Psychological Review
2017, Vol. 124, No. 1, 21–59

0033-295X/17/$12.00 http://dx.doi.org/10.1037/rev0000044

An Interference Model of Visual Working Memory

Klaus Oberauer and Hsuan-Yu Lin
University of Zurich

The article introduces an interference model of working memory for information in a continuous similarity space, such as the features of visual objects. The model incorporates the following assumptions: (a) Probability of retrieval is determined by the relative activation of each retrieval candidate at the time of retrieval; (b) activation comes from 3 sources in memory: cue-based retrieval using context cues, context-independent memory for relevant contents, and noise; (c) 1 memory object and its context can be held in the focus of attention, where it is represented with higher precision, and partly shielded against interference. The model was fit to data from 4 continuous-reproduction experiments testing working memory for colors or orientations. The experiments involved variations of set size, kind of context cues, precueing, and retro-cueing of the to-be-tested item. The interference model fit the data better than 2 competing models, the Slot-Averaging model and the Variable-Precision resource model. The interference model also fared well in comparison to several new models incorporating alternative theoretical assumptions. The experiments confirm 3 novel predictions of the interference model: (a) Nontargets intrude in recall to the extent that they are close to the target in context space; (b) similarity between target and nontarget features improves recall, and (c) precueing—but not retro-cueing—the target substantially reduces the set-size effect. The success of the interference model shows that working memory for continuous visual information works according to the same principles as working memory for more discrete (e.g., verbal) contents. Data and model codes are available at https://osf.io/wgqd5/.

Keywords: attention, computational modeling, interference, working memory

The limited capacity of working memory is one of the strongest determinants of individual differences in cognitive abilities (e.g., Kane et al., 2004; Oberauer, Süβ, Wilhelm, & Sander, 2007) as well as of developmental variation in cognitive abilities (Cowan et al., 2005; Hale, Myerson, Emery, Lawrence, & Dufault, 2007). Four hypotheses compete for explaining the limited capacity of working memory.

First, capacity is assumed to reflect a limited number of place holders or *slots* in the working memory system, such that a discrete number of chunks—independent representational units—can be maintained in working memory (WM) at any time (Cowan, 2005; Zhang & Luck, 2008). Second, WM capacity is assumed to be limited by a fixed *resource* that can be flexibly assigned to maintenance of any number of representations, and to processing of representations (Bays & Husain, 2008; Just & Carpenter, 1992). A third account explains the limited capacity of WM as arising from rapid *decay* of representations over time, which can be counteracted by rehearsal. Because rehearsal can strengthen only a limited amount of information before it decays beyond recovery, WM capacity is limited to what can be rehearsed before being lost through decay (Baddeley, Thomson, & Buchanan, 1975; Barrouillet, Bernardin, & Camos, 2004). A fourth family of theories explains the capacity limit of WM as reflecting *interference* between representations in WM (Nairne, 1990; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012).

Interference models have been very successful in explaining a large range of phenomena from experiments on WM, including phenomena reflecting the system's limited capacity (Jonides et al., 2008; Lewandowsky & Farrell, 2008; Nairne, 1990; Oberauer & Kliegl, 2006; Oberauer, Lewandowsky, et al., 2012). Nevertheless, interference models have been conspicuously absent from the current debate about models of the capacity of visual WM. In this debate, variants of the slot model (Cowan, Blume, & Saults, 2013; Fukuda, Awh, & Vogel, 2010; Zhang & Luck, 2008) and variants of the resource model (Bays, Catalao, & Husain, 2009; Palmer, 1990; van den Berg, Shin, Chou, George, & Ma, 2012) are the strongest contenders. A recent systematic model-comparison effort considered only variants of these two theory families (van den Berg, Awh, & Ma, 2014) and a recent review of the theoretical questions about visual WM likewise considers only these two options (Suchow, Fougnie, Brady, & Alvarez, 2014).

One consequence of this limited theoretical perspective is that most current models of visual WM only account for the quality and quantity of *item* information maintained in WM (i.e., information of which items have been seen in a memory array), disregarding *binding* information (e.g., information of which item has been in which location, or in which sequential position of a series of items). Although binding errors are empirically well documented in visual WM tasks (Bays et al., 2009; Pertzov, Dong, Peich, & Husain, 2013; Rerko, Oberauer, & Lin, 2014; van den Berg et al.,

This article was published Online First November 21, 2016.

Klaus Oberauer and Hsuan-Yu Lin, Department of Psychology, University of Zurich.

This research was supported by a grant from the Swiss National Science Foundation (project 100014_135002) to Klaus Oberauer. We thank Sonja Peteranderl and Justus Spengler for collecting the data.

Correspondence concerning this article should be addressed to Klaus Oberauer, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: k.oberauer@psychologie.uzh.ch

2014), none of the models currently competing for a quantitative account of visual WM provides a mechanistic explanation for why they occur. In contrast, binding errors are a natural consequence of interference models of WM, as we will explain below.

One obstacle for applying interference models to visual WM is that this class of models so far has been developed only for the recall of discrete memoranda such as digits, letters, words, or spatial locations in a grid. In contrast, the most informative tests of visual WM involve recall of visual features on a continuous dimension, such as color or orientation. These memory tests return a measure of deviation on a continuous scale, rather than a discrete score of correct or incorrect recall. For instance, one popular paradigm involves the brief presentation of an array of visual objects, to be remembered for one or a few seconds, after which the location of one object is probed for recall. Participants' task is to report a feature of that object on a continuous response scale, such as clicking on the remembered color on a color wheel, or reproducing the remembered orientation with a dial (Bays, Wu, & Husain, 2011; Wilken & Ma, 2004; Zhang & Luck, 2008). This continuous-reproduction paradigm is particularly useful for studying WM because it provides information not only about whether or not an item is remembered but also about the quality of the memory, reflected in the precision of response distributions around the correct feature value.

The purpose of the present article is to present an interference model of visual WM that can be applied to tests of memory for features in a continuous feature space. The model inherits the core features of interference models that have been successful in accounting for data from paradigms testing memory for discrete items, both over the short term (Jonides et al., 2008; Lewandowsky & Farrell, 2008) and over the long term (Surprenant & Neath, 2009). Specifically, the model incorporates the following principles: (a) Access to contents of visual WM relies on cue-based retrieval. In most contemporary visual WM paradigms, visual objects are presented in different locations of an array, and the test requires remembering which object was in a given location. In these paradigms spatial location is the retrieval cue for accessing representations of WM contents, such as objects or their features. (b) Both memory contents and their potential cues are represented in a distributed fashion, such that their similarity is reflected in the degree of overlap between representations. (c) Memory performance is limited by interference, which can arise from multiple sources. One source of interference is confusion between memory contents that are bound to similar, or overlapping, cues. Another source of interference is the superposition of multiple overlapping memory representations within the same representational medium, so that they blur or distort each other (Oberauer, Farrell, Jarrold, Pasiecznik, & Greaves, 2012).

Measurement Models and Explanatory Models

The literature on visual WM currently offers two kinds of mathematical models that serve different purposes. One kind of model serves primarily as *measurement model*, aiming at measuring latent variables of theoretical interest, such as the number of objects a person can hold in WM, or the precision of feature representations. A second kind of model serves as *explanatory model*, aiming primarily at explaining the variability of behavior—and of latent variables measured by measurement models—across multiple experimental conditions. Reflecting these different aims, measurement models are applied to data from each experimental condition separately, yielding parameter estimates that serve as measures of latent variables in each condition. In contrast, explanatory models are applied across multiple conditions simultaneously with the aim to explain the effects of these conditions with a common set of parameters.

The difference between measurement models and explanatory models can be illustrated with the two models Zhang and Luck (2008) have proposed for the continuous-reproduction paradigm. Their measurement model is a 2-component mixture model that describes the distribution of responses around the true value of the target feature as a probability mixture of two distributions: One distribution reflects responses based on memory of the target feature with a certain precision, and is modeled as a normal distribution around the true value. The dispersion of the normal distribution reflects the precision of memory for the target feature. The other component of the mixture is a uniform distribution reflecting random guessing. Bays et al. (2009) extended the mixture model by a third component reflecting responses based on erroneous recall of nontarget features, that is, features of objects that have been in the memory display but are not the one probed for recall. These mixture models can be used to measure hypothetical processes—such as the precision of memory representations or the probability of guessing—but they do not explain these processes, and why they change across experimental conditions.

Zhang and Luck (2008) also proposed an explanatory model—the Slot-Averaging model—which offers an explanation of the set-size effect on performance in continuous reproduction. The set-size effect refers to the finding that, as the set of objects to be remembered increases, performance declines. This decline is commonly regarded as reflecting the limited capacity of WM. The Slot-Averaging model explains the set-size effect by the assumption that WM has a discrete number K of slots, and as set size increases beyond K, the probability that the target object is not represented in a slot increases. For set sizes below K, multiple slots can be assigned to the same object, and in that case, the precision of report is increased by averaging the output of all slots representing the target object. A competing class of explanatory models are resource models (Bays & Husain, 2008; Palmer, 1990; van den Berg et al., 2012), which assume that a limited resource is divided among all objects to be remembered, so that the average resource share of each object declines with increasing set size.[1]

Here we propose an alternative explanatory model of visual WM based on interference. In our interference model (IM) there is no limited commodity to be distributed among memory representations, be it in the form of discrete slots or a continuous resource. The set size effect must be explained differently. Interference provides a partial explanation: The more representations must be maintained simultaneously in WM, the more each of them suffers from interference by all others. As we will show, interference alone is not sufficient to explain the set size effect in visual WM. Therefore, our model includes a further assumption, based on our

[1] Other explanations of the set-size effect have been proposed (Orhan & Ma, 2015; Orhan, Sims, Jacobs, & Knill, 2014) but they have not yet been developed into explanatory models to be fit to visual WM data, and therefore we do not consider them here.

general theoretical framework for working memory (Oberauer, 2009; Oberauer & Hein, 2012): The working memory system consists of three embedded components serving different functions. One component is the *activated part of long-term memory*, which among other things is responsible for maintaining an above-baseline activation of all currently relevant representations. A second component is the *region of direct access*, which holds available a small number of content elements (e.g., visual objects) by creating and maintaining temporary bindings between each content element and its context (e.g., its spatial location). The third component is the *focus of attention*, which serves to select within the direct-access region one content element (and its context) for processing. Below we specify the role of each component in our mathematical model. We will show that with the addition of the three-component structure, the interference model gives a satisfactory account of the set-size effect in visual WM. Moreover, the addition of the focus of attention enables the model to also explain the effects of precues and retro-cues on performance in visual working memory tests (Lepsien, Thornton, & Nobre, 2011; Souza & Oberauer, 2016), thereby expanding its scope of explanation beyond that of other contemporary models.

The remainder of this article is organized as follows: We first develop the interference model (IM), and briefly introduce two competing explanatory models published to date, the Slot-Averaging model and a resource model, the Variable-Precision model. Next we competitively fit these three models to three experiments designed to challenge the explanatory power of these models, and to test predictions of the IM. We also report a fourth experiment that tests a further unique prediction of the IM. In the final section we develop several new models that implement theoretical ideas that have been discussed in the literature on visual WM, or suggested to us, and compare them to the IM. The raw data from the experiments, as well as the Matlab code for all models, is available on the Open Science Framework at https://osf.io/wgqd5/.

An Interference-Based Explanatory Model

Current explanatory models of WM for visual materials focus on explaining the set-size effect. Therefore, our first aim is to develop an explanatory interference model that explains how performance in a continuous-reproduction task changes as memory set size is increased.

Features and Bindings

The starting point of our modeling effort is the family of 2-layer neural network models that has been used with much success in the literature on (working) memory for lists (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 2006; Lewandowsky & Farrell, 2008; Oberauer, Lewandowsky, et al., 2012; Sederberg, Howard, & Kahana, 2008). The model architecture, illustrated in Figure 1, consists of two layers of units fully interconnected by a matrix of connection weights. One layer represents memory contents (e.g., words or colors), and the other represents their contexts. The context is the information that is used as retrieval cue to retrieve a specific content. In models of list memory, the context is usually a temporal context (i.e., the time at which a list item has been presented, or its serial position in the list). A list is encoded by binding each item (in the content layer) to the time or the list position at which it occurred (in the context layer). Bindings are formed by modifying the connection weights, for instance through a Hebbian learning rule. List items are retrieved by reactivating their context in the context layer, which reactivates the content bound to it in the content layer.

