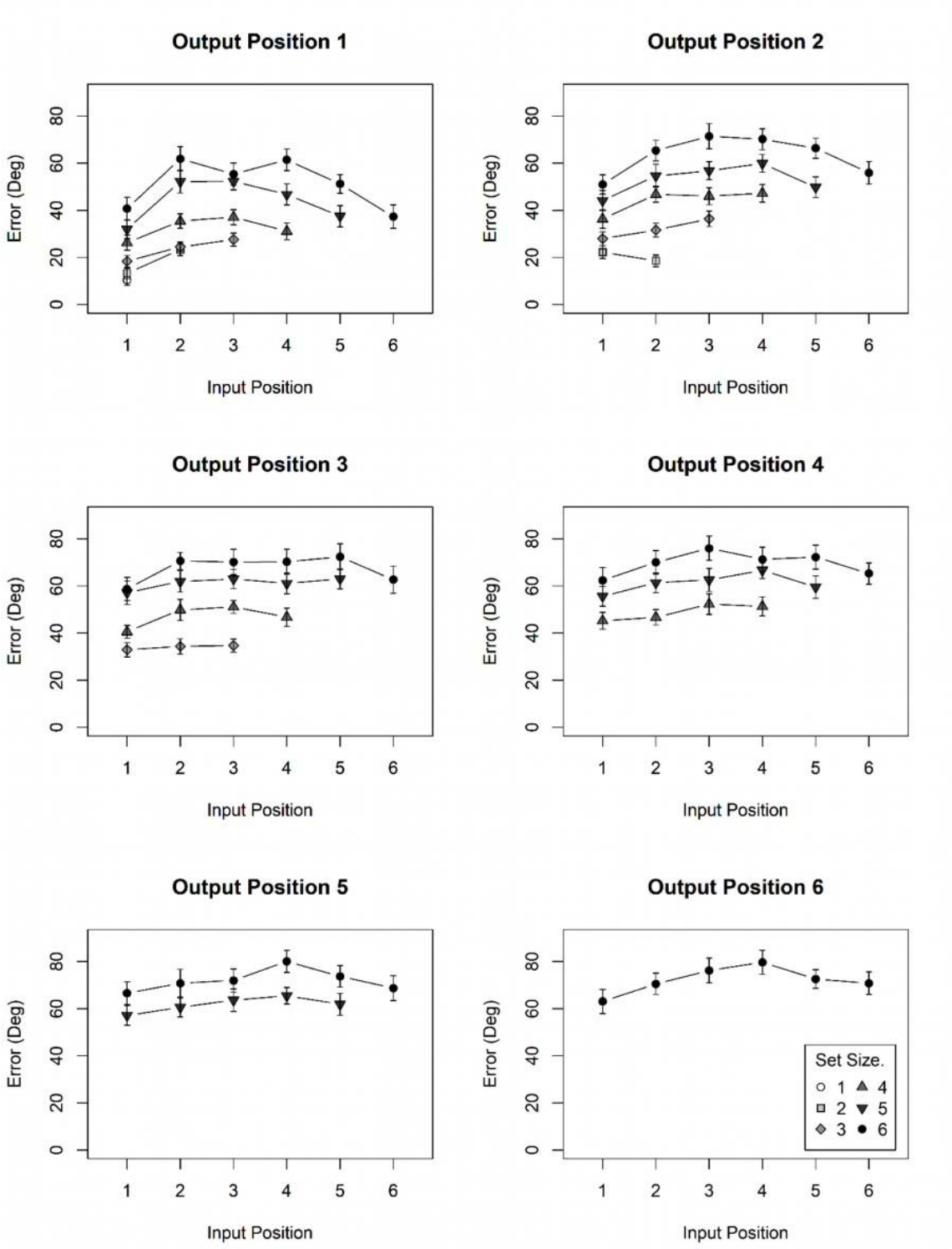


*Figure 3: Observed mean errors in Experiment 1 as a function of input position and set size, for each output position. Error bars are 95% confidence-intervals for within-subject comparisons (Bakeman & McArthur, 1996).*

*Applying the Interference Model*

As foreshadowed, we compared several variants of the extended IM. The main division is between variants implementing noise as background activation as in the original IM, and variants implementing noise in an SDT framework; these variants correspond to the two versions of Luce's choice rule.

Within each of these divisions, we investigated combinations of a direct retrieval route – using the target's spatial location to retrieve its feature through direct feature-location bindings – and two indirect retrieval routes. The first, which we refer to as *mediated*, uses the target's location as retrieval cue to its input position, and then the input position as cue to the location bound to it. The second, which we refer to as *scanning*, scans through the input positions in forward order, retrieving the location bound to each input position in turn, until it finds the given target location. Then it uses that input position as cue to retrieve the feature bound to it. We compared the models with two retrieval routes to models using only the direct, or only the mediated route (i.e., constraining the $w_{direct}$ parameter to 1, or to 0, respectively).

Due to the LiFo constraint on using the high-precision representation in the focus of attention, the focus of attention contributes to performance only when the last-presented item is tested first. This limited contribution raises the question whether the assumption of such a contribution has sufficient support in the data to warrant the extra model complexity from two further free parameters. To test this, we ran *no-focus* variants that eliminate the high-precision representation in the focus of attention (parameter $\kappa_f$), as well as the boost of memory strength for target items in the focus (parameter $r$).

Finally, we considered model variants with stronger constraints on noise. In the original IM, we have assumed that uniform background noise is mixed in with the population code of each encoded stimulus, and as such the background-noise component $A_b$ increases with set size. Alternatively, background noise could arise in the WM system at the time of test, and then it would

not be expected to increase with set size. To test this possibility, we ran a *fixed-noise* model variant in which the background noise component $A_b$ remained constant across set sizes. For the new implementation of noise in the SDT framework, we also tried a constraint on the noise, in which we fixed σ, the standard deviation of the noise distribution, to 1.

We initially tried to fit the IM to both experiments in a hierarchical Bayesian framework, using Jags (Plummer, 2016), and also using Stan (Stan-Development-Team, 2021), but encountered persistent problems with not converging MCMC chains. Therefore, we eventually settled for fitting the model separately to the data of each individual in a maximum-likelihood framework. To minimize the risk of being stuck in a local minimum of the error surface, we ran the model 10 times with different starting values. This was done once with the Nelder-Mead (Simplex) algorithm using the *nmkb* function of the *dfoptim* package for R, and once with the L-BFGS-B algorithm using the *optim* function in basic R.[3] Both algorithms minimized the Deviance = -2 x log-Likelihood as an indicator of badness of fit. We kept the best fit of the 20 runs.

For model selection we computed AIC and BIC, two fit indicators that combine the Deviance with a penalty term for model flexibility due to the number of free parameters:

$$AIC = -2\log(L) + 2N_{par}$$
$$BIC = -2\log(L) + N_{par}\log(N_{obs})$$

Model recovery simulations with the IM and other models of the continuous-reproduction task revealed AIC to perform better than BIC on this family of models (Oberauer & Lin, 2017; van den Berg & Ma, 2014). Therefore, in case of discrepancy, we rely on AIC for model selection.

*Modeling Results*

[3] For the 1st run we set the starting values by hand to plausible values. For the 2nd to 10th run we drew starting values from normal distributions with the mean set to the starting values for the 1st run, and SD set to 0.25 times the starting values for the 1st run, with the constraint that the values don't exceed the parameter boundaries.

As the absolute values of Deviance-based fit indices are uninformative, we computed the differences in fit between each model and the best-fitting model, separately for AIC and BIC. Table 1 shows these differences, averaged across participants.

*Table 1: Mean AIC and BIC Differences of All Model Variants to the Best-Fitting Variant, Experiment 1*

| Sub-Variant | ΔAIC | | ΔBIC | |
|---|---|---|---|---|
| | Uniform background noise | SDT Noise | Uniform background noise | SDT Noise |
| Direct and mediated retrieval | 57.5 | 2.9 | 67.7 | 13.2 |
| Direct retrieval only | 43.1 | 2.1 | 48.1 | 7.1 |
| Mediated retrieval only | 49.1 | 12.2 | 48.8 | 11.9 |
| Direct and Scanning | 42.6 | 0 | 52.8 | 10.2 |
| Direct and mediated, no focus | 77.2 | 2.5 | 76.9 | 2.2 |
| Direct and Scanning, no focus | 61.6 | 0.3 | 61.3 | 0 |
| Direct and mediated, fixed noise | 61.6 | 8.2 | 71.9 | 13.2 |
| Direct and Scanning, fixed noise | 46.1 | 9.6 | 56.3 | 14.5 |

*Note: For each fit index (AIC or BIC), the best of the 16 model variant is identified, and its fit index subtracted from those of all variants. That difference was divided by the number of subjects.*

The model versions with SDT noise generally fared better than the versions with uniform background noise. This observation converges with the outcome of a comparison of measurement models for continuous reproduction (Oberauer, 2021), extending it to explanatory models. The best model was one that combined the direct retrieval route (from the target's spatial location to its color) with access by scanning through input positions to find the target's location, then retrieving

the color bound to that input position. However, a model combining direct with mediated retrieval did not fit much worse. Both models have in common that they combine the direct retrieval route with one that uses the input position as a retrieval cue to link target location and color. Whereas by the AIC, the model including the focus of attention fared best, according to the BIC, the balance between parsimony and fit was better for model versions without a focus of attention, implying that the addition of a last-in, first-out focus of attention does not contribute substantially to explaining the present data.

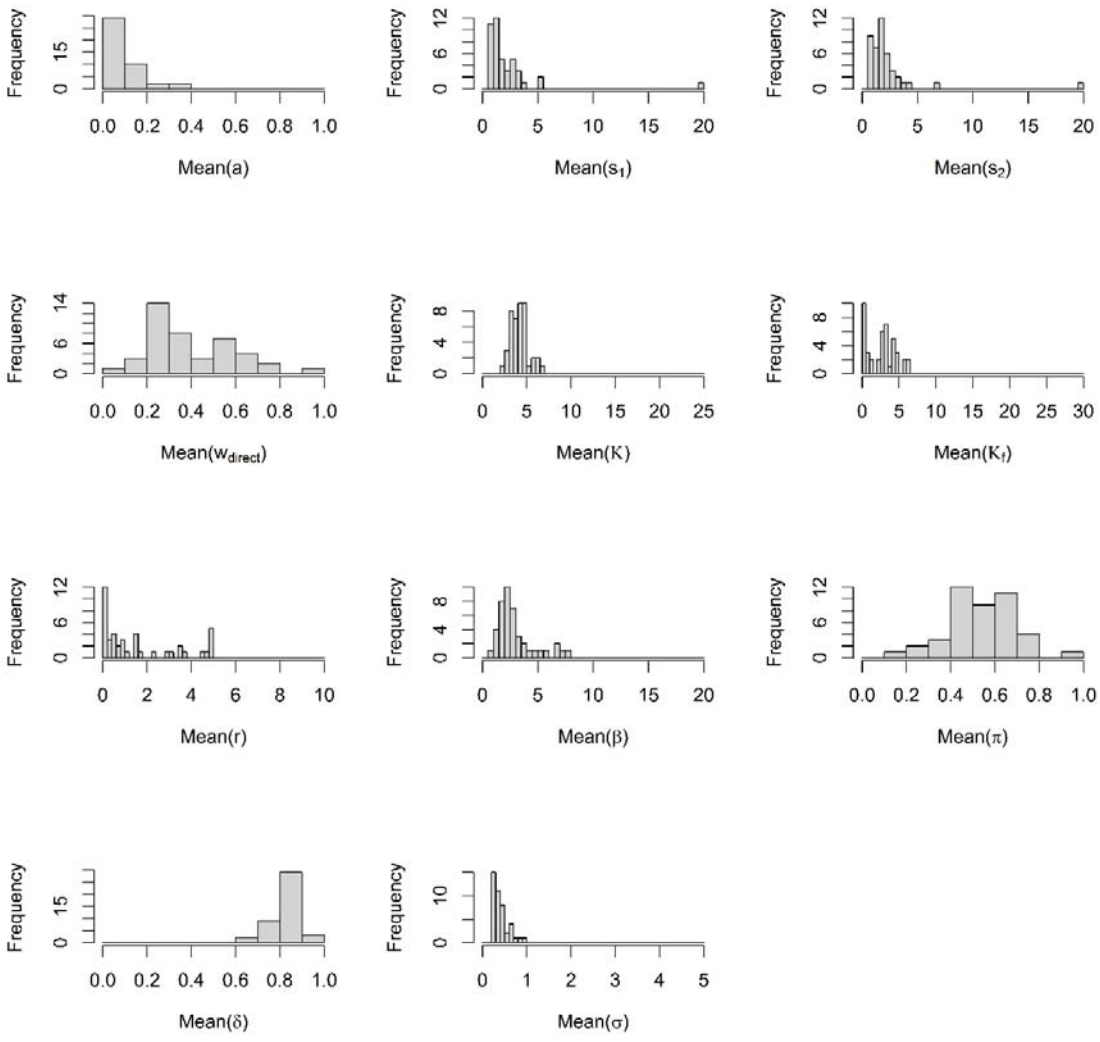


*Figure 4: Distribution of best-fitting parameter estimates over participants for the best-fitting model in Experiment 1.*

Figure 4 shows the distribution of best-fitting parameter estimates across participants. Figure 5 shows the model predictions for the set-size effect, averaged over output positions. Figure 6 presents the model predictions for the serial-position curves over input position. The model reproduced the trends in the data: A steep, approximately linear increase of error with set size; inverse-U shaped input serial-position curves, and a monotonic increase of error with output position. Figure 7 shows the error distributions in feature space for all combinations of input and output positions at set size 4.