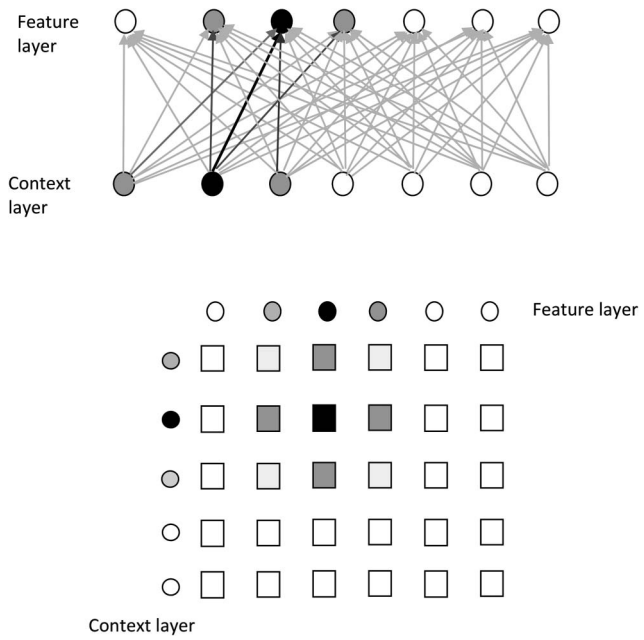

Figure 1. Schema of a 2-layer connectionist network for encoding lists by binding items (represented in the Feature layer) to their list positions (represented in the Context layer). The top panel shows the network structure; the bottom panel shows the same model in matrix layout to facilitate mapping to Figure 2.

In visual WM tasks, memory items are usually presented simultaneously in a spatial array, and the target (i.e., the item to be reproduced) is identified by its spatial location. In that case, spatial location is the retrieval cue, and accordingly, the primary context we need to consider is spatial location. Temporal context probably also plays a role for discriminating the memory set of the current trial from those of previous trials—a point to which we return in the General Discussion—but it cannot play a role to cue retrieval of a specific item in a memory set presented simultaneously. A second special feature of visual WM tasks is that they often use features varying on a single continuous dimension as contents. We can therefore think of both the content and the context layer as a population code of units ordered by their preferred value on the feature dimension and the context dimension, respectively. The representation of each feature (e.g., color) and each context (e.g., spatial location) is a unimodal distribution of activation over the respective layer (see Figure 1), centered on the unit that has the represented feature or location as its preferred value.

The IM models visual WM on a more abstract level than neural-network models (for network models incorporating many of the principles of the IM see Matthey, Bays, & Dayan, 2015; Swan & Wyble, 2014). In the IM we represent context and content as continuous dimensions, which we refer to as *context dimension* and *feature dimension*, respectively. The connection-weight matrix

is accordingly replaced by a continuous 2-dimensional *binding space* (see Figure 2). Features and their contexts are modeled as continuous distributions over feature and context dimension, respectively. In keeping with previous models of the continuous-reproduction task, we describe representations of individual features as a von-Mises distribution on the feature dimension. The von-Mises distribution is a normal distribution on the circle, which is appropriate for circular feature dimensions such as colors in a color wheel, or orientations. The precision of feature memory is governed by the concentration parameter κ, which is inversely related to the variance of the von-Mises distribution. Encoding consists of binding each item's feature representation to the corresponding context representation; this generates a bivariate distribution of binding strengths in binding space (Figure 2A). The binding-strength distribution has limited precision on both the feature and the

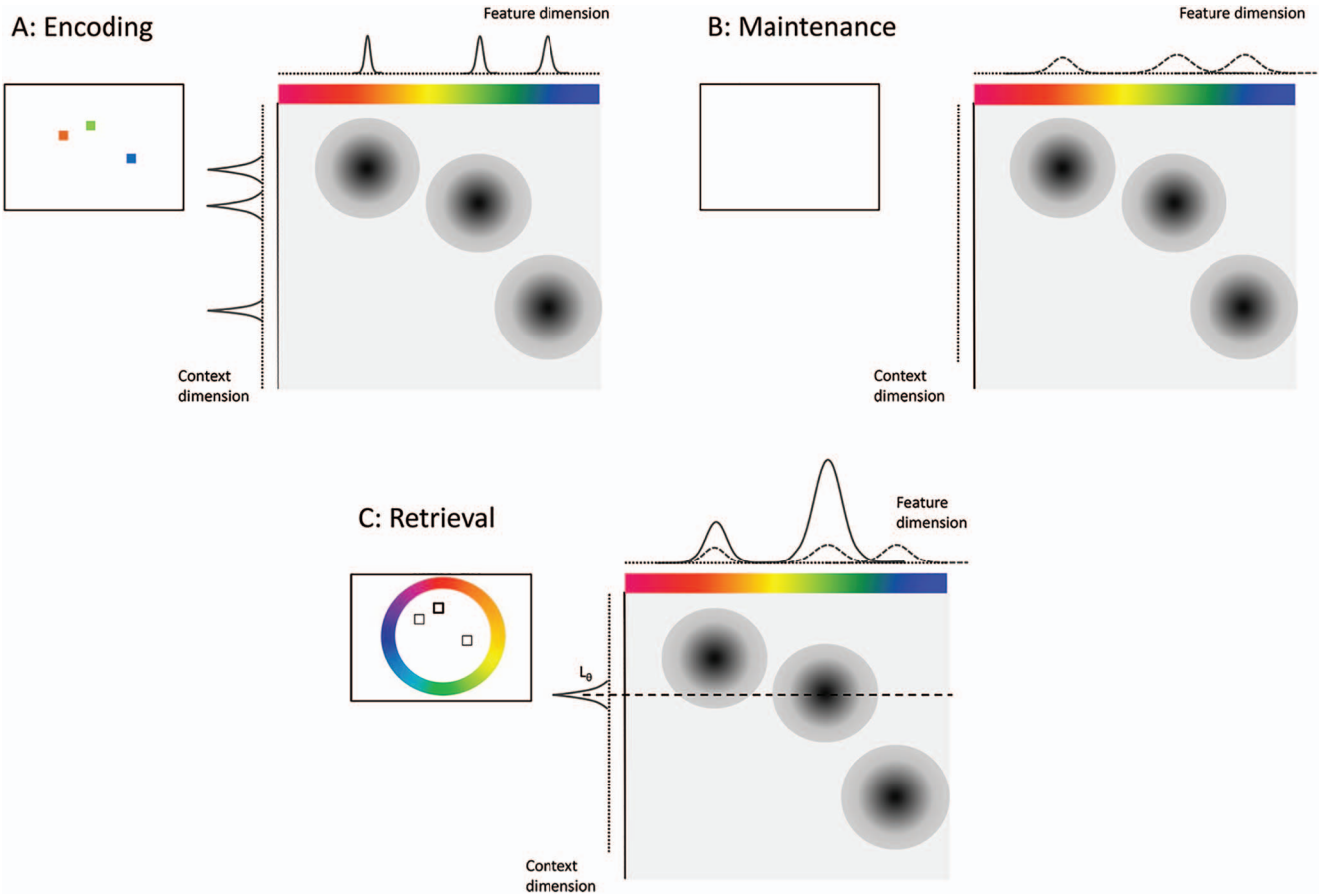

Figure 2. Schema of the hypothetical memory system underlying the interference model, displayed at three moments during a trial. A: During encoding, perceptual representations of contents (e.g., colors) are distributions of activation in feature space (horizontal); representations of context (e.g., spatial location) are distributions of activation in context space (vertical, on the left). Both representations are accompanied by a low level of baseline noise modeled as uniform (dotted lines). Bindings between contents and their contexts are shown as bivariate distributions of binding strength in binding space. As long as they are supported by sensory input, representations in feature space and context space are highly precise, but bindings are represented with lower precision. The baseline noise in feature and context space translates into uniform noise in binding space (light gray background). B: During maintenance, activations in feature and context space dissipates, but background noise remains. The binding information is fully maintained in binding space. Background noise in context space feeds through binding space into feature space, reactivating low-precision distributions of the memory features (broken lines). C: At retrieval, the target location L_θ is activated in context space with high precision. Its activation in context space is multiplied with the binding strength in binding space to re-create a profile of activation in feature space, corresponding to a horizontal cross-section of binding space at the target location. The resulting activation profile in feature space consists of three components: The background noise projected from binding space into feature space (uniform distribution, dotted line), the activation of features generated by background noise in context space, which contributes information about features independent of their context (broken lines), and the activation from cue-based retrieval, which provides information about the feature at the target location, and its spatial neighbors (continuous lines). See the online article for the color version of this figure.

context dimension, which limits the precision with which a feature can be reactivated, given its context.

In the IM we do not model the binding space explicitly—rather, the IM models the distribution of activation that is recreated in the feature dimension at test. The values on the feature dimension (or a subset of them) are the retrieval candidates, and the activation distribution reflects evidence from memory in favor of choosing each candidate. In the IM each recall response is a probabilistic choice from the set of candidate responses, based on the relative strength of activation $A(x)$ of each candidate x at retrieval. The probability of choosing each response x out of N response options is given by Equation 1, which normalizes the activation distribution.

$$P(x) = \frac{A(x)}{\sum_{j=1}^{N} A(j)} \tag{1}$$

The activation distribution over response candidates generated at retrieval is a weighted sum of three components:

$$A(x \mid L_\theta) = cA_c(x \mid L_\theta) + aA_a(x) + bA_b(x) \tag{2}$$

First, each feature value is activated to the extent that it is bound to the given context cue. This component is weighted by parameter c. It depends on the cue indicating the location of the target, hence it is conditional on the target location, L_θ. The second component is information about which feature values were in the current memory set, irrespective of their context (e.g., all colors in the current array, regardless of their locations). At retrieval, all candidates matching any array feature receive activation from this source of information, the strength of which we represent by the parameter a. The third component is background noise in memory, which contributes equal activation to all recall candidates; we represent its strength by parameter b. We next explain each component in detail.

Component A_c reflects the activation from cue-based retrieval. Cue-based retrieval starts from a representation of the cued context in context space, which generates a distribution of activation in feature space through the bindings in binding space. Each feature receives activation according to the strength of its binding to the context cue. Thus, the feature of the item that had been in the cued context is activated most strongly. Because of the limited precision of bindings along the context dimension, a retrieval cue indicating the target location in context space also reactivates features of other items in the memory set to the degree that they are bound to contexts similar to the cued one. For instance, visual objects presented in close spatial proximity to a cued object are strongly coactivated by a spatial cue, and therefore are more likely to be recalled erroneously than more distant objects. Evidence for such a spatial gradient on intrusions from nontarget items has been found in both recognition and recall from visual WM (Rerko et al., 2014).

We model the A_c component as a sum of the activation distributions across all n items in the memory set, weighted by how strongly each item is cued through its binding to the context cue:

$$A_c(x \mid L_\theta) = \sum_{i=1}^{n} \exp[-s \cdot D(L_i, L_\theta)] \cdot vMises(x, x_i, \kappa) \tag{3}$$

Here and in subsequent equations, we use *vMises* as a shorthand to refer to the von-Mises density:

$$vMises(x, x_i, \kappa) = \frac{\exp(\kappa \cdot \cos(x - x_i))}{2\pi \cdot I_0(\kappa)} \tag{4}$$

where I_0 is the modified Bessel function of order 0.[2]

Following precedents (e.g., Brown, Neath, & Chater, 2007), we model cueing strength as exponentially declining with the distance D between the location of item i in context space, L_i, and the location that serves as retrieval cue. In the present experiments, the retrieval cue is the location of the target, L_θ, which is given as the recall probe. Hence, for the target item $D = 0$, so that the von Mises distribution reflecting that item's feature enters with weight 1 into A_c. For nontarget items, the distance is larger than zero, resulting in a smaller weight for their feature distributions. How steeply the weight declines with distance in context space is governed by the spatial gradient parameter s. Thus, s reflects the precision of bindings along the context dimension; smaller values of s reflect bindings that are more spread out in the context dimension, so that the effect of cue-based retrieval spreads more broadly to neighboring items.[3]

The activation component A_a reflects the location-independent feature information from all items in the memory set. Location-independent information about item features could be maintained through sustained activation of feature representations in feature space (e.g., sustained firing of feature-coding neurons in sensory cortex, Emrich, Riggall, LaRocque, & Postle, 2013), or by associating those representations to a global context of the entire memory set, such as a temporal context, or a location-unspecific (e.g., uniform) distribution of activation in context space. The information about the feature of each item i is modeled as a von Mises distribution in feature space, centered on the true feature of that item, x_i, for instance its actual color or orientation in the memory display. A_a is the unweighted sum of the activation over all items in the memory set.

$$A_a(x) = \sum_{i=1}^{n} vMises(x, x_i, \kappa) \tag{5}$$

Finally, component A_b reflects the background noise, which captures all contributions to activation that carry no information about features of items in the present memory set. Background noise could have several sources, including enduring activation of stimuli encountered in all previous memory sets (for related proposals in recognition models see Brockdorff & Lamberts, 2000; Nosofsky, Cox, Cao, & Shiffrin, 2014), or biases unrelated to the current memory set. Because all model predictions are for deviations of responses from the target or the nontargets in the memory set, and all sources of background noise are by definition uncorrelated to items in the memory set, we can simplify the model and

[2] Equations 3 to 6 give the activation density in a continuous feature space. In practice, the feature space is discretized into a finite number N of response options by the resolution of the response scale. For instance, in the experiments reported below, participants chose their responses from a response space subdivided into $N = 360$ colors or orientations.

[3] The idea that nontarget intrusions depend on the nontarget's distance to the target has first been explored with a variant of the 3-parameter mixture model mentioned in the Supplementary Material of van den Berg et al. (2012). This model assigns a separate free parameter for the mixture weight of nontargets at each possible spatial distance from the target. In contrast, the IM imposes a stronger constraint on the spatial gradient and models it with a single free parameter.

represent background noise as a uniform distribution on the circle, assigning equal activation to all response options. In the IM we assume that each item in a memory array is associated with a constant amount of background noise. Therefore, component A_b increases linearly with set size n:

$$A_b(x) = \frac{n}{2\pi} \quad (6)$$

Each object's representation in feature space is bound to that object's representation in context space. This binding process includes the background-noise components in the feature and the context representation, thereby creating noise in binding space, which we represent as a bivariate uniform distribution. Each additional objects adds to the background noise in binding space, so that the contribution of background noise to the activation from memory, A_b, increases linearly with set size. Appendix A presents one possible mechanistic explanation of how location-independent activation of features (component A_a) and uniform background noise (component A_b) arise in the visual WM system.

We assign each array item its own background noise because we assume that background noise is added to each item representation already in an early, location-specific stage of visual processing (this could be a retinotopically organized feature map, Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; or a fragile, location-specific visual short-term store, Pinto, Sligte, Shapiro, & Lamme, 2013). The representations in each location carry their own background noise. Thus, background noise is added to each item representation already before, at a later stage of visual processing, it is projected into a common feature space and bound to its location in context space. We make no specific assumption about the source of background noise. One possibility is that it arises from attention to the entire target feature dimension (Müller, Reimann, & Krummenacher, 2003; Müller et al., 2010). For instance, when colors are the features to be remembered, participants attend to the color dimension, which could raise the tonic activation or gain of all color-coding neurons, increasing the chance of these neurons firing independently of the stimulus color (Pollmann, Weidner, Müller, & von Cramon, 2006; Töllner, Müller, & Zehetleitner, 2012). Another possibility is that background noise reflects a more passive (i.e., attention-independent) form of priming of all feature values due to the fact that all possible feature values are presented frequently across trials (Kruijne, Brascamp, Kristjánsson, & Meeter, 2015; Thomson & Milliken, 2013).

Readers might wonder at this point why explicit bindings of features to locations in context space are necessary if features are already implicitly bound to their locations in a spatiotopic map. Implicit bindings in spatiotopic feature maps can only bind features to spatial locations, whereas visual WM can also maintain and use bindings of features to nonspatial contexts. For instance, colors can be used as retrieval cues for orientations bound to them (Gorgoraptis, Catalao, Bays, & Husain, 2011; Kalogeropoulou, Jagadeesh, Ohl, & Rolfs, 2016, and the present Experiment 4), and pictures of objects can be used as retrieval cues for colors bound to them (Brady, Konkle, Gill, Oliva, & Alvarez, 2013). Therefore, visual WM needs a more general binding mechanism than the implicit binding to spatial locations provided by feature maps.

The Focus of Attention in Visual Working Memory

One additional assumption in the IM draws on our earlier work on the architecture of working memory (Oberauer, 2003, 2009; see also Olivers, Peters, Houtkamp, & Roelfsema, 2011): The working memory system includes a focus of attention that holds a single object (including its context, such as its location in the array). During encoding, the focus is directed to one object in one location, and that object's feature (on the to-be-recalled dimension) is encoded with a higher precision than all other objects. Therefore, the IM has two precision parameters, κ for representations outside of the focus and κ_f for the representation of one object i in the focus.

In the absence of a precue we assume that the focus selects one object in the memory array at random and holds it until the test. Therefore, the probability that a randomly selected test target θ is in the focus of attention equals $P(F_\theta) = 1/n$. Thus, in the continuous-reproduction task the chance of reporting the target's feature with heightened precision decreases hyperbolically with increasing set size. If the target happens to be the one encoded and maintained with precision κ_f in the focus of attention, activation component A_c is augmented by an additional component reflecting information from the focus of attention, A_f.

$$A_f(x) = vMises(x, x_i, \kappa_f) \quad (7)$$

As long as the focus of attention remains on the object that was focused during encoding, that object's feature is represented with precision κ_f, but if the focus shifts to another object, the previously focused object loses its extra precision and remains represented outside the focus with precision κ. Therefore, even when the memory array is presented long enough to enable the focus to sequentially scan several or all objects, only one of them—the one encoded last—can be maintained with increased precision.

If an object is encoded with precision κ_f, but then the focus shifts away from it during the retention interval, then none of the objects is represented with heightened precision. If the focus shifts to an object during the retention interval, that object does not gain in precision, because a gain in precision implies a gain of information, and in the absence of informative sensory input it is hard to conceive where that information should come from. Yet, there is a solid body of evidence that directing attention to an item in WM after encoding improves memory for that item: When a retro-cue identifying the target's location in the array is presented during the retention interval, one to several seconds after offset of the memory array, then memory for the target is improved compared to a condition where the target location is identified only at the time of testing (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; for a review see Souza & Oberauer, 2016). This so-called retro-cue effect consistently increases the probability of recalling the correct object's feature, and has less consistent effects on the precision (Gunseli, van Moorselaar, Meeter, & Olivers, 2015; Murray, Nobre, Clark, Cravo, & Stokes, 2013; Souza, Rerko, Lin, & Oberauer, 2014; Souza, Rerko, & Oberauer, 2016). Therefore, focusing attention to an item in WM must have an effect over and above maintaining a high-precision representation.

To incorporate this second effect of the focus of attention we assumed that when the target is in the focus of attention, the contributions of nontargets (A_a) and of background noise (A_b) to the activation is reduced (see Appendix A for a possible mecha-

nism). The proportional reduction of A_a and A_b for a focused target (relative to the standard case of a nonfocused target) is captured by a further free parameter r. This reduction applies regardless of whether the target is focused already during encoding or only later during the retention interval, for instance in response to a retro-cue indicating the target's location.

To incorporate the assumptions outlined above about the focus of attention, Equation 2 is modified as follows:

$$\begin{aligned} A(x\,|\,L_\theta, \neg F_\theta) &= cA_c(x\,|\,L_\theta) + aA_a(x) + bA_b(x).\\ A(x\,|\,L_\theta, F_\theta) &= cA_c(x\,|\,L_\theta) + cA_f(x) + rbA_b(x) + raA_a(x).\\ A(x\,|\,L_\theta) &= (1 - P(F_\theta)) \cdot A(x\,|\,L_\theta, \neg F_\theta) + P(F_\theta) \cdot A(x\,|\,L_\theta, F_\theta). \end{aligned} \quad (8)$$

The first line of Equation 8 reflects the activation in case the target is not in the focus of attention (denoted $\neg F_\theta$), and the second line reflects the activation in case the target happens to be in the focus of attention (F_θ). In the latter case, the contribution of noise and of nontarget features is reduced by factor r, and the low-precision activation from cue-based retrieval, A_c, is augmented by the high-precision activation from the focus of attention, A_f. Activation from the focus is added to activation from cue-based retrieval, rather than replacing it, because we assume that cue-based retrieval is obligatory. The last line of Equation 8 combines the expressions for the two cases, weighting the first with the probability that the target is not in the focus of attention, and the second with the probability that the target is in the focus. In the absence of reasons to believe that the focus of attention is biased toward a particular item or subset of items, the probability of the target to be in the focus is $P(F_\theta) = 1/n$.

In summary, the IM has seven parameters: *a, b, c*, κ, κ_f, *s*, and *r*. Parameter estimation requires that one of the three weight parameters (*a*, *b,* and *c*) is fixed to set the scale of the other two. For mathematical convenience we fixed $c = 1$. Thus, the IM has six free parameters.

There are two equally valid theoretical interpretations for the reduction parameter *r*. One is that it reflects an actual reduction of the contribution of nontarget information and noise to the activation when the focus is directed to the target during the retention interval, that is, a reduction of parameters *a* and *b*. For instance, directing focused attention to an item's location in the array might imply inhibition of all information unrelated to that location, or removal of nontargets from WM (Souza, Rerko, & Oberauer, 2014; Williams, Hong, Kang, Carlisle, & Woodman, 2013). The alternative interpretation is that focusing on an item strengthens its binding to its context, that is, its location in the array (Rerko & Oberauer, 2013; Souza, Rerko, & Oberauer, 2015). Strengthening an item's binding to context facilitates access to that item through cue-based retrieval, and in the IM that would be reflected in an increased value of the *c* parameter. Because the model's behavior depends only on the relative size of parameters *a*, *b*, and *c*, a reduction of *a* and *b* is equivalent to an increase in *c*. Because we fixed $c = 1$ as the scaling parameter for *a* and *b,* a strengthened content-context binding for a focused relative to a nonfocused target can only be reflected in a (proportionally equal) reduction of *a* and *b*.

With the addition of the single-item focus of attention, the structure of the IM maps onto the architecture of the three-embedded components framework of WM (Oberauer, 2009). The framework describes WM as consisting of three embedded components: The activated part of long-term memory (LTM), the region of direct access, and the focus of attention. The activated LTM has several roles, among them to keep a subset of LTM representations that are likely to be relevant for ongoing cognition activated above baseline, thereby making them relatively easy to retrieve. In the IM, the activation of the relevant set of representations is reflected in the context-independent component of activation, A_a. The direct-access region is responsible for creating and holding a limited set of temporary bindings between contents and their contexts. This function is reflected in the cue-based retrieval component of activation in the IM, A_c. Cue-based retrieval is based on temporary bindings between objects (or their features, such as color) and their retrieval-relevant context in the current memory array (e.g., their spatial location). These bindings are temporary—they are set up within about 100 ms per object (Vogel, Woodman, & Luck, 2006; Woodman & Vogel, 2008), and need to be reset in every trial. The focus of attention in the three-embedded components framework has the function to select an item as the input for a cognitive or overt action (e.g., recall), or as the object of manipulation. Under most circumstances this selection function is best served by selecting one item at a time, because the focus of attention as conceptualized in this framework does not have a binding mechanism—hence, selecting multiple items involves the risk of blending or confusing them (Oberauer, 2013; Oberauer & Hein, 2012) In the context of the IM we make the additional assumption that a visual object is represented with higher precision as long as it remains in the focus of attention since its encoding.

Fitting the Interference Model to Experimental Data

We first fit the IM to data from three experiments that we carried out to collect benchmark data for model comparisons. The experiments are described in detail in Appendix B. All three experiments tested recall of colors using a continuous color wheel as response scale (Wilken & Ma, 2004; Zhang & Luck, 2008). Each memory array consisted of a variable number of colored squares placed on a randomly selected subset of 13 equidistant locations on a virtual circle centered on the screen center. We measured spatial distance *D* between two objects as their angular distance in radians (i.e., ranging from 0 to π). In a fourth experiment we generalized the IM to the recall of orientations, and tested the unique prediction of the IM that targets are most likely to be confused with their neighbors in context space.

Overview of Experiments

In each trial of Experiments 1 to 3 participants saw an array of a variable number of color patches and were asked to remember the colors and their locations. The color patches disappeared after a brief delay. After a 1-s retention interval, empty frames were presented at the locations of color patches, and the target location was marked by a thicker frame. Participants reproduced the target's color by selecting it from a color wheel.

Experiment 1 varied set size through all levels from 1 to 8. Set size is the most important independent variable for studying capacity limits of WM because the capacity limit is revealed by the decline of performance with increasing set size. The detailed aspects of this set-size effect have been used to support or challenge competing claims about the nature of the capacity limit (Bays & Husain, 2008; Mazyar, van den Berg, & Ma, 2012; Zhang & Luck, 2008).

Experiments 2 and 3 varied set size again over a wide range (1, 2, 4, 6, 8), and in addition included cues pointing to one of the visual objects as the one most likely to be tested (Griffin & Nobre, 2003; Zhang & Luck, 2008).[4] The cueing manipulations served to control the orientation of the focus of attention, which plays an important part in the IM. In Experiment 2 we presented a precue preceding display of the memory array, and in Experiment 3 we used a retro-cue following offset of the memory display but preceding the test. In both experiments, the cue indicated the target object with 67% validity. In addition to the two cueing conditions (valid cues and invalid cues) there was a neutral condition in which an uninformative stimulus was presented instead of the cue. In the cue conditions of Experiment 3, the retro-cue was presented 1 s after offset of the memory display, and the probe was presented 0.3 s after the cue. In contrast, in the no-cue condition the probe appeared 1 s after the memory display, at the same time as the cue in the cue conditions. Therefore, any beneficial effect of a valid retro-cue cannot be explained by protection of the cued item from decay.