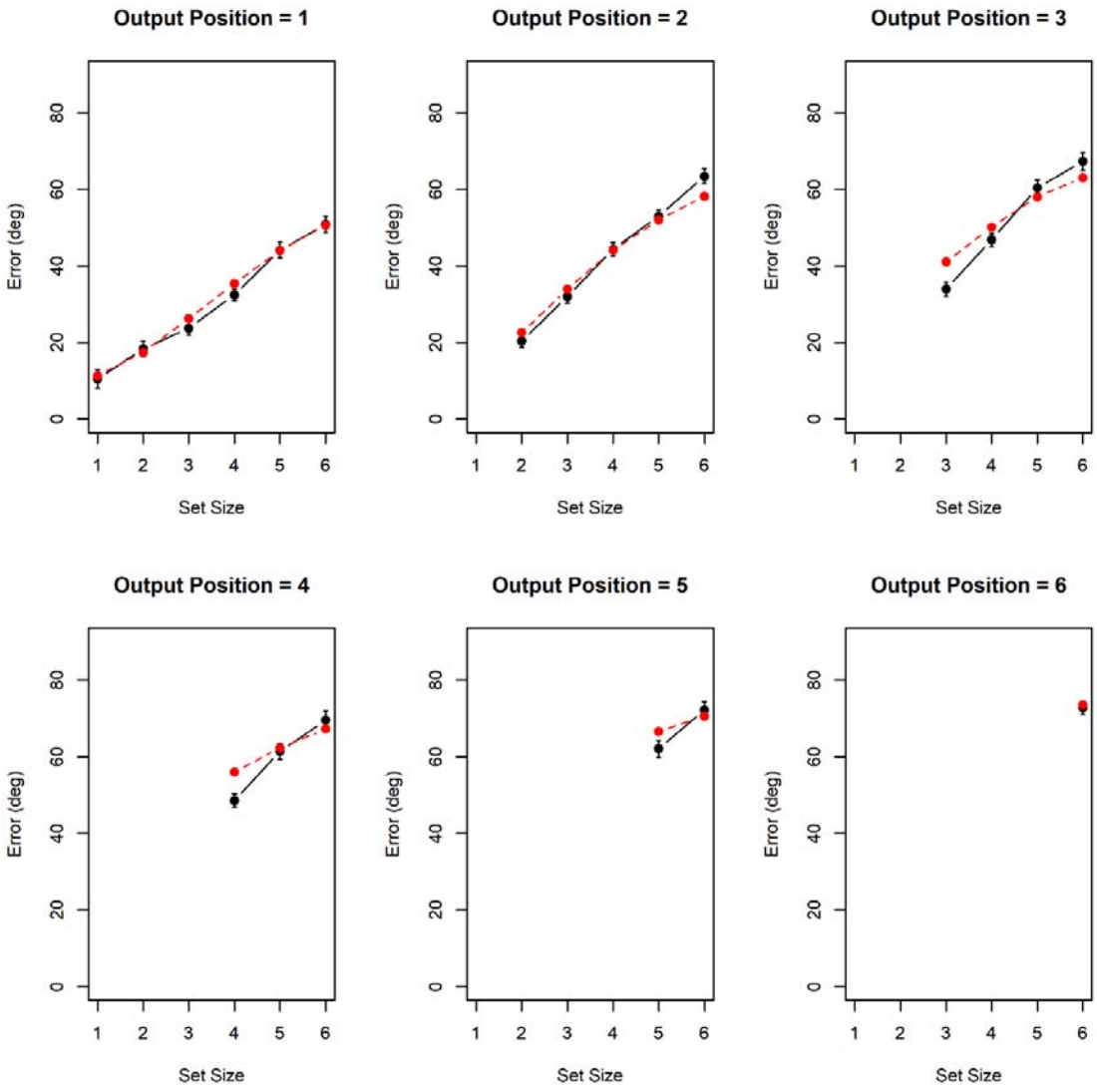


*Figure 5: Observed mean errors by set size at each output position (black) together with predictions from the best-fitting model (red), Experiment 1. Error bars are 95% confidence intervals for within-subjects comparisons.*

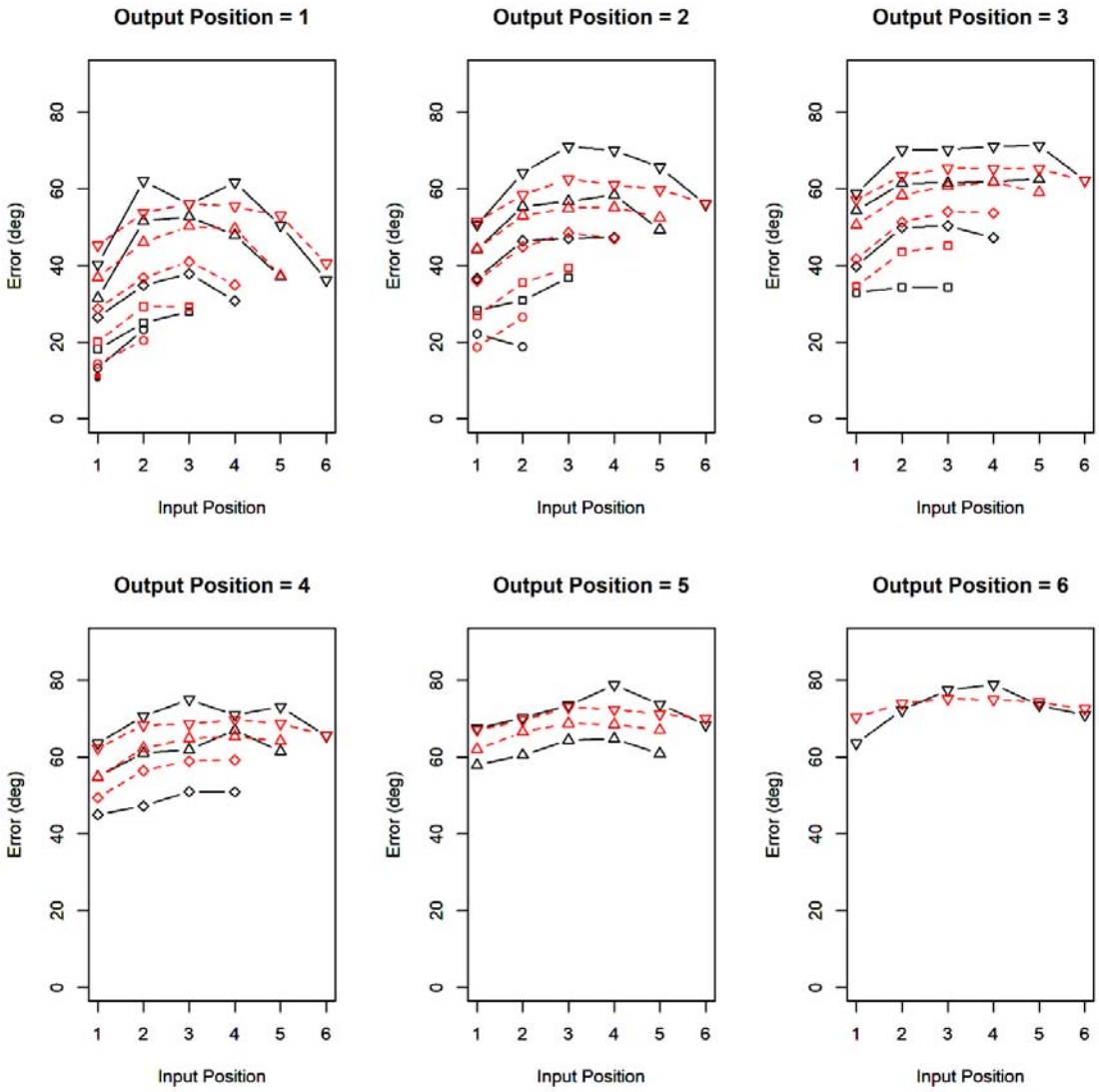


*Figure 6: Observed mean errors by input serial position and set size for each output position (black) together with predictions from the best-fitting model (red), Experiment 1.*

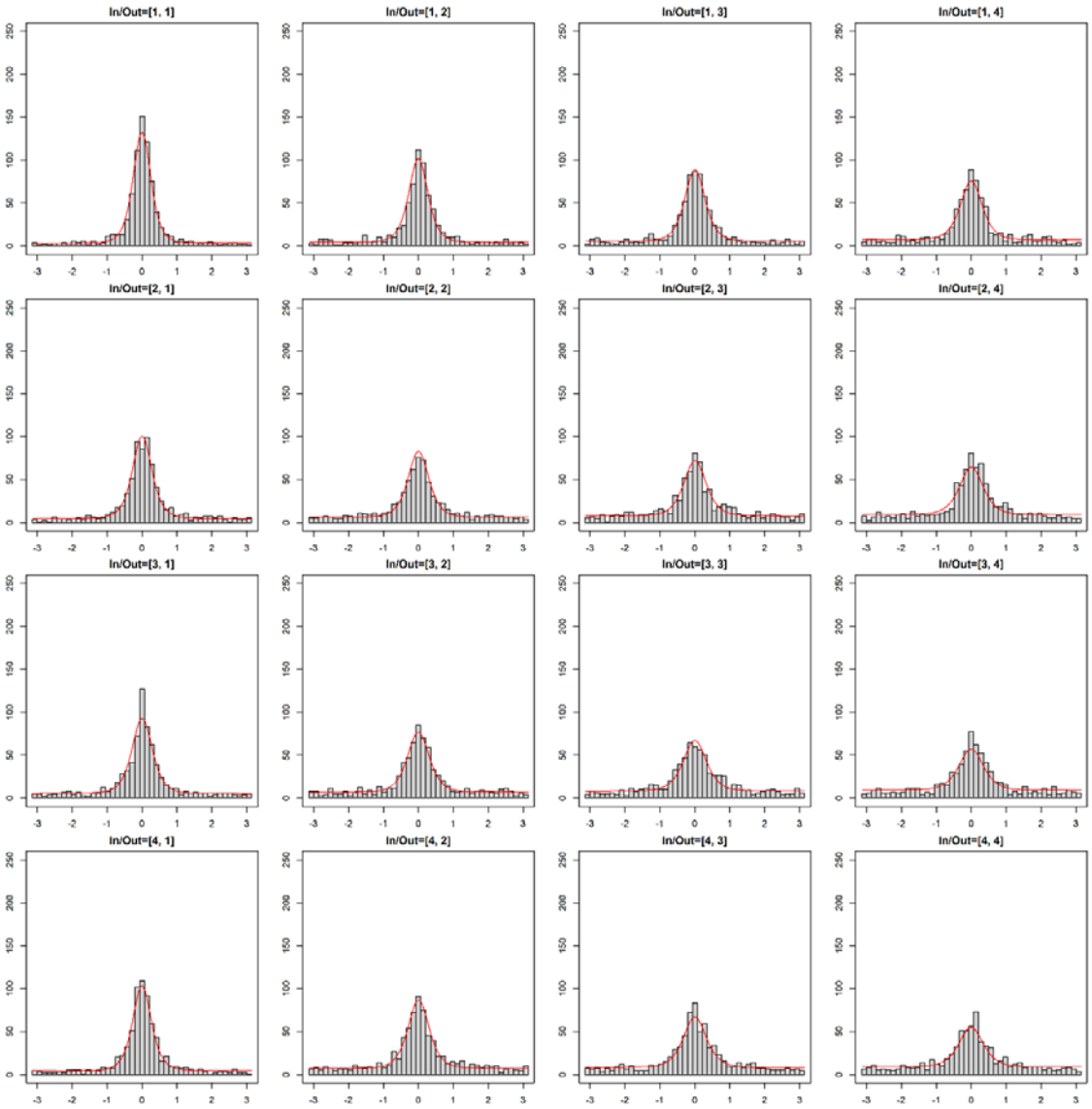


*Figure 7: Observed error distributions (black) and predictions of the best-fitting model (red) for all combinations of input and output position with set size 4 in Experiment 1.*