Experiment 4 served to test a key assumption of the IM: because of the limited precision of item-context bindings, retrieval cues reactivate not only the target feature but also to some extent features of nontarget objects that are close to the target object on the retrieval-relevant context dimension (e.g., spatial location). In principle, any feature dimension could be used in the role of the context dimension. For instance, Gorgoraptis et al. (2011) presented sets of oriented bars sequentially and probed the target orientation by its unique color. In that experiment, color served as the context dimension, whereas orientation was the dimension of the feature to be recalled.

In Experiment 4 we used two kinds of retrieval cues, location and color, and asked participants to report orientation. We chose color as a second retrieval cue because a previous study showed that color and location were approximately equally good retrieval cues for orientations (Pertzov & Husain, 2014). Participants saw arrays of 6 disks, each with a rectangular gap (Anderson, Vogel, & Awh, 2011). The orientations of the gaps (chosen at random from a uniform distribution of 360 possible orientations) was the feature to be remembered. The disks were distributed among 6 randomly selected locations out of 13 equally spaced locations on an invisible circle. Analogously, the disks had 6 different colors, selected at random from 9 colors equally spaced on the color circle (i.e., the same color circle that served as the response dimension in Experiments 1 to 3). At test the target (i.e., the disk the orientation of which was to be reported) was identified by its location, its color, or a combination of location and color. The IM predicts that nontarget intrusions tend to come from spatial neighbors of the target when location was used as a retrieval cue, and from color neighbors when color was used as a cue.

Explaining the Set-Size Effect

To investigate the ability of the IM to explain the set-size effect we fit the IM, as well as a number of competing explanatory models, to the data from Experiment 1. For this and the subsequent model comparisons we selected the competing explanatory models by two criteria: First, they must have a closed-form expression of the likelihood of continuous-reproduction data so that they can be fit to data of individual subjects with reasonable computational costs. Second, they must be interpretable in terms of theoretical mechanisms and processes, so that they actually offer a theoretical explanation of the phenomena in question, rather than a mere statistical description of the data.

We first compare the IM with two published alternative explanatory models meeting these criteria: The Slot-Averaging model (Zhang & Luck, 2008), and the Variable-Precision model (van den Berg et al., 2012). The Variable-Precision model includes the original Resource model (Bays & Husain, 2008) as a special case, and therefore we do not report fits of the original Resource model. We next briefly introduce these models. In a later section we test the IM against a number of improved variants of the published models.

Slot-Averaging model. The Slot-Averaging model assumes that working memory has a limited number of slots, K, for storing visual objects. If the set size exceeds K, the remaining objects are not stored at all, and if one of them is tested, the person has to guess by drawing a response at random from a uniform distribution. If K is larger than the set size, some or all of the objects are stored in more than one slot, and at test the information from all the slots coding the target are read out, and their information is averaged. Averaging the information from several slots increases precision because pooling independent noisy samples representing the same information reduces noise. Specifically, the standard deviation of the response follows the equation

$$\sigma = \frac{\sigma_1}{\sqrt{m}}, \tag{9}$$

where σ_1 is the standard deviation of information in a single slot, and m is the number of slots assigned to the target.

The Slot-Averaging model has two free parameters, K and σ_1. K was allowed to vary continuously between 0 and the 8 (the largest set size in our experiments); we interpret noninteger values as averages across trials over which the number of slots could vary. We assumed that slots are allocated as evenly as possible to objects, such that K objects are represented in one slot each when $K \leq n$. In case $K > n$, we calculated two possible values of m by rounding K/n up and down to the nearest integer, computed σ for both values of m, and averaged these values, weighted by the probability that a randomly chosen target had the higher or the lower number of slots. For instance, with $K = 3.5$ and $n = 2$, the number of slots on any one trial is 3 or 4 with equal probability; in the first case, the target has an equal chance of receiving $m = 1$ or $m = 2$ slots, and in the second case it always receives $m = 2$ slots. Across trials, $p(m = 1) = .25$ and $p(m = 2) = .75$, and the two values of σ computed from Equation 9 are weighted accordingly. The model likelihood is a probabilistic mixture of a von-Mises distribution with standard deviation σ centered on the target and a uniform distribution; the weight of the von-Mises distribution is the probability that the target is represented in at least one slot, $min(1, K/n)$.

Variable-Precision model. In line with the Resource model by Bays and Husain (2008), the Variable-Precision model (van den

[4] The cues presented in these experiments must not be confused with the retrieval cues discussed so far. Precues (presented before the memory display) and retro-cues (presented in the retention interval between encoding and test) precede the probe given at test that serves as the retrieval cue. When precues and retro-cues are valid, they point ahead of time to the target item that is eventually indicated by the probe, but when they are invalid, they point to a nontarget.

Berg et al., 2012) starts from the assumption that WM has a limited resource quantity that is flexibly assigned to all representations that need to be maintained. The resource quantity assigned to each representation determines its precision. Whereas the Resource model assumes a fixed precision for a given participant and set size, the Variable-Precision model assumes that the precision with which each item is represented varies randomly both within and between trials. The precision of each representation is formalized as the Fisher information J. For a Gaussian normal distribution, $J = 1/\sigma^2$. For the von Mises distribution this is true only to a close approximation, but the relationship between J and σ^2 clarifies how the Variable-Precision model is related to other resource models. The average Fisher information for a given item depends on its resource share R according to

$$\begin{aligned} \bar{J} &= \bar{J}_1 R^{\alpha}; \\ \bar{\sigma}^2 &\approx \frac{\bar{\sigma}_1^2}{R^{\alpha}}; \end{aligned} \tag{10}$$

where $\bar{J}_1$ refers to the mean precision, and $\bar{\sigma}_1^2$ to the mean variance, at set size 1. When there is no reason to assume any bias in the resource assignment to specific items, as in our Experiment 1, $R = 1/n$.

To implement variability of precision, for each individual item in a given trial, J is drawn from a Gamma distribution with mean $\bar{J}$ and scale parameter τ. When fitting the model, we simulated the precision distribution by drawing 5000 samples of J from a Gamma distribution with the current estimates of $\bar{J}$ and τ, computed the model's likelihood function—a von-Mises centered on the target feature with precision set by the sampled value of J—and averaged it across the samples. The model's three free parameters are $\bar{J}$, τ, and α.

An explanatory model explains a phenomenon only if it incorporates a theoretical explanation of that phenomenon, as opposed to a merely descriptive mathematical function capturing it. For the Slot-Averaging model, the theoretical assumption of a constant number of slots, together with the assumption of averaging during read-out, directly leads to the model's mathematical form. The relation between theoretical assumptions and formalization is less straightforward for the Variable-Precision model. Why should the relation between R and σ be a power function? Bays and Husain (2008) offer an explanation based on the sample-size model of resources (Lindsay, Taylor, & Forbes, 1968; Palmer, 1990): Each item's feature is represented by a sample of noisy codes. Informally, we can think of each code as the information encoded in a single neuron. The resource is the total number of coding units available in the system, which must be divided among the items in the memory set, so that R is the proportion of coding units assigned to an individual item. Assuming uncorrelated noise, the inverse of the variance of the mean signal across all units coding the same feature, $1/\sigma^2$, increases linearly with R. Therefore,

$$\sigma^2 = \frac{\sigma_1^2}{R}. \tag{11}$$

Thus, the sample-size model implies that variance is related to $1/R$ by a power function with a power constant of 1. By introducing α as a free parameter, Bays and Husain (2008) made their Resource model more flexible than the sample-size model. In their Footnote 20 they justify this step by pointing out that variability of neural firing rates is often correlated; the free parameter is assumed to relax the assumption of uncorrelated neuronal variability. We argue that this justification is not sound. The variance of the mean of m correlated variables is not related to m through a power function with a free parameter α. Rather,

$$\sigma^2(\bar{x}) = \frac{\sigma^2(x)}{m} + \frac{m-1}{m}\rho\sigma^2(x), \tag{12}$$

with ρ for the correlation. A more sophisticated alternative comes from modeling work by Moreno-Bote et al. (2014) who investigated the effect of correlated noise in biologically realistic models of neural networks. They found that the loss of Fisher information due to correlated noise is given by:

$$J = \frac{J_0}{1 + \varepsilon J_0}, \tag{13}$$

with J and J_0 for the Fisher information with and without reduction by correlated noise, respectively, and ε for the weight of the information-reducing correlation.

To conclude, the power function with a free parameter α in the Variable-Precision model (and the original Resource model) has no theoretical foundation; it appears to be motivated primarily by the observation that such a function fits the decline of $1/\sigma$ with increasing set size well (Bays & Husain, 2008). We nevertheless fit the Variable Precision model in its published form, including the free parameter α. In addition, we fit versions of the Variable Precision model in which we fixed $\alpha = 1$, and used Equation 12 (with ρ as a free parameter) or Equation 13 (with ε as a free parameter) to incorporate correlated noise.

Competitive model fitting. We fit the explanatory models to individual trial responses of Experiment 1 separately for each participant.[5] We evaluated relative model fit through the Akaike Information Criterion (AIC), which combines $-2log$ (*Likelihood*) with a penalty for the number of free parameters (Akaike, 1974). The choice of AIC as fit index for this and all subsequent model comparisons was based on a model recovery study (described in Appendix C), which showed that for the class of models investigated in this article, AIC recovered the data-generating model better than the Bayesian Information Criterion (BIC; for converging results see Donkin, Nosofsky, Gold, & Shiffrin, 2015; van den Berg et al., 2014).

The fit indices are summarized in the first column of Table 1. The IM fit the data best, followed by the Variable-Precision model in its original version. The Variable-Precision model versions with correlated-noise Equations 12 or 13 fit worse than the original (by 968 and 38 AIC points, respectively), and we therefore do not consider them further. Both the IM and the Variable-Precision model far outperformed the Slot-Averaging model. Table 2 summarizes the best-fitting parameter estimates for the IM, averaged across participants. We conclude that the IM is a viable account of the data of Experiment 1, and the Variable-Precision model remains a reasonable competitor, whereas the Slot-Averaging model is decisively ruled out.

[5] Models were fit with the *fminsearch* function of Matlab, keeping the best result of 10 runs with different starting values. The starting values for the first run were set by hand to values found to converge to a good fit, and those for the remaining nine runs were drawn from a uniform distribution over a broad but plausible range.

Table 1
Model Fits for Explanatory Models Applied to Experiments 1, 2, and 3

Model	Experiment 1		Experiment 2		Experiment 3	
	N(Param)	ΔAIC	*N*(Param)	ΔAIC	*N*(Param)	ΔAIC
Slot-Averaging	2	1739	6	2092	6	2461
Slot-Averaging (3C)			3 × 2	3361	3 × 2	2366
Variable Precision	3	877	7	695	7	1401
Variable Precision (3C)			3 × 3	662	3 × 3	1390
Interference	6	0	6	0	6	0

Note. *N*(Param) is the number of free parameters, and ΔAIC is the difference of the Akaike Information Criterion (AIC, summed across participants) from the best-fitting model (lower values indicate better fit). "3C" refers to models applied separately to the 3 cueing conditions of Experiments 2 and 3.

The left column of Figure 3 presents the mean deviation of responses from the correct color for the eight set-size levels of Experiment 1, overlaid by the predictions of the three competing models with their best-fitting parameters. The figure shows that all three models capture the increase of response deviations with increasing set size. Whereas that increase was approximately linear in the data, all three models predict a decelerated increase. The second column of Figure 3 shows the response distributions from set size 6 of Experiment 1, centered on the target color, together with fits of the three models.[6]

The right column of Figure 3 shows distributions of the same responses, centered on the colors of the nontargets. These plots show why the IM fit the data better than the other two models: The IM is the only model that accurately predicts the distribution of responses centered on nontargets. People report the features of nontargets more often than chance (Bays et al., 2009; van den Berg et al., 2014), and the IM is the only model that has a mechanism for explaining this tendency: Activation of the features of nontargets arises from the context-independent activation component A_a, and from the fact that bindings have limited precision along the (spatial) context dimension, so that a probe at the spatial location of the target also cues nontargets spatially close to the target.

To conclude, the IM passed a first test in accounting for continuous-reproduction data with a parametric variation of set size better than competing explanatory models. It is worth noting that the IM fit the data not much worse than the three-parameter mixture model (ΔAIC = 47). The three-parameter mixture model is a measurement model that fits the response distributions—including responses related to nontargets—separately at each set size (Bays et al., 2009). As such, it is free to adjust its parameters to capture any effect of experimental conditions, such as set size, perfectly. We can assess how well an explanatory model such as the IM explains the effect of the experimental manipulation by comparing its fit to that of the best available measurement model (i.e., the three-parameter mixture model). The comparison of the IM to the mixture model shows that the IM's mis-prediction of the set-size effect on response deviations apparent in Figure 3 comes at only a small cost in terms of quantitative fit.