The distribution of $w_{direct}$ (Figure 4) implies that participants differed substantially in their preference for the direct retrieval route from spatial location to color, and the indirect route through input serial position. This is reflected in the transposition gradients over spatial and over input distance, shown in Figure 8. The transposition gradients show the probability of reporting an item as a function of its distance to the target item. To determine which item was reported, we selected for each response the array item that the chosen color was most similar to (unless the response differed by more than 30 degrees from the most similar item, in which case we classified that response as an

extra-array response).[4] The transposition gradients over spatial distance show a clear decline, as predicted from the use of spatial retrieval cues and limited memory precision on the spatial dimension. The transposition gradients over input-position distance also show a decline, as predicted from using input position as a retrieval cue – as assumed for the scanning route – together with limited precision on the input-position dimension.

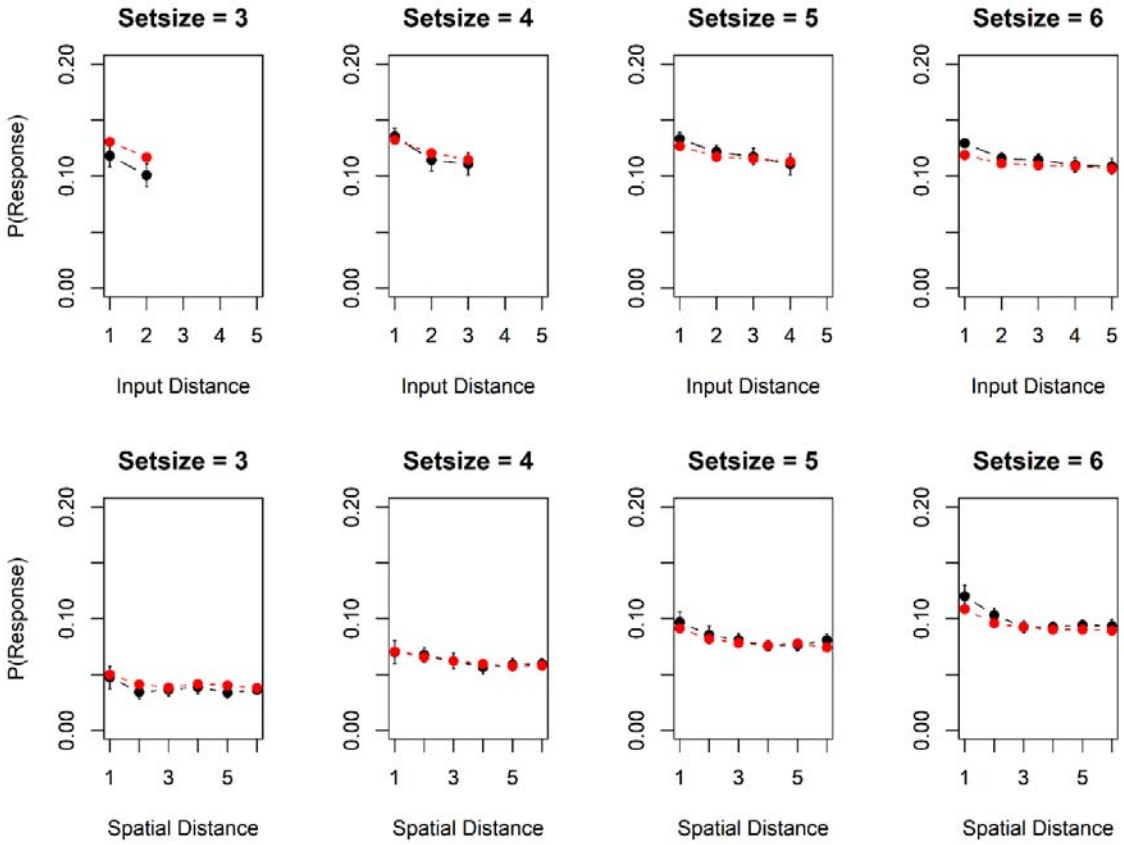


*Figure 8: Transposition gradients, showing the probability of erroneously recalling the color of a non-target item as a function of the distance between the target and the non-target along the dimension of input serial positions (top row), and their spatial distance in the array (bottom row). Observed data are in black; model predictions in red. Spatial distance is measured in units of 1/13 of the circular*

[4] For input distance, the probabilities were averaged over all possibilities of realizing a given input distance (e.g., distance 2 could be realized by recalling item 1 in position 3, recalling item 2 in position 4, and so on), thereby controlling for the fact that shorter distances have more ways of being realized.

*arrays, corresponding to the 13 equidistant place holders for colors in an invisible circle that were used in the experiment.*

## Experiment 2: Sequential Presentation of Nonwords

With Experiment 2 we wanted to measure the precision of verbal representations in WM through a continuous-reproduction procedure. To that end we tried to build a continuous similarity scale for verbal materials. Some previous research took steps in this direction by varying individual vowels (Joseph et al., 2015) or consonants (Hepner & Nozari, 2019) on a continuous scale varying one phonemic feature. The limitation of these similarity scales is that they are extremely inhomogeneous: A few points on the scale correspond to vowels or consonants in the language, whereas the points in between do not represent any linguistic units, and this leads to strong response biases towards the linguistic category centers (Hepner & Nozari, 2019). A further limitation is that these similarity scales cannot be used to measure the similarity between two verbal stimuli that differ in more than one phonemic feature. We therefore built an approximately continuous similarity scale of stimuli that are all linguistically legitimate units of the participants' language: Sets of 19 nonwords were arranged in a closed loop of similarity relations, such that neighboring nonwords differed in a minimal number of phonemic features from each other. In this way, the set of nonwords densely covers a hypothetical continuous circular dimension of phonological similarity, analogous to the dense coverage of the continuous color-hue dimension by the 360-color scale used in Experiment 1.

### Methods

*Participants*

Participants were 22 students of the University of Zurich, who took part in exchange for partial course credit or 15 Swiss Francs (~15 USD) per hour as reimbursement, for three one-hour sessions.

*Design*

Set size *n* was varied from 1 to 6. To achieve a constant number of 15 observations (i.e., responses) in each cell of the input-position x output-position matrix, the number of trials per set size increased with set size: 15, 30, 45, 60, 75, and 90 for *n* = 1 to 6, respectively. The trials were administered in a random order.

*Materials*

We constructed sets of 19 nonwords with a CVCV structure (C for consonant, V for vowel), composed to form a closed loop of of similarity relations: Each CVCV differed from its neighbors in only one consonant, such that the difference between them was only one or at best two phonemic features (see Figure 9 for an example). Phonemic features were determined according to the analysis of features for German consonants by Wiese (2000). We constructed 12 sets of 19 nonwords each, using different vowels and consonants in each set.

*Procedure*

For each trial *n* nonwords were sampled from one set of 19 nonwords that formed a similarity scale, where *n* is the set size. The nonwords were presented in frames in an approximately circular arrangement centered on the screen center. Each nonword of a trial was allocated to a different frame, chosen at random from the 13 frames. The nonwords were presented in a random order, each for 1 s. Testing commenced 1 s after offset of the final nonword. At test, the response set of 19 nonwords was presented in an approximately circular arrangement centered on the screen center, ordered according to their similarity structure, analogous to the color wheel. All n nonwords were tested in a random order. The frame of the tested nonword was highlighted, and participants selected the nonword they remembered for that frame by clicking on it with the mouse in the surrounding set of 19 candidates. 0.3 s after each response, the next target location was highlighted. After the last response, the screen went blank for 0.5 s before the next trial started. Participants completed 105 trials in each of three sessions, preceded by 10 practice trials per session.

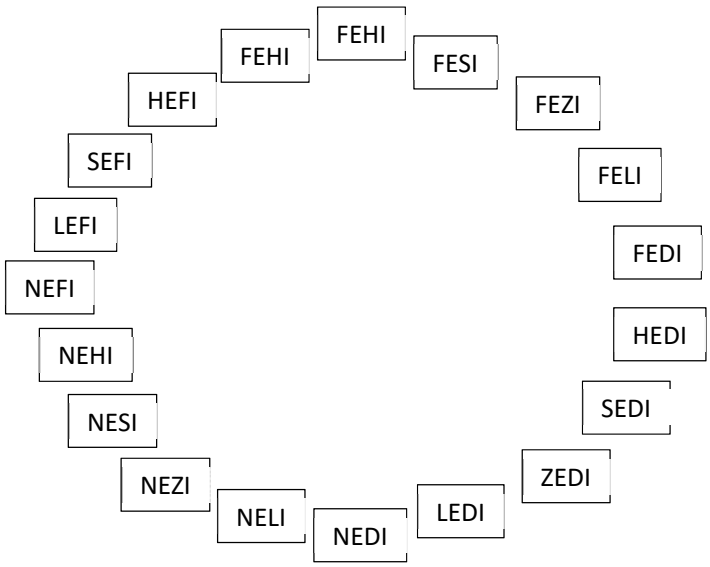


*Figure 9: Example of a circle of phonologically similar non-words constructed for Experiment 2.*

**Results**

*Descriptive Results*

We measured performance in two ways: (1) the reproduction error, defined as the mean absolute angular deviation of the response from the correct nonword on the phonological similarity scale, in degrees, and (2) The proportion of correct nonwords chosen. Figure 10 shows the average error by set size, input position, and output position. Figure 11 plots the distributions of errors for each set size. The distributions show that our attempt to vary phonological similarity continuously did not work as well as we had hoped. The error probability did not fall off gradually with increasing angular error, but rather fell off steeply from zero (i.e., the correct response), and then spread out nearly equally over all error levels, with only a very shallow decline of frequencies with increasing angular error. As a consequence, the continuous error measure contained little information over and above the categorical proportion-correct score, which is plotted in Figure 12.