We next turn our attention to three tests of predictions of the IM that distinguish it from the competing models considered so far. The first two prediction pertains to the basic mechanisms of the IM, which entails opposite effects of interitem similarity on the feature dimension and on the context dimension. The third set of predictions arises from the assumption of a high-precision focus of attention in the IM.

Table 2
Parameter Estimates for Interference Model

Parameter	Experiment 1		Experiment 2		Experiment 3	
	M(*SD*)	Median	*M*(*SD*)	Median	*M*(*SD*)	Median
b	0.16 (0.11)	0.15	0.15 (0.15)	0.11	0.12 (0.10)	0.08
a	0.21 (0.15)	0.14	0.23 (0.16)	0.16	0.19 (0.15)	0.16
s	7.7 (5.7)	5.6	7.9 (6.0)	5.23	3.9 (1.8)	3.5
SD	22.2 (2.7)	21.6	22.0 (5.1)	20.3	21.1 (1.7)	20.9
SD_f	9.1 (2.1)	8.7	9.1 (1.1)	8.7	9.0 (1.8)	8.9
r	.12 (.17)	.0	.12 (.14)	.06	.56 (.21)	.57

Note. Model parameters for the precision of memory outside (κ) and inside the focus of attention (κ_f) have been converted into the corresponding circular standard deviations (in degrees), *SD* and SD_f, respectively, using the *k2sd* Matlab function provided by Paul Bays: http://www.paulbays.com/resources.htm.

Testing Predictions From the Interference Model, 1: The Effect of Similarity Between Targets and Non-Targets

The core assumption of the IM, which distinguishes it from most other models of visual WM, is that access to individual memory contents depends on cue-based retrieval, which gives rise to interference. Interference arises when the retrieval cue does not uniquely reactivate the target representation but also, to some extent, other information in WM. As a consequence, the retrieved information is a superposition of the target representation with nontarget representations and noise (i.e., their weighted sum in feature space). Superposition distorts the distribution of activation in feature space generated at retrieval: With a single item in memory, the activation distribution is narrowly centered on the item's true feature. In contrast, with multiple items in memory, activation of features of nontargets are mixed into the activation distribution, which is thereby broadened and distorted away from the target feature.

One characteristic of interference by superposition is that its effect is less severe when the superimposed content representations are similar to each other than when they are dissimilar,

[6] Analogous plots for other set sizes show the same pattern and are therefore not shown to avoid redundancy.

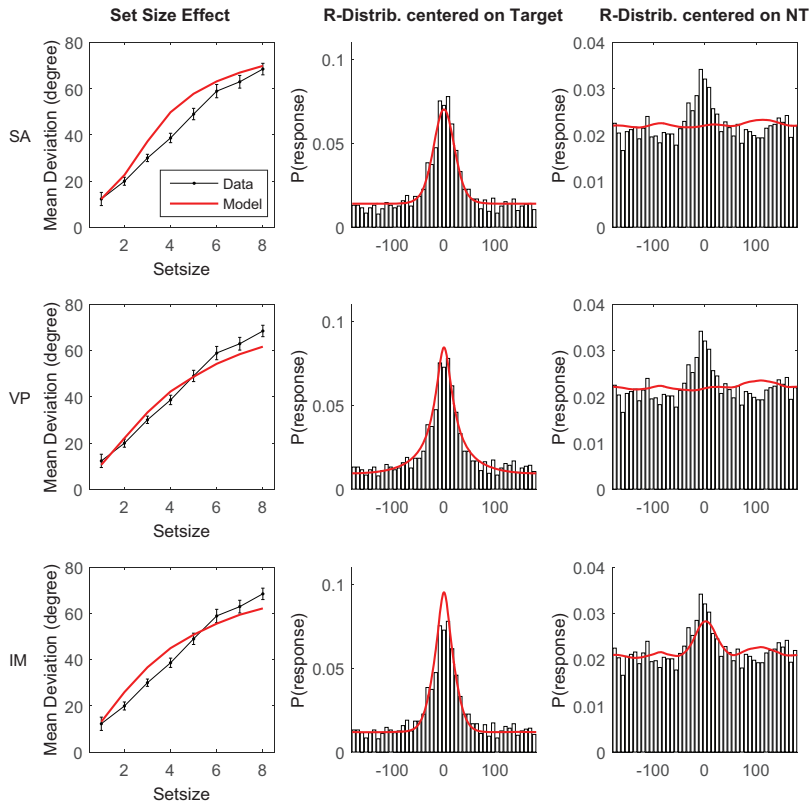

Figure 3. Left column: Mean (absolute) deviation of responses from true color as a function of set size in Experiment 1; experimental data are plotted as black dotted lines, with error bars reflecting 95% confidence intervals for within-subject comparisons (Bakeman & McArthur, 1996). Model predictions from the Slot-Averaging model (SA), the Variable-Precision model (VP), and the Interference Model (IM) are overlaid as continuous red lines. Middle colum: Response distributions from Experiment 1, set size 6, centered on the target color, with predictions from the three models (red lines). Right column: The same response distributions centered on nontargets. These distributions were generated by centering the response distributions on each nontarget in turn, adding them for all nontargets, and dividing by the number of nontargets. See the online article for the color version of this figure.

because the degree of mutual distortion increases with the difference between representations. A prediction following from this fact is that distractors inadvertently entering WM are less damaging when they are similar to the to-be-remembered items than when they are different. This prediction has been confirmed for verbal WM (Oberauer, Farrell, et al., 2012) and for visual WM (Bona & Silvanto, 2014; Huang & Sekuler, 2010; Magnussen & Greenlee, 1999). Huang and Sekuler further showed that a single feature held in WM is distorted in the direction of a single distractor, as predicted by superposition.

For the same reason, interference by superposition of multiple items in WM is less severe when the items are more similar to each other. Simulations with the IM showed that memory arrays with high similarity between the target and other memory items result in better performance—measured as mean absolute deviation of report—than more heterogeneous arrays. The beneficial effect of similarity between target and nontargets arises because the distorting effect of superposition is more benign in homogenous sets: Nontargets similar to the target contribute activation close to the correct feature in feature space, whereas dissimilar nontargets contribute activation of distant features. This effect is illustrated in Figure 4.

To test this new prediction of the IM we analyzed the effect of target-to-nontarget similarity in Experiments 1 to 3. For each trial we measured the average distance in color space between the target and all nontargets in the array. We then sorted trials into six equal-sized bins of average target-to-nontarget similarity. Figure 5 plots performance in Experiment 1 as a function of set size and target-to-nontarget similarity, together with the corresponding predictions of the IM with its best-fitting parameter values. Figure 6 shows equivalent plots for Experiments 2 and 3. Both the model and the data display an increase of reproduction deviation with decreasing similarity between the target and the nontargets, an effect that becomes larger with increasing set size. In the IM the effect of similarity between the target and nontargets arises from the superposition of nontarget information upon the target information at retrieval. None of the competing models incorporates the assumption of superposition of information from several items, and therefore, none of them can produce the observed effect of similarity.

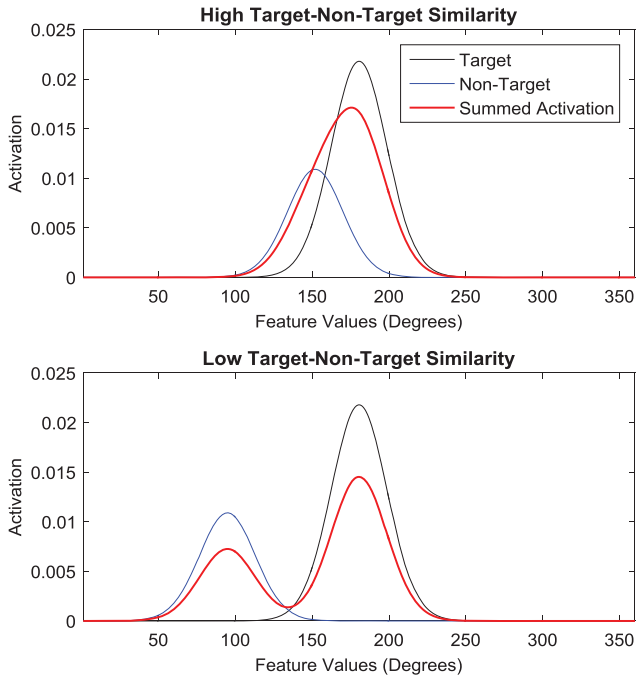

Figure 4. Schematic illustration of the beneficial effect of target-to-nontarget similarity. When a similar nontarget is superimposed on a target (top), the summed activation is shifted only slightly away from the target. A more dissimilar nontarget (bottom) distorts the activation distribution further away from the target. See the online article for the color version of this figure.

The beneficial effect of similarity confirms a prediction from the binding-pool model of visual WM by Swan and Wyble (2014). The binding-pool model is a neural-network model that incorporates cue-based retrieval and superposition, and as such shares fundamental principles with our IM, though differing in many details. The converging predictions from the IM and the binding-pool model confirm that the similarity benefit does not arise from specific assumptions in one of these models but from their shared principles, in particular, interference by superposition. By contrast, the beneficial effect of similarity observed here is the opposite of what is predicted by a model in which the capacity limit arises from competition among representations that are close to each other in feature space. Such a model has been proposed by Franconeri, Alvarez, and Cavanagh (2013).

Our finding of a benefit of similarity within the memory array converges with findings from recognition and change-detection tests of visual WM, which have revealed higher sensitivity with more similar arrays (Lin & Luck, 2009; Makovski, Watson, Koutstaal, & Jiang, 2010; Peterson & Berryhill, 2013). Brady and Tenenbaum (2013) describe systematic accuracy differences between color arrays in change detection, which they attribute to the encoding of ensemble statistics. Their Figure 13 shows that the easier arrays are the ones with higher interstimulus similarity, suggesting that at least part of the systematic variability between arrays can be explained by superposition of object representations without the need to assume additional ensemble representations.

The similarity benefit observed here differs from the benefit of environmental variability reported by Sims, Jacobs, and Knill (2012). Based on an ideal-observer analysis, Sims and colleagues predicted—and confirmed experimentally—that visual WM performance is better when the stimulus features are consistently drawn from a distribution with smaller variability. This effect arises in their model because the optimal Bayesian decoding function combines information from memory with prior information about the environment, and thereby provides more accurate responses when the prior contains more information. When the features to be remembered are drawn from a narrower distribution and the observer knows that distribution, they can use that information to improve performance. In contrast, in the present experiments the colors of each array were drawn from a uniform distribution over the entire color circle, so that the prior knowledge observers could use did not differ for trials with higher or lower similarity between targets and nontargets. The present finding of better memory with higher similarity is therefore not predicted by the model of Sims et al. (2012), unless it is assumed that observers adjust their prior on a trial-to-trial basis depending on the variability of the current array.

Testing Predictions From the Interference Model, 2: The Effect of Context Similarity

The IM makes a unique prediction about people's tendency to report the feature of a nontarget: That tendency should increase with the similarity between the retrieval-relevant context of the target—that is, the context presented as the retrieval cue—and the context of the nontarget. In Experiments 1 to 3, as in most published experiments on visual WM, the retrieval-relevant context was spatial location. These experiments already provide evidence that the tendency to retrieve information from nontargets

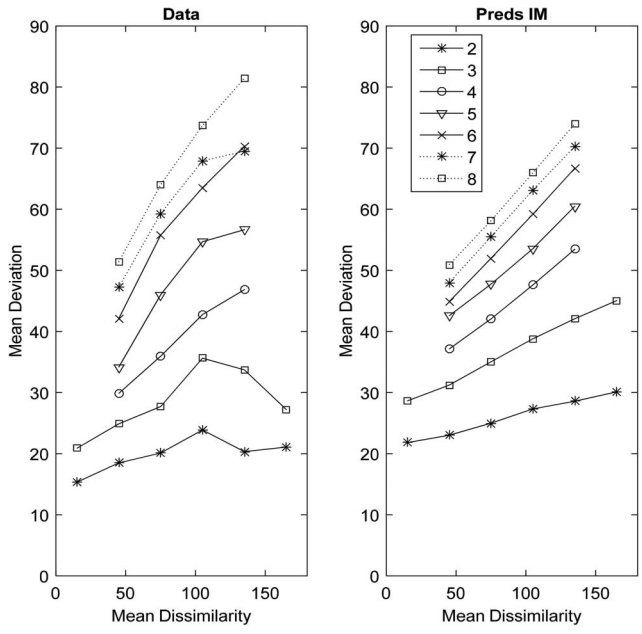

Figure 5. Experiment 1: Mean (absolute) deviation of responses from true color as a function of set size and average dissimilarity between target and nontarget colors. Data from bins with fewer than 50 observations are omitted.

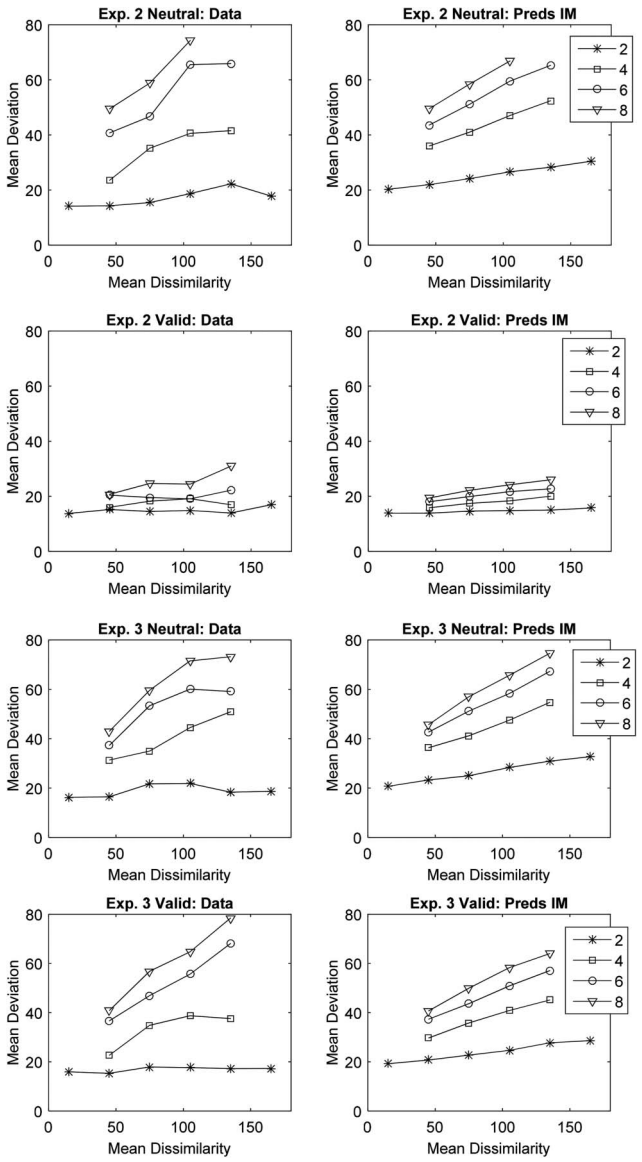

Figure 6. Experiments 2 and 3: Mean (absolute) deviation of responses from true color as a function of set size and average dissimilarity between target and nontarget colors, for neutral and valid-cue trials. Data from bins with fewer than 50 observations are omitted.

increases with their spatial proximity to the target. Figure 7 shows that responses reflecting nontarget features are more prevalent for nontargets spatially close to the target than for more distant nontargets. The IM captures this trend through the spatial-gradient parameter s that reflects the precision of bindings on the context dimension. To test this gradient statistically, we fixed s to 20, which implies virtually zero cueing of spatial neighbors of the target. This constraint reduced the model's fit to Experiments 1, 2, and 3 (ΔAIC = 17, 118, and 209, respectively). Hence, these data provide evidence for the assumption in the IM that nontarget intrusions arise in part from the limited precision of bindings along the context dimension.

With Experiment 4 we tested whether this prediction generalizes to another context dimension when it becomes retrieval relevant: We tested memory for orientations by probes identifying the target through its location, its color, or a combination of both.[7] Experiment 4 used a single set size, and had three conditions: Orientations were probed by location only, by color only, or by a conjunction of location and color. We created a simplified version of the IM that we fit as a measurement model to estimate parameters separately for each condition. The simplified version omitted the focus of attention, because its contribution is difficult to measure without experimental manipulations of set size or pre-/retrocueing. We fit the data of Experiment 4 using two separate s parameters, one for the gradient over spatial location (used in the condition with location probes), and one for the gradient over color (for the condition with color probes). For the condition with combined location-color probes we determined the strength of cueing for each item by a weighted average of the contributions of location cue L_θ and color cue C_θ, with a new parameter w determining the relative weight of the location cue:

$$
\begin{aligned}
A_c(x \mid L_\theta) &= \sum_{i=1}^{n} \exp[-s_{loc} D(L_i, L_\theta)] \cdot vMises(x, x_i, \kappa) \\
A_c(x \mid C_\theta) &= \sum_{i=1}^{n} \exp[-s_{col} D(C_i, C_\theta)] \cdot vMises(x, x_i, \kappa) \\
A_c(x \mid L_\theta, C_\theta) &= \sum_{i=1}^{n} \{w \exp[-s_{loc} D(L_i, L_\theta)] \\
&\quad + (1 - w)\exp[-s_{col} D(C_i, C_\theta)]\} \cdot vMises(x, x_i, \kappa)
\end{aligned}
\tag{14}
$$

Because both the spatial locations and the colors of the objects were equidistantly placed on a circular dimension, we measured distances on both dimensions in radians to enable a comparison of the contextual generalization gradients s_{col} and s_{loc}.

We first fit the IM to each condition separately to test whether we could simplify the model further by fixing some of its remaining free parameters (a, b, s, and κ). A model version fixing $a = 0$ provided the best fit (see Table 3). Figure 8 shows the distributions of response errors in the three probing conditions. Responses were slightly more concentrated around the target orientation in the location-probe than the color-probe condition, with performance in the double probe condition in between. Parameter estimates for the three probing conditions differed only moderately (see Table 4). To statistically test these differences, we constrained all parameters to be the same across probing conditions. The constrained model provided a slightly better fit than the unconstrained IM (ΔAIC = -32). This means that the differences between conditions can be attributed entirely to the different gradient parameters s_{loc} and s_{col}. Fixing these two parameters to be equal led to a loss of fit, ΔAIC = 58. Hence, the data provide evidence for a difference in the generalization gradients of location and color probes: The smaller value for s_{col} compared to s_{loc} implies that cueing retrieval by color generalizes more to neighboring colors than cueing by location generalizes to neighboring locations. Figure 9 shows the responses centered on the orientation of nontargets, separately for nontargets at three levels of distance on the retrieval-relevant context dimension. The prevalence of nontarget responses decreased over spatial distance with location probes, it decreased over color distance (or dissimilarity) with

[7] We thank Nelson Cowan for suggesting this experiment.

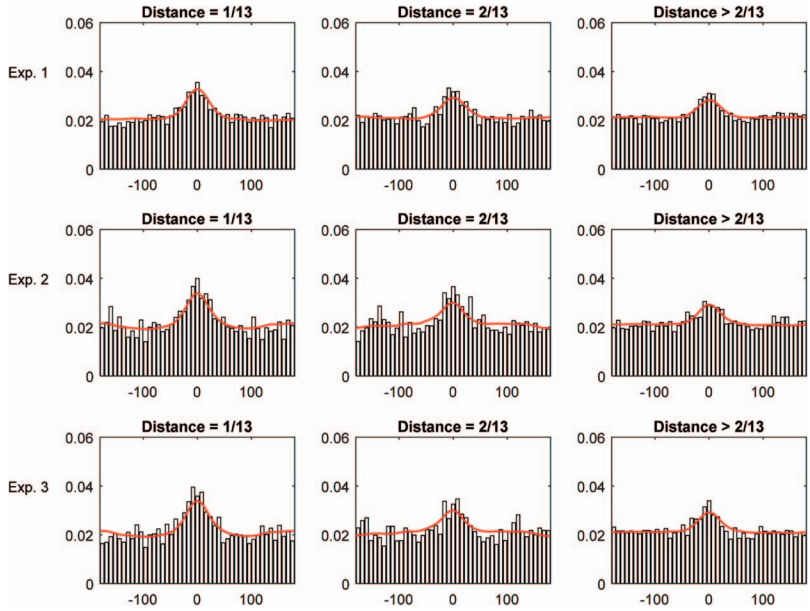

Figure 7. Experiments 1 through 3: Response distributions from all subjects, averaged across set sizes 4 to 8, centered on nontargets at different spatial distances from the target. Left: Smallest possible distance (1/13 of the virtual circle on which objects were arranged), middle: second smallest possible distance; right panel: all larger distances. For Experiments 2 and 3, only neutral trials were included. See the online article for the color version of this figure.

color probes, and it declined with distance along both dimensions with double probes.

It might be surprising that performance in the double-probe condition did not surpass that on the location and color probe conditions, because the double probe provided two redundant retrieval cues. Simulations with the simplified IM showed that two redundant retrieval cues lead to better performance if they are combined multiplicatively (i.e., multiplying the two exponential terms in Equation 14) but not when they are combined by averaging. With a multiplicative combination of retrieval cues, the simplified IM fit the data of Experiment 4 slightly worse than with the weighted-averaging combination. The success of the weighted-averaging equation could mean that people split their attention on each double-probe trial between location and color information, such that their combined contribution to cue-based retrieval is as strong as that of a probe providing a single feature. Alternatively, weighted averaging could also reflect a mixture of trials on which people attend only to the location, and other trials on which they attend only to the color of the double probes, with parameter w reflecting the proportion of trials on which they use location as retrieval cue.

Table 3
Model Fits for Versions of the Interference Model for Experiment 4

Condition	Model Version (parameters fixed) ($a = 0$, $s = 20$)	($s = 20$)	($a = 0$)	Full version
Exp. 4, location probe	0	48	2	92
Exp. 4, color probe	220	98	0	84
Exp. 4, double probe	42	16	0	97

Note. Table entries are differences of AIC values (summed across conditions and participants) relative to the best-fitting model. Higher AIC values reflect worse fit.

To conclude, across all four experiments we obtained evidence for a unique prediction of the IM: The effect of a retrieval cue generalizes to neighbors on the retrieval-relevant context dimension. For Experiments 1 to 3, the IM with a free parameter s for generalization of spatial cueing provided a superior fit to a model with s fixed to a value eliminating generalization (i.e., $s = 20$). The spatial generalization parameter captures the finding that spatially close objects are more likely to be confused than spatially distant objects. This effect is not due to perceptual limitations at encoding because it is found also with sequential presentation of items (Rerko et al., 2014). Moreover, Experiment 4 generalized the effect to another dimension by which the target can be identified at test: color. When location was used as the retrieval cue provided by the probe, participants tended to retrieve the orientation of a spatially nearby nontarget; when color was the retrieval cue, they tended to retrieve the orientation of an object with a similar color, and when both retrieval cues were provided jointly, the tendency to confuse the target with a nontarget showed a gradient along both spatial location and color.

No other model of visual WM captures this aspect of the data. The spatial gradient and color gradient on confusion errors is theoretically important because it mirrors the temporal confusion gradient observed in memory for sequentially presented memory sets (Farrell & Lewandowsky, 2004; Murdock & vom Saal, 1967). From the perspective of interference theories of memory, such a gradient is to be expected: Retrieval of a memory content such as a list item or an object's visual feature is driven by a context cue

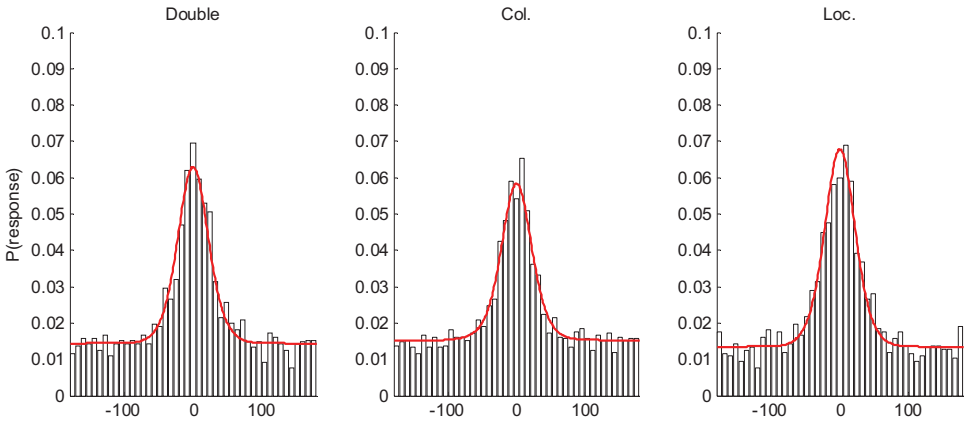

Figure 8. Experiment 4: Observed distributions of responses across all subjects (black bars) and predictions from the simplified IM, centered on the target. See the online article for the color version of this figure.

such as the item's temporal list position, the object's location in space, or—in our Experiment 4—its color. Representations of neighboring list positions overlap more than those of positions further apart, explaining why list items are more often confused with close neighbors than with more distant items (Farrell & Lewandowsky, 2004; Henson, 1998). Representations of neighboring spatial locations likewise overlap more than those of more distant locations, so that when retrieval of an object is cued by its spatial location, nearby objects are more likely to be erroneously recalled than more distant objects. The same gradient of confusion errors is to be expected along any dimension on which the retrieval cue distinguishes the target from nontargets.