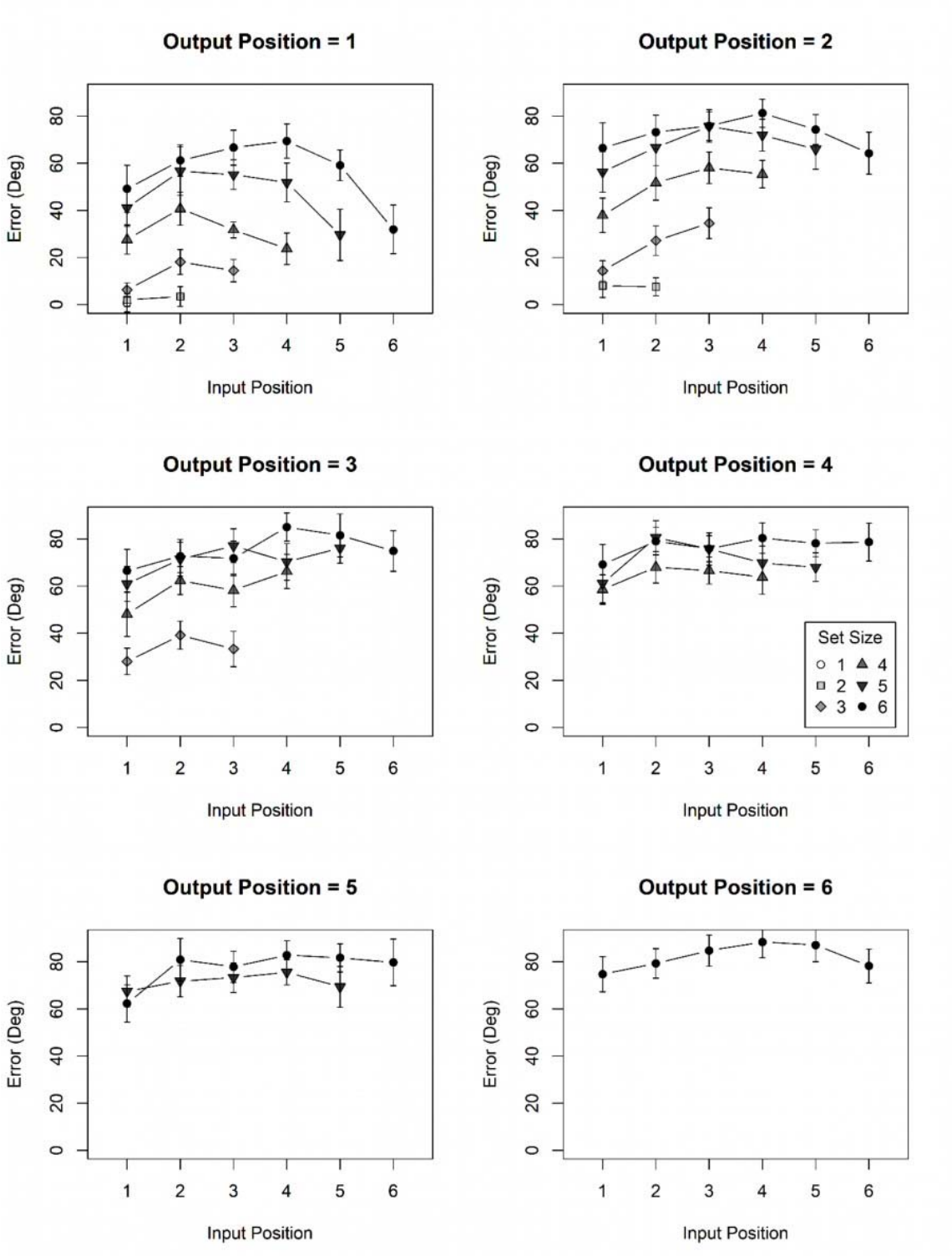


*Figure 10: Mean errors of reproduction in Experiment 2 as a function of set size and input position.*

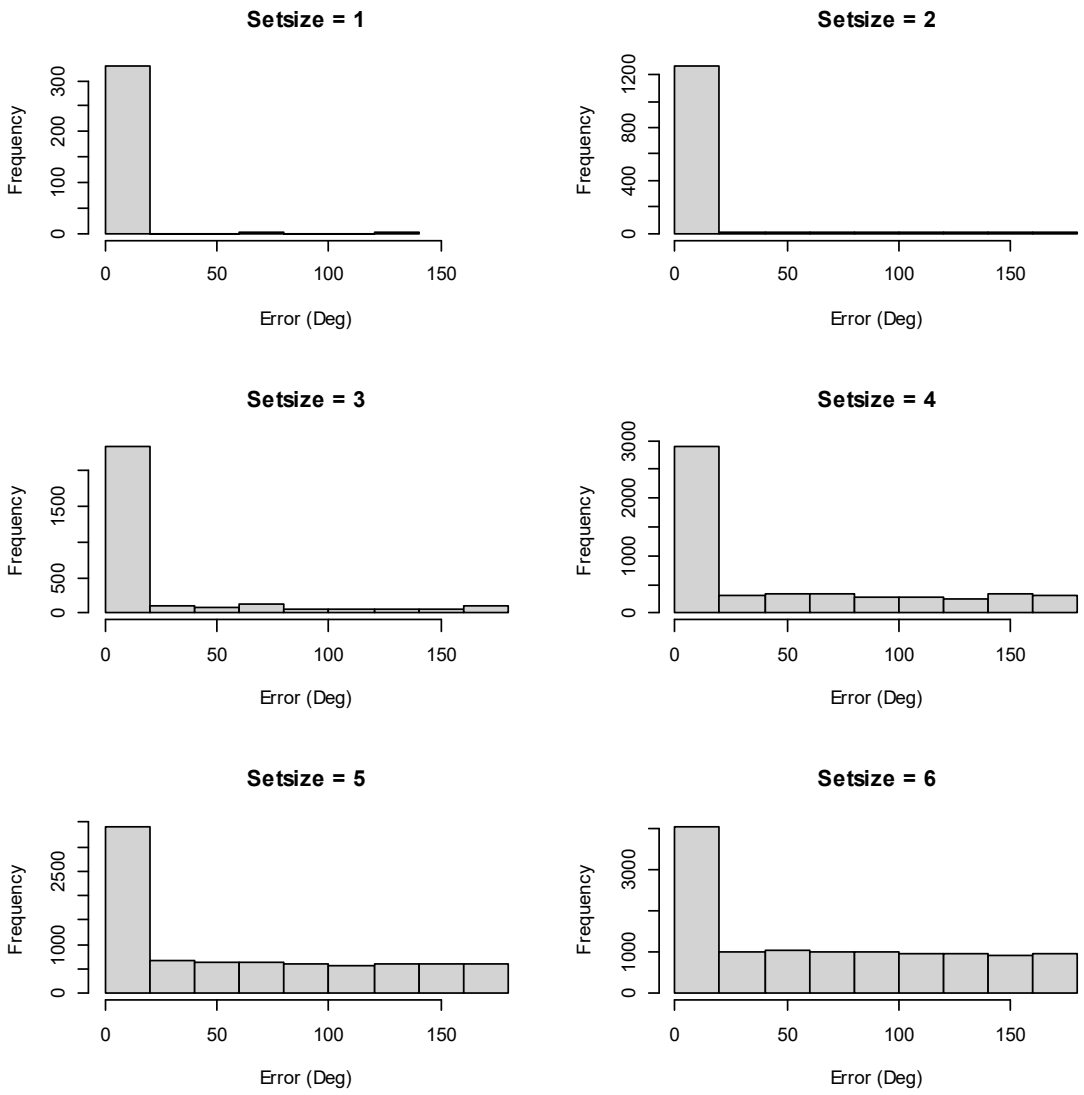


*Figure 11: Distributions of errors in Experiment 2 as a function of set size.*

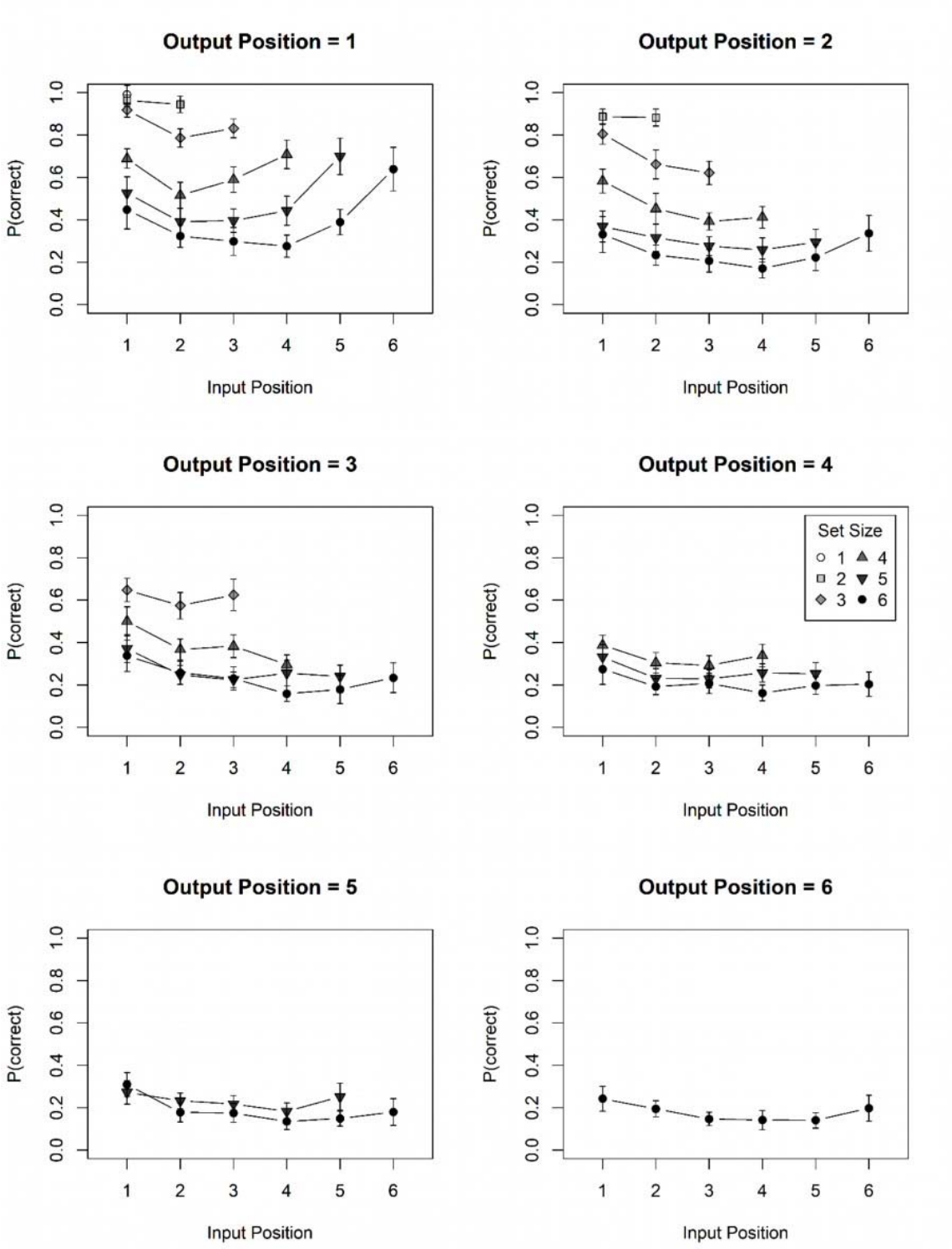


*Figure 12. Proportion of correct responses as a function of set size and input position in Experiment 2.*

The effects in the data mirror those from the visual WM experiment: In addition to a large set-size effect, there was a primacy effect over input position and, at higher set sizes, also a recency effect, which was more pronounced at the first than at later output positions. In addition, performance declined with increasing output position.

*Applying the IM*

We first applied the IM to the data in the same way as for Experiment 1. That resulted in extreme estimates for the precision parameter κ. This is a consequence of the shape of the error distribution: There was no gradual decline in frequency with larger values of error, so that there is no information in the data to localize a finite precision value. We therefore gave up on the idea of measuring the precision of phonological representations with the present data set, and developed a version of the IM for categorical responses (IMcat) to fit the present data. The IM for categorical data combines the assumptions about the primacy gradient and recency gradients of memory strength, and the focus of attention, with the response-selection assumptions of the $M^3$ framework (Oberauer & Lewandowsky, 2019). The $M^3$ is a framework for building simple memory measurement models. The measurement models are designed to be fit to each condition of an experiment separately, and therefore incorporate no assumptions about the effects of experimental manipulations such as set size or serial position. The $M^3$ shares with the IM the assumption that response candidates receive activation from three sources at retrieval -- cue-based reactivation, persistent activation, and background noise (Equation 1) -- and the assumption that the activation distribution is translated into a probability distribution of responses through Luce's choice rule (Equation 5). To derive IMcat from IM, we merely need to change the equations for the distributions of the three components (Equations 2-4) so that they express an activation distribution over a discrete set of response candidates. To that end we replace the von-Mises distribution by the Kronecker delta function, $\delta_{x,x_i}$ which is 1 when $x = x_i$, and 0 otherwise.