Testing Predictions of the Interference Model, 3: Controlling the Focus of Attention

One key assumption in the IM is the notion of a focus of attention directed to a single item in WM. In standard visual WM paradigms researchers have no control over which item is in the focus, and therefore we have to model such data as a probability mixture of cases where the target is in the focus and cases where it is not. Experiments 2 and 3 were designed with the aim to control which item is in the focus of attention. In Experiment 2, a precue indicated the target with 67% validity, enabling participants to direct the focus to the target during encoding and holding it there until test, so that the target can be represented with high precision, and in addition be shielded from nontarget information and noise. In Experiment 3, the target was indicated by a retro-cue 1 s after offset of the memory array. The focus therefore could be shifted to the target only after encoding—this implies that the target is again shielded from nontarget information and noise, but at this point in time the focus cannot turn a low-precision representation into one with higher precision, because there is no more precise information available than that in the representation of the target generated at encoding.

Based on these considerations, the IM makes a number of specific predictions for Experiments 2 and 3. First, the IM treats precueing and retro-cueing differently, and predicts that only precueing increases the representational precision of a validly cued target. The IM also predicts that for a validly precued target the set-size effect will be much diminished, compared with the standard no-cue condition. The reason for this prediction is that a validly precued item is always in the focus of attention from the time of encoding, benefiting both from high precision and a reduction of nontarget information and noise. Thus, most of the sources of the set-size effect are eliminated for a validly precued target. For a validly retro-cued target the IM predicts a much less pronounced reduction of the set-size effect because one major driving force of the set-size effect, the diminishing probability of the target being in the focus of attention since encoding, is still in place.

Applying the Models to the Pre-Cue and Retro-Cue Paradigms

The interference model. Application of the IM to the precue paradigm of Experiment 2 is straightforward: We can use Equation 8

Table 4
Median Parameter Estimates (Standard Deviation in Parentheses) for Experiment 4

Condition	*b*	s_{loc}	s_{col}	*SD*	w
Location probe	.83 (4.74)	2.46 (3.14)	—	25.3 (15.2)	—
Color probe	1.26 (3.77)	—	1.0 (4.07)	21.0 (11.0)	—
Double probe	1.11 (1.38)	1.90 (6.08)	1.50 (5.22)	24.0 (9.22)	.56 (.40)
Joint fit to all conditions	.66 (.53)	1.71 (4.10)	1.14 (4.14)	25.5 (6.4)	.77 (.27)

Note. The first three rows contain parameter estimate from fitting the IM (version with *a* fixed to 0) separately to each condition; the last row reflects parameters from the joint fit, constraining all parameters to be equal across conditions. The constrained model had a better fit (ΔAIC = 32).

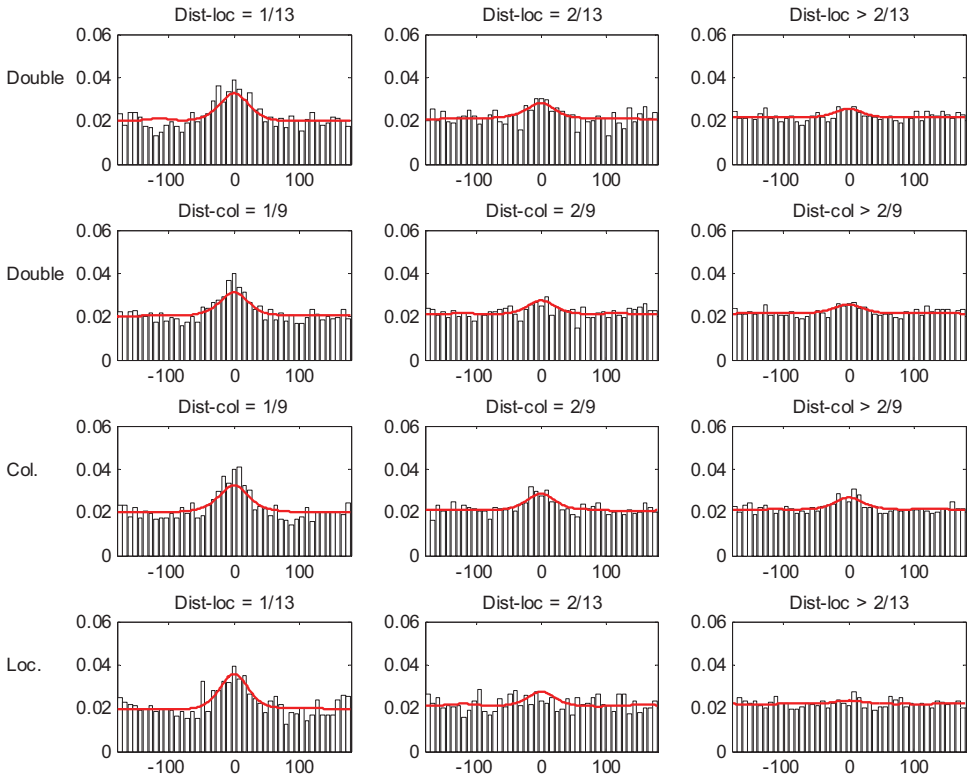

Figure 9. Experiment 4: Observed distributions of responses across all subjects (black bars) and predictions from the simplified IM (red lines), centered on nontargets at three levels of distance from the target in space (loc) and in color space (col). Distance is measured by steps in the equidistant placement of 13 locations on the virtual circle of the array, or 9 colors in the color circle, respectively: The left-most panels show responses centered on the features of immediate neighbors of the target, the middle panel shows responses centered on the features of nontargets that were neighbors once removed, and the right panel shows responses centered on the features of distant nontargets. The distributions were generated by centering the response distributions on each nontarget in turn, adding them separately for nontargets in the three distance categories, and dividing by the number of nontargets in each distance category. See the online article for the color version of this figure.

as for Experiment 1, setting $P(F_t) = 1$ for the valid-cue condition, to 0 for the invalid-cue condition, and to $1/n$ for the neutral condition. These settings reflect our assumption that the focus of attention is always directed to the cued item when a cue is presented.

For the retro-cue paradigm of Experiment 3 we need to refine Equation 8 by distinguishing between the case where the target is in the focus of attention at encoding, and is still in the focus at test—this is the case denoted F_θ—and the case where the target is not in the focus during encoding—denoted $\neg Fe_\theta$—but the focus shifts to it during the retention interval, Fr_θ:

$$\begin{aligned}
&A(x \mid \neg F_\theta) = cA_c(x) + aA_a(x) + bA_b(x).\\
&A(x \mid F_\theta) = cA_c(x) + cA_f(x) + raA_a(x) + rbA_b(x).\\
&A(x \mid \neg Fe_\theta, Fr_\theta) = cA_c(x) + raA_a(x) + rbA_b(x).\\
&A(x) = [1 - P(F_\theta)] \cdot A(x \mid \neg F_\theta) + P(Fe_\theta) \cdot P(F_\theta \mid Fe_\theta) \cdot A(x \mid F_\theta)\\
&\qquad + [1 - P(Fe_\theta)] \cdot P(Fr_\theta \mid \neg Fe_\theta) \cdot A(x \mid \neg Fe_\theta, Fr_\theta).
\end{aligned} \tag{15}$$

Equation 15 expresses the activation distribution as a probability mixture of three cases: In the first case, the target is not in the focus of attention at test; this occurs with probability $1 - P(F_\theta)$. In that case, no benefit from focusing applies. In the second case, the target is in the focus of attention at encoding, and stays there until test, in which case it benefits from increased precision and a reduction of the influence of A_a and A_b by factor r. The probability of this case equals the probability that the target happens to be in the focus at encoding, $P(Fe_\theta)$, times the conditional probability that it is still in the focus at test, $P(F_\theta \mid Fe_\theta)$. In the third case, the target is not in the focus at encoding, but the focus shifts to it during the retention interval in response to a retro-cue. In this case, which occurs with probability $[1 - P(Fe_\theta)] \times P(Fr_\theta \mid \neg Fe_\theta)$, the target still benefits from a reduced impact of A_a and A_b, but not from increased precision. Because in Experiment 3 participants did not know the target during encoding in any of the cueing conditions, $P(Fe_\theta)$ was $1/n$ for all conditions. The retro-cue could affect only $P(F_\theta \mid Fe_\theta)$ and $P(Fr_\theta \mid \neg Fe_\theta)$. For the valid-cue condition, we set both conditional probabilities to 1 because we assumed that the retro-cue always directs the focus of attention to the target before the test. For the invalid-cue condition, we set both to zero because the retro-cue misleads the focus of attention away from the target. For the neutral condition, we assumed as before that the focus of attention stays where it was during encoding, so $P(F_\theta \mid Fe_\theta) = 1$ and $P(Fr_\theta \mid \neg Fe_\theta) = 0$.

The Slot-Averaging model. How could the Slot-Averaging model be applied to the cueing paradigms? Zhang and Luck (2008) report an experiment in which a cue appeared simultaneously with the memory array. They assumed that participants devote a large proportion of their slots to the cued item. This is plausible when the cue is available at encoding, but it is difficult to imagine how anything could be gained from shifting slots to an item retro-cued during the retention interval. Either the cued item has not been encoded into a slot, then it is not represented at all, and it is logically impossible to shift a slot to it. Or else, the cued item has already been encoded into a slot, and then shifting another slot to it does not increase its probability of recall. One could argue that devoting one or more additional slots to the cued item increases its precision, but it is unclear where the necessary information for a more precise representation should come from if it is not represented anywhere during the time between the memory array and the retro-cue.

We decided to give the Slot-Averaging model maximal flexibility in applying it to Experiments 2 and 3. To that end, we added four more free parameters to represent the average number of slots devoted to the cued item, one for each set size exceeding 1 (i.e., c_2, c_4, c_6, c_8 for set-sizes 2, 4, 6, and 8, respectively). For the precue experiment we expect that the cued item receives more slots than what would be expected for an average item in the absence of a cue. In contrast, for the retro-cue experiment the theory behind the Slot-Averaging model implies the prediction that the cued item receives no more slots than the average item in the no-cue condition, that is, $c_n = K/n$. We fit the model with free c parameters to see whether their estimates bear out this prediction.

The Variable-Precision model. When an item is cued before or during display of the memory array, it is plausible to assume that a larger resource share is allocated to it than to other items, and this is how Bays, Gorgoraptis, Wee, Marshall, and Husain (2011) explain the benefits for a valid cue, and the cost of an invalid cue, presented during display of the memory array. When a retro-cue is presented during the retention interval, it is still possible to reallocate some resource quantity to the cued item, but it is difficult to conceive how this might increase the representation's precision without access to additional information.

As for the Slot-Averaging model, we gave the Variable-Precision model maximal flexibility to account for the cueing effects in Experiments 2 and 3: For each set size exceeding 1, we added a free parameter representing the proportion of the resource that is assigned to the cued item (i.e., p_2, p_4, p_6, p_8). The remainder is divided equally among all items (including the cued one). We note that by introducing separate resource-shifting parameters (or slot-shifting parameters for the Slot-Averaging model), we give the models additional flexibility to account for the set-size effect in the cue conditions. For instance, the Variable Precision model can compensate any deviation from the power function at a given set size by adjusting the p_n parameter for that set size.

Competitive model fitting. The two right-most columns of Table 1 show the AIC values for the competing models when applied to Experiments 2 and 3.[8] The table also includes fit indices for the Slot-Averaging and the Variable-Precision model when applied separately to each cueing condition. These model applications serve to test the models' fit when they are not required to explain the cueing effects. For both experiments, the IM clearly provided the best fit among all competitors. The parameter estimates for the IM are included in Table 2.