The $A_c$ component of IMcat, representing the contribution of cue-based retrieval, is again a sum of the activation distributions across all *n* elements in the memory set, weighted by how strongly each element is cued through its binding to the context cue:

$$A_c(x \mid L_\theta) = \sum_{i=1}^{n} \exp[-s \cdot D(L_i, L_\theta)]\delta_{x,x_i}. \quad (14)$$

The $A_a$ component simply assigns an activation of 1 to all elements of the memory set:

$$A_a(x) = \sum_{i=1}^{n} \delta_{x,x_i} \quad (15)$$

The $A_b$ component representing background noise adds a constant value of 1 to all elements in the candidate set, multiplied with the set size:

$$A_b(x) = n \quad (16)$$

*Modeling Results*

Table 2 summarizes the model comparison results. The kind of noise implementation had only a small and unsystematic effect on the goodness of model fits. Models without a focus of attention, as well as models with fixed noise, were clearly inferior to the less constrained model versions. The best-fitting model according to AIC was the model with a signal-detection theoretical implementation of noise, and a mixture of direct retrieval and a scanning strategy for accessing the target item, though the variant using mediated retrieval instead of scanning did not fit much worse. According to BIC, a simpler model with only mediated retrieval was even slightly preferred.

*Table 2: AIC and BIC Differences of All Model Variants to the Best-Fitting Variant, Experiment 2*

| Sub-Variant | ΔAIC | | ΔBIC | |
|---|---|---|---|---|
| | Uniform background noise | SDT Noise | Uniform background noise | SDT Noise |
| Direct and mediated retrieval | 7.0 | 2.1 | 8.3 | 3.4 |
| Direct retrieval only | 5.7 | 5.8 | 1.9 | 1.9 |
| Mediated retrieval only | 9.1 | 13.6 | 0 | 4.5 |
| Direct and scanning | 14.8 | 0 | 16.2 | 1.3 |
| Direct and mediated, no focus | 16.2 | 8.2 | 12.3 | 4.3 |
| Direct and scanning, no focus | 16.2 | 6.9 | 12.3 | 3.0 |
| Direct and mediated, fixed noise | 14.9 | 9.1 | 16.3 | 10.4 |
| Direct and scanning, fixed noise | 14.9 | 9.0 | 16.2 | 10.4 |

*Note: For each fit index (AIC or BIC), the best of the 16 model variant is identified, and its fit index subtracted from those of all variants.*

Figure 13 shows the distribution of parameter estimates across participants for the model version that fits best according to AIC (i.e., the model with SDT noise, and a mixture of direct retrieval and scanning). Figure 14 shows the model predictions for the effect of set size. Figure 15 presents the predictions alongside the observed proportion correct by set size, input position, and output position. The model accurately reproduces the systematic trends in the data: The set-size effect, the U-shaped serial-position curve over input position, the monotonic decline of performance over output positions, and the reduction of the recency effect (in the input position curves) at later output positions. The latter interaction arises from the focus of attention, which gives the last-presented item an advantage when it is tested first. It is probably for that reason that the model versions

omitting the focus of attention fit substantially worse. By contrast, in Experiment 1 the model versions with and without the focus of attention had very similar fits. This is because, in Experiment 1, the interaction between input and output position was much less pronounced than in Experiment 2 – there was little specific advantage for the last-presented item when output first – and therefore relatively weak evidence for a last-in-first-out focus of attention.

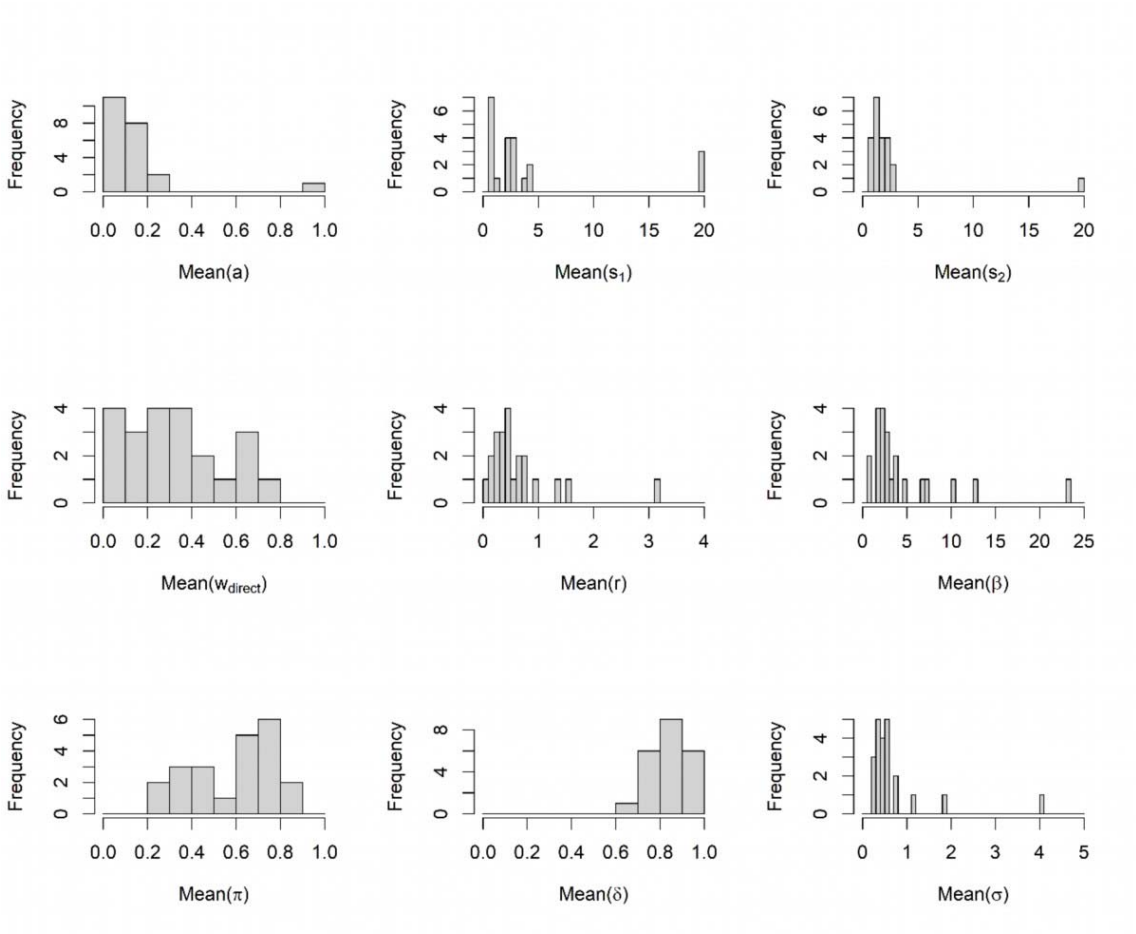


*Figure 13: Distribution across participants of best-fitting parameter values of the IMCat model for Experiment 2.*

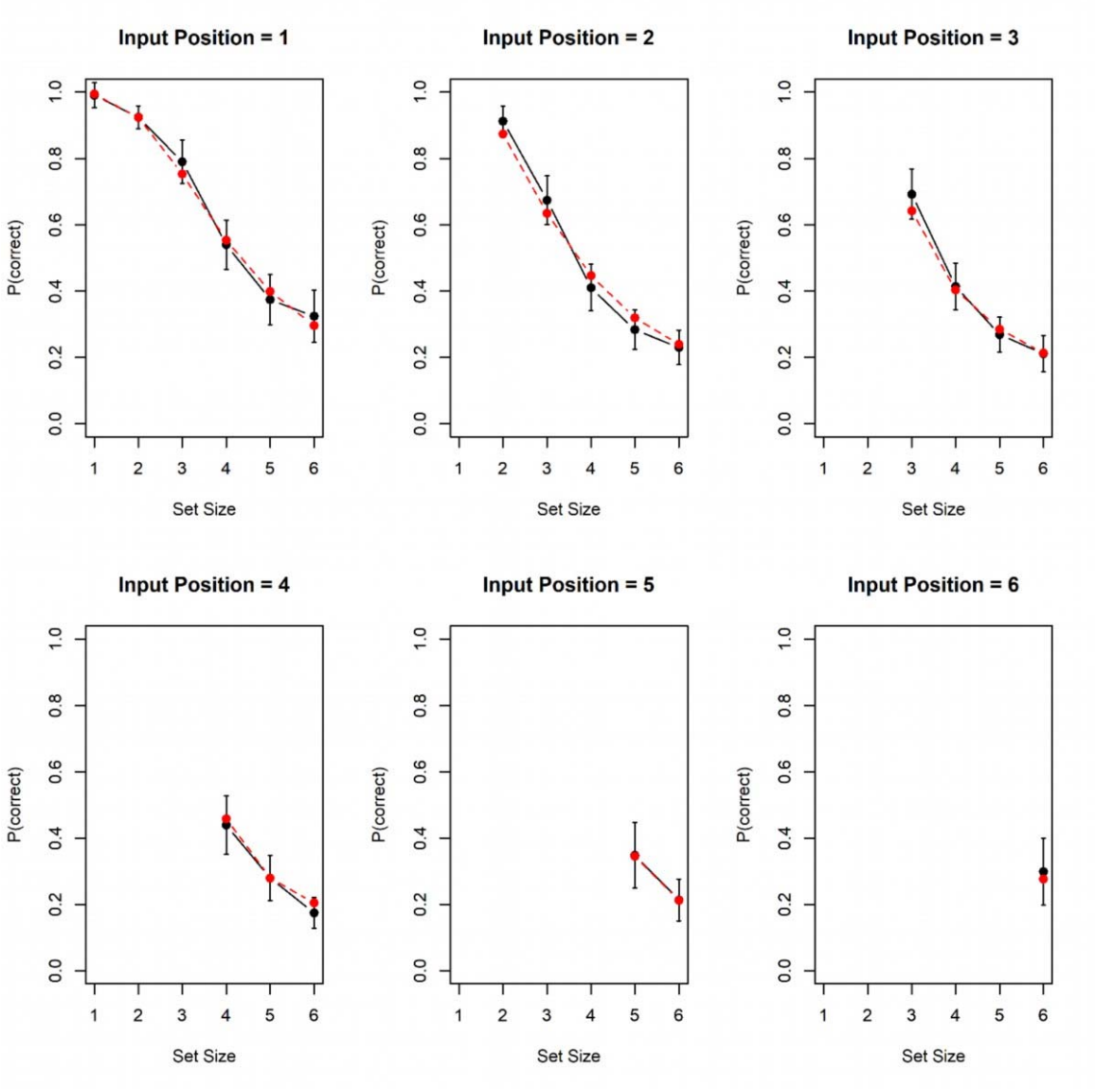


*Figure 14: Observed proportion correct (black) and predictions of the best-fitting IMCat model version (red) for the proportion of correct responses as a function of set size, averaged over output position, for Experiment 2.*

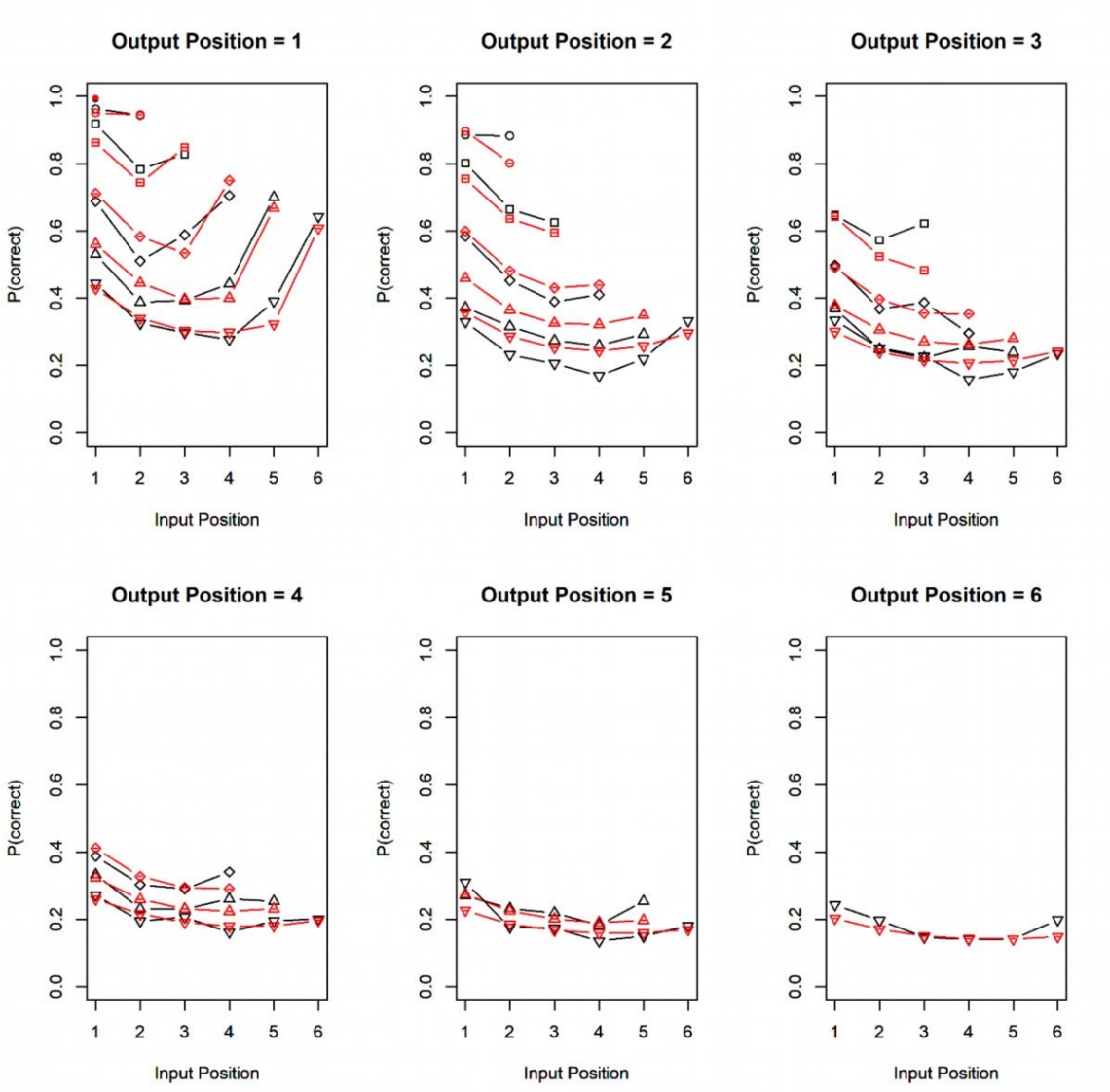


*Figure 15. Observed proportion of correct responses (black) and predictions by the best fitting IMCat version (red).*

## Discussion

With the present study we took two steps to bridge between research on visual WM for continuously varying features on the one hand, and on WM for lists of discrete (typically verbal) items on the other hand. First, we conducted experiments with both kinds of materials using the same experimental design, covering the effects most investigated in each area of research. The shared design enables a comparison between materials unconfounded by features of the experiment. Second, we developed a computational model accounting for the benchmark effects measured with this design for both kinds of materials. To this end we extended the scope of application of the Interference Model (IM) of visual working memory (Oberauer & Lin, 2017) to sequentially presented lists, and incorporated an account of serial-position effects over input positions when items are presented sequentially, and for effects of output position when multiple items are tested sequentially. In addition, we attempted to develop verbal materials varying on an approximately continuous dimension of similarity, so that we could apply the IM, including its predictions for the precision of WM representations, to verbal materials.

The model reproduced the qualitative pattern of effects in both experiments, and achieved a reasonably good quantitative fit. Similar model versions won the competition for both experiments, and their parameter distributions were very similar for the two experiments (compare Figures 4 and 13). We conclude that the model passed a first test of generalizability between two kinds of materials, which have been preferentially investigated in two WM research traditions. We next discuss the new assumptions we added to the IM to account for serial-position effects, their implications for the explanation of the set-size effect, and our failed attempt to measure phonological similarity on a continuous scale.

### Serial-Position Effects

The assumption of a primacy gradient of encoding strength builds on precedents and strong evidence (Farrell & Lewandowsky, 2004). Explaining a primacy gradient of accuracy by a primacy

gradient of encoding strength is not very satisfying, as the explanation adds little to the observation it explains – though it does add some theoretical commitments: The primacy gradient arises during encoding, independent of the information encoded, and its strength is controlled by the person's intention. Still, a more complete explanation would require spelling out a plausible mechanism that generates the primacy gradient of encoding strength.

We can think of two such mechanisms. One is that participants intentionally reduce the strength of encoding of each subsequent item in a memory set because they are aware of the fact that, by default, the automatic updating mechanism of WM generates a strong recency gradient. The intentional primacy gradient partially counteracts the recency gradient and thereby protects the accessibility of list-initial items. There is evidence that people can control to some extent how strongly an event is encoded into WM (Oberauer, 2018; Oberauer & Lewandowsky, 2019), and there is also evidence that the primacy gradient depends strongly on people's expectations of whether or not they will need to recall a list of items from the beginning (Duncan & Murdock, 2000; Palladino & Jarrold, 2008), and on their intention to remember a list (Oberauer & Greve, 2021).

A second mechanism for generating the primacy gradient could be that encoding of an event into WM depletes a resource for memory encoding that needs more time to be replenished than the typical presentation rate of items in a WM task, and therefore the available encoding resource decreases from each item to the next. Popov and Reder (2020) have made that assumption for encoding into episodic long-term memory; recent work provides support for such an encoding resource also in WM (Mızrak & Oberauer, 2021; Oberauer, 2022). If each item recruits a constant proportion of the resource, the amount of encoding resource used declines by a constant proportion per item. If the amount of encoding resource invested in encoding an item is under the person's control, this mechanism would be compatible with the fact that primacy gradient depends on people's expectations.

Our proposed explanation of the recency effect by a proportional decrement of the strength of cueing for earlier list positions is novel. We make this proposal because existing explanations for

the recency effect in serial-order memory are insufficient. There are two such explanations: First, when items are recalled in forward order, and each recalled item is suppressed afterwards, then the number of not-suppressed response candidates decreases towards the end of the list, improving performance (Farrell & Lewandowsky, 2012). Response suppression, however, predicts a recency effect over output position, not input position, when the two are unconfounded, as in our experiments.

The second existing explanation is that, if list position is used as a retrieval cue, a recency effect arises as an edge effect because the last item has no neighbor following it. This explanation does not apply to memory tests in which the input position is not used as a retrieval cue. One such task is global recognition (e.g., the Sternberg paradigm), which is characterized by strong recency (Donkin & Nosofsky, 2012). Another such task is a probed-recall task in which each item is probed by a retrieval cue that is uncorrelated with its input position, as in the present Experiment 1, and several other experiments with verbal (Jones & Oberauer, 2013) and visual materials (Gorgoraptis et al., 2011). One way in which an edge effect could arise in probed-recall tasks is through an indirect retrieval route, mediated through input position (Jones & Oberauer, 2013). To assess this possibility, we modelled the data of Experiment 1 as a mixture of the direct retrieval route (from spatial location to color) and the mediated retrieval route (from spatial location to input position to color). The mixture weight $w_{direct}$ was estimated values across the entire range between 0 and 1 for different persons, implying that the mediated retrieval route was used at least sometimes. However, this was not enough to fully explain the recency effect as an edge effect: The estimates of the δ parameter, which controls the recency gradient, were all < 1.

Our proposal for explaining the recency effect has three arguments in its favor. First, it helps to explain the present data. This does not mean that it is without alternative, but we were not able to find an alternative solution to achieve a fit even remotely as good, suggesting that there is at least no obvious viable alternative. Second, it has three theoretical interpretations that are mathematically equivalent on the abstract level on which the IM is described (see Appendix A). This could be seen as

a weakness because we cannot adjudicate between the three interpretations with the present data. We see it as a strength because it lends our proposal a higher prior probability than it would otherwise have: If one of the three hypothetical mechanisms that imply our formal expression for the recency gradient is correct, our proposal would be correct, so it inherits the sum of the priors of the three hypothetical mechanisms. The third argument is that the multiplicative combination of primacy and recency gradients of strength contributes substantially to the explanation of the set-size effect, as shown in Figure 2: With larger set sizes, the strengths at all serial positions decrease.

**The Set-Size Effect**

The set size effect is the primary empirical signature of the capacity limit of WM (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), and the main target of explanation for competing models of visual WM (van den Berg et al., 2014). Most of these models explain it as arising from a limited resource that is either infinitely divisible among all memory items, or discrete, providing a limited number of "slots" for storing a number of items. The IM demonstrates that the set-size effect can be explained without appeal to a limited resource. In the original version of the IM, the set-size effect emerges from a combination of three effects: First, with larger set sizes more non-target items contribute to the distribution of activation over feature space at retrieval, resulting in a larger proportion of transposition or swap errors. Second, the background noise added to binding space at encoding, which is responsible for the uniform activation component *Ab*, increases with each additional item. Third, with larger set sizes, the chance that the tested item is in the focus of attention decreases. In the present version of the IM for sequential presentation, the serial-position effects make a large contribution to the set-size effect.

The contribution of the focus of attention became less important for the full-report version of WM tasks used here, because it provides a performance boost only if the last-presented item is tested first. The best-fitting model for Experiment 1 does not even include the focus of attention. The increase of background noise with set size also became unnecessary for explaining the set-size effect:

The best-fitting model uses the formalization of background noise from signal-detection theory, as introduced by Schurgin et al. (2020), and in this model version the noise is constant across set sizes.

Visual WM performance with sequential presentation has been found to be sometimes a bit better than with simultaneous presentation (Mance, Becker, & Liu, 2012; Miller, Becker, & Liu, 2014), sometimes a bit worse (Gorgoraptis et al., 2011), and sometimes equally good (Johnson, Spencer, Luck, & Schöner, 2009; Sewell, Lilburn, & Smith, 2014). If serial-position effects in sequentially presented memory sets contribute substantially to the set-size effect, then why is performance with simultaneously presented arrays about the same as with sequentially presented ones? One plausible explanation is that with simultaneously presented arrays, items are still encoded into WM sequentially. In models – such as the IM – in which content features are bound to contexts through the rapid formation of temporary bindings, sequential encoding would circumvent a computational problem: When *n* items are presented simultaneously in *n* different contexts (e.g., six colors in six different location, or six orientations in six different colors), how can each item be bound to the correct context? Sequential encoding could solve this problem by activating one item and its context at a time, then binding them together (e.g., through a Hebbian learning rule), and sequentially adding the binding-strength distributions to the binding space. One model of visual WM that describes the encoding process explicitly makes this assumption (Manohar, Zokaei, Fallon, Vogels, & Husain, 2019), and Shepherdson, Hell, and Oberauer (2022) provide evidence for it. If simultaneously presented items are encoded sequentially into WM, the primacy and recency effects observed with sequential presentation would apply equally (though not observable as long as we don't know people's order of encoding), and would contribute to the set-size effect of simultaneously presented memory sets as well.

A prominent debate among WM researchers revolves around the question whether the capacity limit is caused by domain-general or domain-specific limiting factors (Cowan, Saults, & Blume, 2014; Fougnie, Zughni, Godwin, & Marois, 2015; Vergauwe, Barrouillet, & Camos, 2010). Many studies have investigated this issue by measuring the decline of memory performance when

the load on WM is increased by adding items from the same domain (e.g., all verbal, or all spatial) versus from another domain (e.g., a combination of verbal and spatial memoranda). The pattern emerging across such studies is that adding any further information to WM impairs performance, but the impairment is larger when the added information comes from the same domain as the tested WM content (Oberauer et al., 2018), pointing to both domain-general and domain-specific factors limiting WM capacity. One important limitation of previous interference-based models of WM (Nairne, 1990; Oberauer et al., 2012) has been that they explain the set-size effect largely through interference between representations within the same feature space, and therefore have difficulties explaining the domain-general sources of the WM capacity limit (Oberauer et al., 2016). In the new version of the IM proposed here, the set size effect arises from some factors that are arguably domain-specific (i.e., the confusion of targets with non-targets, and potentially the background noise), and others that are plausibly domain-general (i.e., the declining strength of encoding across successive items, and the automatic updating decrement, as well as the focus of attention). Therefore, the IM is in a good position to explain both the domain-specific and the domain-general factors contributing to the limited capacity of WM.

To conclude: Rather than aiming to explain the set-size effect as a direct manifestation of a limited mental commodity, as has been dominant in the visual-WM tradition, it might be more parsimonious to explain it as emerging from two factors constraining access to WM: The interference between multiple representations, and the modulation of binding strength as a function of serial position.

**The Precision of Phonological Representations in Working Memory**

Whereas for Experiment 1 we started from a popular paradigm testing WM for continuous (visual) features and developed it in the direction of typical paradigms for discrete-item lists, the

design of Experiment 2 derives from a standard probed-recall paradigm for testing lists of discrete items, and we modified it to approximate a continuous dimension of stimulus similarity.

Contrary to our expectations, we found that participants showed little tendency to confuse the target nonword more often with alternatives closer in our phonological-similarity circle. This was unexpected in light of the well-established phonological-similarity effect (Conrad, 1964; Conrad & Hull, 1964): In serial recall, lists of phonologically similar items lead to more confusions among list items than lists of dissimilar items. In experiments demonstrating the phonological-similarity effect, the difference between similar and dissimilar pairs of items is at least one phoneme (when consonants are used), or several phonemes (in case words are used).

We therefore considered the possibility that memory confusion errors are not sensitive to the more fine-grained gradation of similarity in terms of phonemic features that we used to construct the phonological-similarity scales. We re-analyzed the responses in Experiment 2 as a function of a coarser similarity metric, the number of different consonants (which corresponds approximately to the number of different phonemes). Figure 16 (top row) shows the probability of choosing an erroneous response candidate for candidates differing in one or in two consonants from the target.[5] Across all set sizes, candidates differing in only one consonant had a higher chance of being chosen, after correcting for chance.[6]

The bottom row of Figure 16 shows why this tendency did not translate into a gradual decline of confusions with increasing distance on our similarity scale. Although candidates that differed from the target in only one consonant were usually close to the target in the phonological similarity circle (large orange squares along the diagonal in the bottom-left panel), they were occasionally also far away (arms of small orange square pointing away from the diagonal). The

[5] Two consonants was the maximally possible difference, because the vowels were all the same in each set of 19 candidate nonwords.

[6] To account for the different numbers of candidates differing in one vs. two consonants from the target, we counted the number of choices in each category (1 vs. 2 consonants) and divided them by the number of candidates in each category (left panel), or the number of list items in each category (right panel).

bottom right panel of Figure 16 shows that actual confusions closely followed the distribution of candidates differing from the target in only one consonant. We conclude that for the purpose of predicting confusions in WM, measuring phonological similarity in terms of the number of different consonants, or different phonemes, is more appropriate than measuring the number of changed phonemic features.

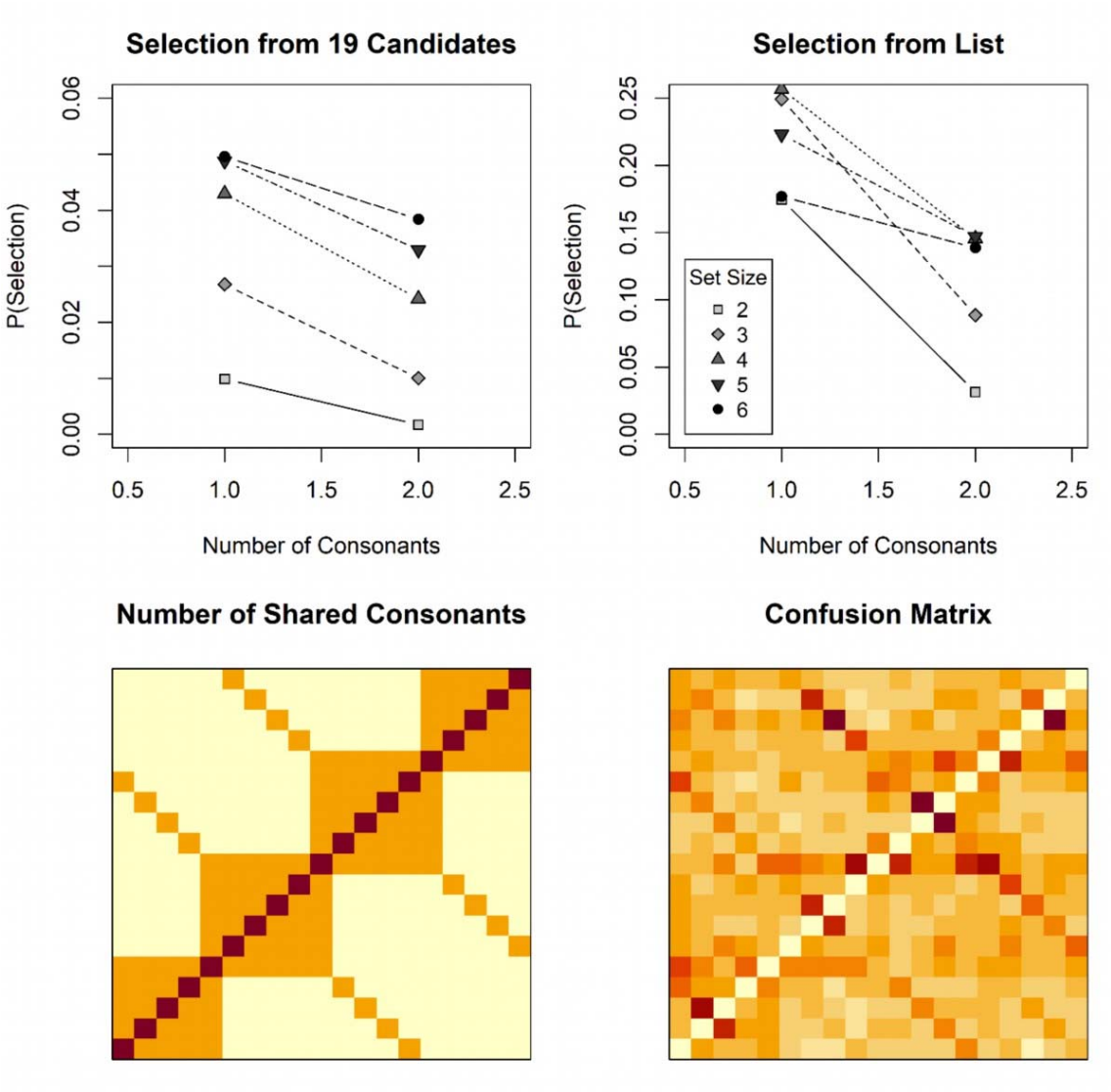


*Figure 16: The top panels show the proportion of selections of response candidates differing in one or two consonants from the target, out of all candidates differing in the same number of consonants (left), or out of all list items differing in the same number of candidates. The bottom left panel shows the number of consonants shared between each pair of candidates, ordered by their place on the*

*phonological-similarity circle on each axis (purple = 2, orange = 1, yellow = 0). The bottom right panel shows the frequencies of confusing each target (y-axis) with each other candidate (x-axis).*

**Toward an Integrative Model of Working Memory**

The IM incorporates theoretical assumptions that we argue hold for WM generally regardless of the kind of material:

(1) Contents are encoded into WM by temporarily binding each content element to a unique context. The content elements can be any object or event that is represented by the person as a unit, either because it is unitized by perceptual principles (such as a closed contour forming a visual object) or because it has been learned as a unit (i.e., a chunk), such as a word. The context can be any information that serves as a distinct retrieval cue to a content element, such as a perceptual or semantic feature, its location in space, or its ordinal position in a list. Access to an element of a memory set relies on cue-based retrieval, using the context information to re-activate the content representation(s) bound to it.

(2) When a context is used as retrieval cue, it activates the content item bound to it, and in addition other content items to the extent that their contexts are similar to the cued context. This leads to the activation of multiple content representations at retrieval. The probability of choosing an item for a response is proportional to its relative activation at retrieval. This assumption is sometimes referred to as *competitive cueing*, and it is a fixture in contemporary models of serial-order memory (Hurlstone et al., 2014). It explains the characteristic transposition gradients along the list-position dimension in serial-recall tasks (reviewed in Hurlstone et al., 2014), along spatial distances in visual-array tasks that use location as retrieval cue (Bays, 2016; Oberauer & Lin, 2017; Rerko et al., 2014), and along feature dimensions in tasks using a visual feature as retrieval cue (Oberauer & Lin, 2017).

(3) When elements of a memory set are encoded sequentially, they are encoded with declining strength, creating a primacy gradient on the strength of content-context bindings. With each encoded item, all current bindings in WM are downgraded in strength by a constant proportion,

creating a recency gradient. Jointly, these mechanisms create a U-shaped serial-position function of binding strength. U-shaped serial-position curves have been observed for both accuracy and RT measures in every kind of WM test of sequentially presented materials (for a review see Oberauer et al., 2018). The set-size effect on accuracy – the most obvious demonstration of a capacity limit of WM – is to a large extent a side-effect of this serial-position curve (Monsell, 1978).

(4) The proportional decrement of WM contents continues with every retrieval event, leading to a continuous decline of performance over successive tests of different items in the array, called output interference. This phenomenon has been observed with verbal materials (Cowan et al., 2002; Oberauer, 2003) and with visual materials (Adam et al., 2017; Peters et al., 2018).

(5) The retrieved contents have a limited precision. This precision is reflected in the dispersion of a distribution of activation over the feature dimension(s) on which the response candidates are distinguished.

(6) Representations in WM have two formats. One is the matrix of temporary bindings between contents and contexts, and the other is the persistent activation of content or context representations. The latter consists of a small amount of persistent activation of all items in the memory set, represented by activation component $A_a$, plus the activation of the last-encoded content and its context in the focus of attention. On the neural level, we can think of the $A_a$ component and the focus of attention as the neurally active representations that can be decoded from patterns of neural activity during the retention interval, whereas the bindings can be thought of as the neurally silent content of WM (Lewis-Peacock, Drysdale, Oberauer, & Postle, 2011; Manohar et al., 2019; Rose et al., 2016; Stokes, 2015). One prediction from this assumption (which the IM shares with the model of Manohar et al., 2019) is that the last-encoded item can be decoded from neural activity much better than all previously encoded items.

## Conclusion and Outlook

The IM was initially developed to account for data from the continuous-reproduction paradigm for studying visual WM of continuously varying features (Oberauer & Lin, 2017). We have meanwhile extended it to the second popular paradigm for investigating visual WM, change detection (Lin & Oberauer, 2022). With the present work we extend the model's scope of application further to sequentially presented memory sets, and to discrete list items. We also revise the model's assumption about how to incorporate noise: Instead of assuming a uniform distribution of background activation over all response candidates, here we adopt the assumption from signal-detection theory that noise is added to the strength of memory-based of evidence for each response candidate at retrieval. This revision is based on the present comparisons of model versions with the two ways of implementing noise, and on a previous comparison of measurement models, in which models with SDT noise fit better than models with the uniform background noise of the original IM (Oberauer, 2021).

With its extensions so far, the IM already has a broader explanatory scope than other formal models of visual WM (Schneegans & Bays, 2017; Schneegans, Taylor, & Bays, 2020; Swan & Wyble, 2014; van den Berg et al., 2012; Zhang & Luck, 2008). In comparison with formal models of list memory that arose from the older tradition of WM research, however, its explanatory scope is still limited. These models account for benchmark findings still outside of the scope of the IM, such as the phonological similarity effect (Farrell, 2006) and the effect of temporal grouping (Burgess & Hitch, 1999), effects of processing distractors in the complex-span paradigm (Oberauer et al., 2012), the acquisition of long-term knowledge of repeated lists (Burgess & Hitch, 2006; Page & Norris, 2009), and the relationship between serial recall and free recall (Farrell, 2012). Covering a larger set of benchmark findings about WM (Oberauer et al., 2018) should be a goal for theory development. As the IM shares many of its assumptions with successful models of list memory, it is in a good position to serve as the basis for developing a more comprehensive model of WM.

## Appendix: Alternative Interpretations of Recency

In addition to the automatic updating described in the main text, we consider two alternative interpretations of the recency gradient controlled by parameter *Δ*: Cueing with the current temporal context, and cueing with an end marker.

### Cueing with Current Temporal Context

At encoding, each item is bound to a context representation that is to be used as the given retrieval cue at test (for instance, the item's spatial location), and the strength of that binding declines according to a primacy gradient, as described in the main text. In addition, each item is bound to the current temporal context. The temporal context changes by a constant proportion δ with each event, so that the similarity of contexts of successive events is *δ*, and their psychological distance is *log(1/δ)*. At retrieval, memory is cued by a combination of the given retrieval cue (e.g., its spatial location) and the current temporal context. The two cues are combined multiplicatively to obtain the cueing strength for each item *i* (considering only the direct retrieval path for simplicity):

$$\begin{aligned} Q(i \mid L_\theta, o) &= (1+\beta P^{i-1}) \exp(-s \cdot D(L_i, L_\theta)) \cdot \exp(-\log(\delta^{-1})(n-i+o-1)) \\ &= (1+\beta P^{i-1}) \exp(-s \cdot D(L_i, L_\theta)) \cdot \delta^{n-i+o-1} \end{aligned} \quad (17)$$

Summing over all array items *i* yields the $A_c$ component:

$$A_c(x \mid L_\theta, o) = \sum_{i=1}^{n} (1+\beta P^{i-1}) \exp(-s \cdot D(L_i, L_\theta)) \cdot \delta^{n-i+o-1}, \quad (18)$$

which is equivalent to the $A_c$ component in the main text.

### Cueing with End Marker

Each item is bound to the context that is to serve as retrieval cue, with a strength declining according to a primacy gradient. In addition, each item is bound to an end marker – a representation of the end of the current episode (e.g., of the current memory set, or the current trial). With each new event in the episode, the existing bindings to the end marker are decreased by multiplication

with $\delta$, so that the strength of binding to the end marker reflects the relative recency of events. The strength with which the end marker cues each item *i* in a memory set of size *n* is $\delta^{n-i}$. If we assume that output events are treated as belonging to the same episode as item presentations, the end marker would continue to be updated during output, and the cueing strength decline over output positions: $\delta^{n-i+o-1}$. At test, the given cue and the end marker are multiplicatively combined, yielding the cueing strength *Q* for item *i* (again considering only the direct retrieval path):

$$Q(i \mid L_\theta, o) = (1 + \beta P^{i-1}) \exp(-s \cdot D(L_i, L_\theta)) \cdot \delta^{n-i+o-1}, \qquad (19)$$

which is again equivalent to the proposal in the main text, and to the first alternative above.

We assume that the cues are combined multiplicatively because that makes them non-compensatory: Both cues need to be strong for a high combined cueing strength for an item. In contrast, if the cues were combined additively, the strong cueing of the last item by the current context (first alternative) or the end marker (second alternative) could overpower the cueing of non-final items by the given retrieval cue. Multiplicative combination of cues has precedents in memory models: In SAM (Raaijmakers & Shiffrin, 1981), the cueing strengths from associations between words in a list and from associations of the general list context to each word are combined multiplicatively. In the SIMPLE model, cueing strength for item *i* is modelled as an exponential function of the distance between the retrieval cue and item *i* in psychological space. Cues on several dimensions are combined by adding their distances to the item along each dimension in the exponent, which comes down to a multiplicative combination of the cueing strengths. For instance, when an item *i* is cued by the spatial position $L_\theta$ and the temporal position $T_\theta$, the combined cueing strength in SIMPLE is:

$$\begin{aligned} Q(i \mid L_\theta, T_\theta) &= \exp\left(-s \cdot \left(D(L_i, L_\theta) + D(T_i, T_\theta)\right)\right) \\ &= \exp\left(-s \cdot D(L_i, L_\theta)\right) \cdot \exp\left(-s + D(T_i, T_\theta)\right). \end{aligned}$$

References

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology, 97*, 79-97. doi:10.1016/j.cogpsych.2017.07.001

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limits on retrieval. *Cognitive Psychology, 30*, 221-256.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. VIII, pp. 47-90). New York: Academic Press.

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavioral Research Methods, Instruments, & Computers, 28*, 584-589.

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

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports, 6*. doi:10.1038/srep19203

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*, 1-11.

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

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 353-369.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539-576.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627-652.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75-84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Pychology, 55*, 429-432.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

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

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General, 143*, 1806-1836. doi:10.1037/a0036814

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153-177.

Cowan, N., Saults, J. S., & Morey, C. C. (2006). Development of working memory for verbal-spatial associations. *Journal of Memory and Language, 55*, 274-289.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*, 415-423.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: empirical and computational investigation of recency effects. *Psychological Review, 112*, 3-42.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science, 23*, 625-634. doi:10.1177/0956797611430961

Duncan, M., & Murdock, B. (2000). Recognition and recall with precuing and postcuing. *Journal of Memory and Language, 42*, 301-313.

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12*, 1-16.

Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55*, 587-600.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*, 223-271.

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51*, 115-135.

Farrell, S., & Lewandowsky, S. (2012). Response suppression contributes to recency in serial recall. *Memory & Cognition, 40*, 1070-1080.

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General, 144*, 30-47. doi:10.1037/a0038211

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology, 20*, 177-182.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience, 31*, 8502-8511.

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience, 15*, 1176-1194.

Grossberg, S., & Stone, G. (1986). Neuronal dynamics of attention switching and temporal order information in short term memory. *Memory & Cognition, 14*, 451-468.

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36*, 73-137.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A*, 80-115.

Hepner, C. R., & Nozari, N. (2019). Resource allocation in phonological working memory: Same or different principles from vision? *Journal of Memory & Language, 106*, 172-188. doi:10.1016/j.jml.2019.03.003

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*, 339-373. doi:10.1037/a0034221

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*, 568-577.

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

Jones, T., & Oberauer, K. (2013). Serial-position effects for items and relations in short-term memory. *Memory, 21*, 347-365. doi:10.1080/09658211.2012.726629

Joseph, S., Iverson, P., Manohar, S., Fox, Z., Scott, S. K., & Husain, M. (2015). Precision of working memory for speech sounds. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2014.1002799

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kool, W., Conway, A. R. A., & Turk-Browne, N. B. (2014). Sequential dynamics in visual short-term memory. *Attention, Perception & Psychophysics, 76*, 1885-1901.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. New York: Academic Press.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43*(149-164).

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54*, 20-38.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: new data and a model. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 49, pp. 1-48). London, UK: Elsevier.

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2011). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24*, 61-79.

Lin, H.-Y., & Oberauer, K. (2022). An interference model for visual working memory: Applications to the change detection task. *Cognitive Psychology, 133*, 101463. doi:https://doi.org/10.1016/j.cogpsych.2022.101463

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

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

Madigan, S. A. (1980). The serial position curve in immediate serial recall. *Bulletin of the Psychonomic Society, 15*, 335-338.

Mance, I., Becker, M. W., & Liu, T. (2012). Parallel consolidation of simple features into visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 429-438.

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience & Biobehavioral Reviews, 101*, 1-12. doi:10.1016/j.neubiorev.2019.03.017

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 46, pp. 155-200). San Diego: Academic Press.

Miller, J. R., Becker, M. W., & Liu, T. (2014). The bandwidth of consolidation into visual short-term memory depends on the visual feature. *Visual Cognition, 22*, 920-947. doi:10.1080/13506285.2014.936923

Mızrak, E., & Oberauer, K. (2021). What is time good for in working memory? *Psychological Science, 32*(8), 1325-1337. doi:10.1177/0956797621996659

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10*, 465-501.

Murdock, B. B., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74*, 137-143.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251-269.

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49*(4), 469-483. Retrieved from <Go to ISI>://000186403100003

Oberauer, K. (2008). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 439-459. Retrieved from <Go to ISI>://000255543300001

Oberauer, K. (2017). What is working memory capacity? / ¿Qué es la capacidad de la memoria de trabajo? *Estudios de Psicología / Studies in Psychology, 38*, 338-384. doi:10.1080/02109395.2017.1295579

Oberauer, K. (2018). Removal of irrelevant information from working memory: sometimes fast, sometimes slow, and sometimes not at all. *Annals of the New York Academy of Science, 1424*, 239–255. doi:10.1111/nyas.13603

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

Oberauer, K. (2022). When does working memory get better with longer time? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0001199

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142*, 758-799. doi:10.1037/bul0000046

Oberauer, K., & Greve, W. (2021). Intentional remembering and intentional forgetting in working and long-term memory. *Journal of Experimental Psychology: General*. doi:10.1037/xge0001106

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science, 21*, 164-169. doi:10.1177/0963721412444727

Oberauer, K., & Lewandowsky, S. (2019). Simple measurement models for complex working memory tasks. *Psychological Review, 126*, 880-932. doi:10.1037/rev0000159

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A. R. A., Cowan, N., . . . Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*, 885-958. doi:10.1037/bul0000153

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

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

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society of London, 364*, 3737-3753.

Palladino, P., & Jarrold, C. (2008). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology, 61*, 392-399.

Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception & Psychophysics, 76*, 1914-1924.

Peteranderl, S., & Oberauer, K. (2018). Serial recall of colors: Two models of memory for serial order applied to continuous visual stimuli. *Memory & Cognition, 46*, 1-16. doi:10.3758/s13421-017-0741-0

Peters, B., Rahm, B., Czoschke, S., Barnes, C., Kaiser, J., & Bledowski, C. (2018). Sequential whole report accesses different states in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 588-603. doi:10.1037/xlm0000466

Plummer, M. (2016). JAGS 4.2.0. Retrieved from http://mcmc-jags.sourceforge.net/

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*, 1-46. doi:10.1037/rev0000161

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93-134. doi:10.1037/0033-295x.88.2.93

Rajsic, J., Swan, G., Wilson, D. E., & Pratt, J. (2017). Accessibility limits recall from visual working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 43*, 1415-1431. doi:10.1037/xlm0000387

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and de-focused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1175-1196. doi:10.1037/a0031172

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatially imprecise representations in working memory. *Quarterly Journal of Experimental Psychology, 67*, 3-15. doi:10.1080/17470218.2013.789543

Ricker, T. J., & Hardman, K. O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*. doi:10.1037/xge0000346

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*, 1136-1139. doi:10.1126/science.aah7011

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, 105*, 5975-5979.

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience, 37*, 3913-3925. doi:10.1523/jneurosci.3493-16.2017

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

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour, 4*(11), 1156-1172. doi:10.1038/s41562-020-00938-0

Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2014). An information capacity limitation of visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 2214-2242. doi:10.1037/a0037744

Shepherdson, P., Hell, L., & Oberauer, K. (2022). How does visual working memory solve the binding problem? *Journal of Experimental Psychology: Human Perception and Performance, 48*, 1137-1152. doi:10.1037/xhp0001044

Simons, D. J., & Lewin, D. T. (1998). Change blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Smith, P. L., Lilburn, S. D., Corbett, E. A., Sewell, D. K., & Kyllingsbæk, S. (2016). The attention-weighted sample-size model of visual short-term memory: Attention capture predicts resource allocation and memory load. *Cognitive Psychology, 89*, 71-105. doi:10.1016/j.cogpsych.2016.07.002

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception & Psychophysics, 78*, 1839–1860. doi:10.3758/s13414-016-1108-5

Stan-Development-Team. (2021). *Stan Modeling Language Users Guide and Reference Manual, Version 2.26*. Retrieved from https://mc-stan.org

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences, 19*, 394-405. doi:10.1016/j.tics.2015.05.004

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception & Psychophysics*. doi:10.3758/s13414-014-0633-3

Tamber-Rosenau, B. J., Fintzi, A. R., & Marois, R. (2015). Crowding in visual working memory reveals its spatial resolution and the nature of its representation. *Psychological Science*. doi:10.1177/0956797615592394

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

van den Berg, R., & Ma, W. J. (2014). "Plateau"-related summary statistics are uninformative for comparing working memory models. *Attention, Perception & Psychophysics*. doi:10.3758/s13414-013-0618-7

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *eLIFE, 7*, e34963. doi: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, 109*, 8780-8785.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science, 21*, 384-390.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92-114.

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

Wiese, R. (2000). *The phonology of German* (Vol. Oxford University Press): Oxford.

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

Yellott, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology, 15*, 109-144.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233-236.