Figure 10 shows the mean deviations in the data of Experiment 2 (top panel) and Experiment 3 (bottom panel) alongside the predictions of the IM. Both valid precues and retro-cues improved

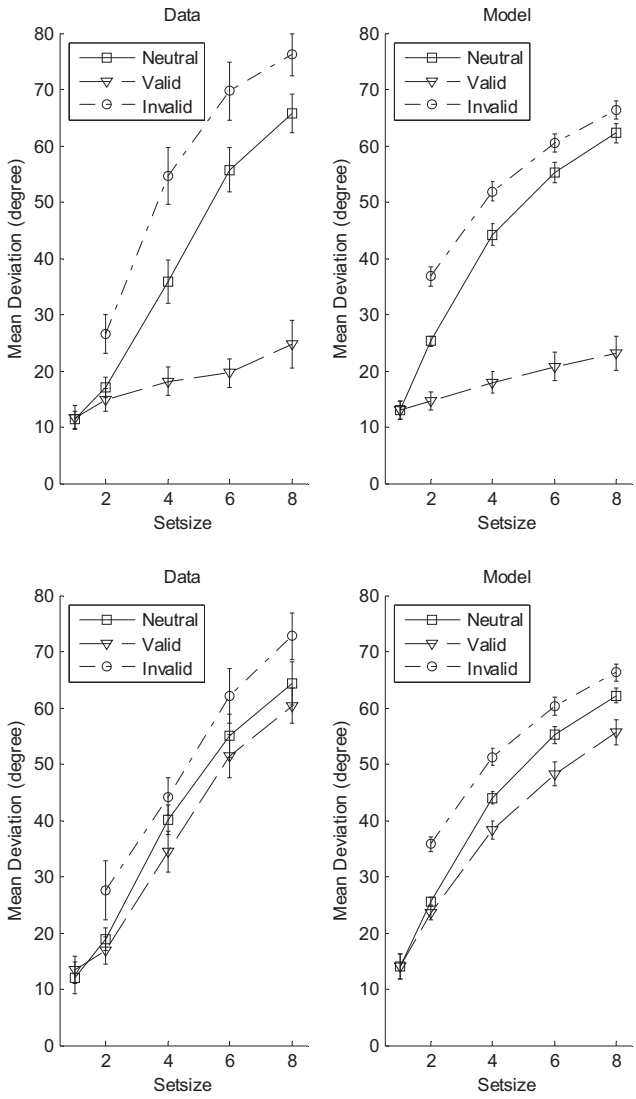

Figure 10. Mean (absolute) deviation of responses from true color as a function of set size and cue condition in Experiments 2 (top) and 3 (bottom); experimental data on the left, and predictions from the IM on the right. Error bars are 95% confidence intervals for within-subject comparisons.

performance, in line with previous findings (Lepsien et al., 2011; Murray et al., 2013; Souza, Rerko, Lin, et al., 2014), but the beneficial effect of valid precues was much larger. As predicted by the model, a valid precue substantially reduced the set-size effect. In contrast, a valid retro-cue reduced the set-size effect very little.

When the precue or retro-cue was invalid, performance was worse than in the neutral conditions without a cue. In the data, the set-size effects for the invalid and the neutral condition were largely parallel

[8] All models were fit with the Simplex algorithm implemented in the *fminsearch* function of Matlab, keeping the best of 10 runs with different, randomly chosen starting values. We also fit the 7-parameter Variable-Precision model with the *patternsearch* and the *ga* (genetic-algorithm) functions of Matlab, and obtained the same fits.

in both experiments. The IM predicts the set-size effect to be flatter for invalidly cued targets because in that condition the probability that the target is in the focus of attention does not diminish with set size (i.e., it is always zero). There is no evidence, however, that this deviation between model prediction and data is systematic, because for both Experiments 2 and 3 the fit of the IM was even better than that of the three-parameter mixture model (ΔAIC in favor of the IM was 1204 for Experiment 2, and 1121 for Experiment 3). As the mixture model was fit separately to each combination of set-size and cueing condition, it was free to capture any main effect and interaction of those two variables by appropriate parameters. If there was a systematic effect that the IM is missing, this would show as a better AIC for the mixture model than the IM.

The response distributions from Experiment 2 at set size 6 are shown in Figure 11, together with the fits of the IM. The bottom two rows of the figure show response distributions centered on nontargets. The IM accurately reproduced the increased prevalence of nontarget features in the neutral and the invalid-cue conditions, and their virtual absence in the valid-cue conditions. One unexpected observation—not predicted by the IM—is that when the cue was invalid, the cued item's feature tended to intrude with higher probability than the feature of a not-cued nontarget.

Figure 12 shows the corresponding response distributions from Experiment 3. Again there was a tendency for reporting nontarget features, and this tendency was visible even for the validly retro-cued trials. When the nontarget intrusions in invalidly cued trials were broken down by intrusions from the cued and from noncued nontargets, there was again a tendency for more intrusions from the cued nontarget. This tendency, though weak, was observed consistently at all set sizes (not shown). Nevertheless, because it was not predicted by the IM or any other model, we believe that it needs to be replicated before drawing theoretical conclusions from it.

Alternative Model Versions

So far we have pitted the IM against other published models. We now consider eight further models that are of interest because they

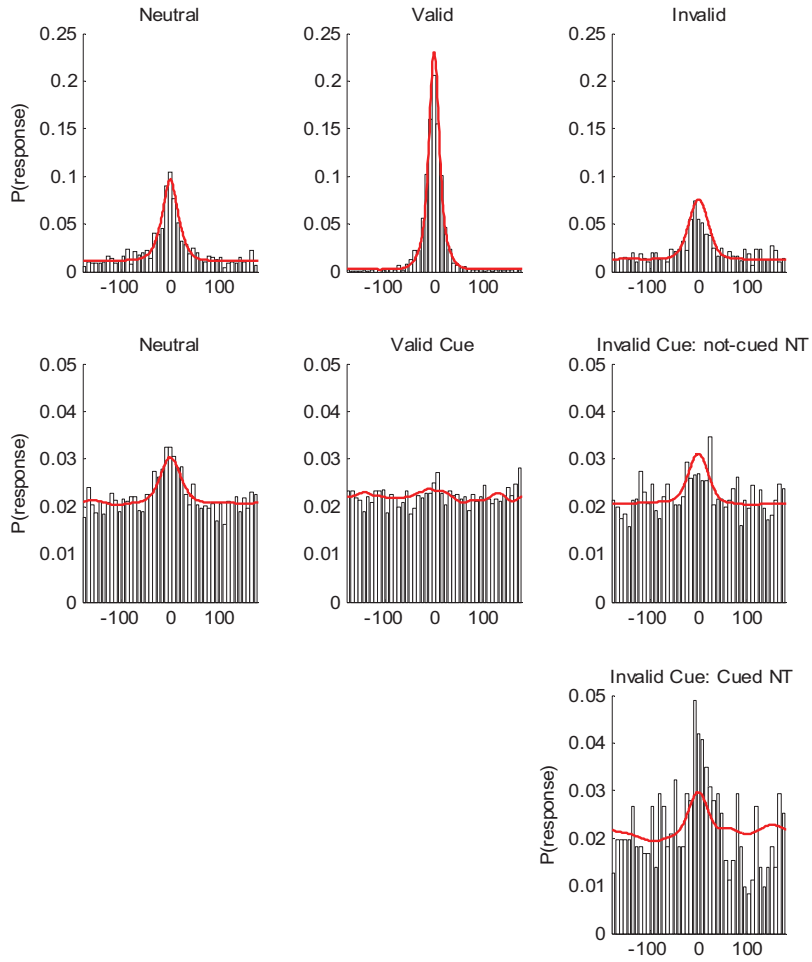

Figure 11. Response distributions from Experiment 2, set size 6, for the three cue conditions. The top row shows response distributions centered on the target color; the bottom two rows show the distribution of the same responses centered on the nontarget colors. For invalid-cue trials, response distributions are shown centered on the cued nontarget (NT), and centered on the not-cued nontargets. See the online article for the color version of this figure.

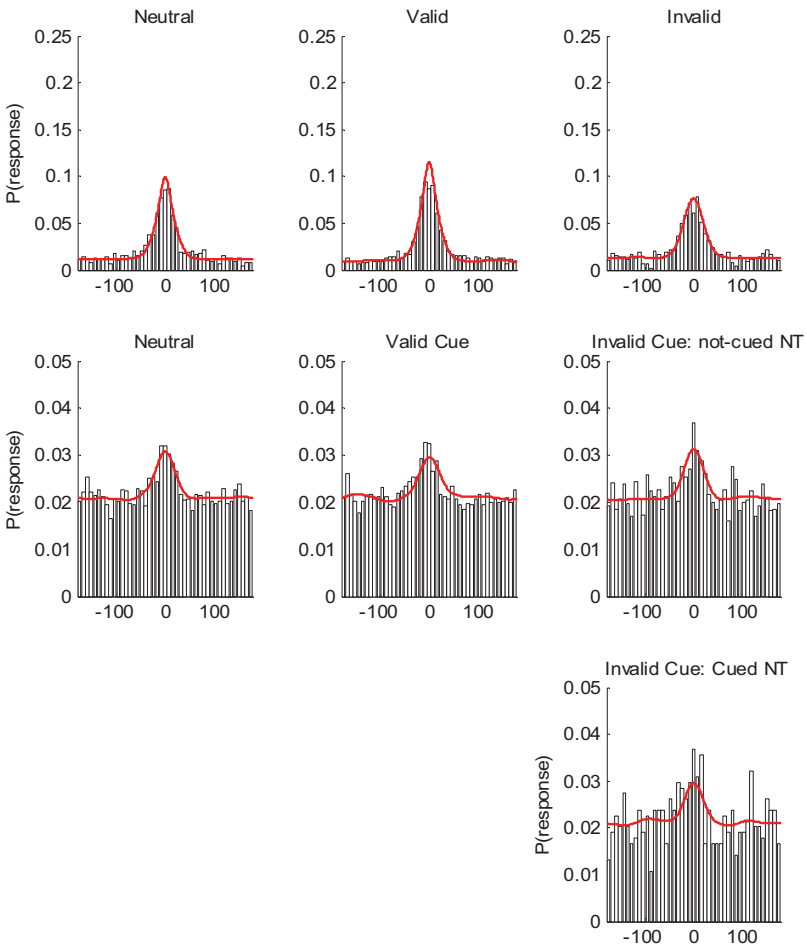

Figure 12. Response distributions from Experiment 3, set size 6, for the three cue conditions. The top row shows response distributions centered on the target color; the bottom two rows show the distribution of the same responses centered on the nontarget colors. For invalid-cue trials, response distributions are shown centered on the cued nontarget (NT), and centered on the not-cued nontargets. See the online article for the color version of this figure.

incorporate important theoretical ideas discussed in the literature on visual working memory. The first six models are variants of the IM that test how adding or changing one theoretical assumption affects the model's ability to account for the data of Experiments 1 to 3: (a) An IM augmented with variable precision, (b) a version of the IM in which some responses are attributed to guessing, (c) a slot-limited interference model, (d) a generalized version of the IM allowing for more than one object to be held in the focus of attention, (e) a version of the IM with decay and sequential refreshing of memory items by the focus of attention, and (f) a model implementing the idea of binding by synchrony, which turns out to be yet another version of the IM. The final two models are attempts to overcome the single most important limitation of the two competing models, their inability to explain nontarget intrusions in recall. To that end we constructed variants of the Slot-Averaging model and of the Variable-Precision model augmented by a mechanism to include binding errors. We fit these eight models to Experiments 1 to 3; their fit indices—in comparison with the IM—are summarized in Table 5, and Table 6 tallies the number of participants for which each model fits better than each other model.

An Interference Model With Variable Precision

The assumption of variable precision has been found to be important for a good fit to continuous-reproduction data: Van den Berg et al. (2014) tested a set of models generated by factorially combining several model features. Holding other features constant, adding random variability in precision improved the fit of all models. As van den Berg and colleagues (2012) noted, variable precision enables the model to accommodate the shape of the response distribution better than a von Mises distribution with constant variability: A mixture of von Mises distributions with different precision parameters is more peaked and has fatter tails than the von Mises distribution itself, and these features accommodate the empirical distributions better, as can be seen in the residuals of the response distribution. To illustrate, we plotted the residuals of the three competing models applied to Experiments 1

Table 5
Model Fits of Additional Model Variants in Comparison With the Interference Model

Model	Experiment 1		Experiment 2		Experiment 3	
	N(Param)	ΔAIC	*N*(Param)	ΔAIC	*N*(Param)	ΔAIC
IM-VP	7	33	7	−24	7	−3
IM-Guessing	6	5	6	−19	6	−143
GIM	7	38	7	31	7	383
IM-DC	7	6	7	−144	7	274
IM-DR	6	−139	6	307	6	253
IM-Phase	5	−102	5	391	5	487
Slot-Averaging Binding	3	1316	7	1894	7	1971
VP-Binding	3	−18	4	812	—	—

Note. *N*(Param) is the number of free parameters, and ΔAIC is the difference of the Akaike Information Criterion (AIC, summed across participants) from the IM (lower values indicate better fit). IM-VP = Interference model with variable precision; GIM = Generalized Interference Model; IM-DC = Discrete-Capacity version of the IM; IM-DR = Decay-Refreshing version of the IM; IM-Phase = version of IM with binding by synchrony in phase space.

to 3, averaged across set sizes, in Figure 13. The Slot-Averaging model, which uses a constant precision parameter, shows the characteristic Mexican-hat shaped residual first observed by van den Berg et al. (2012): Models with constant precision tend to underestimate the peak of the response distribution and overestimate its shoulders. This misfit largely disappears in the Variable-Precision model.

The IM also shows very little, if any, sign of misfit of the shape of the response distribution, as shown by the essentially flat residual. This is so because the predicted distribution of the IM arises from a mixture of von Mises distributions with two different precisions, κ and κ_f. Different from the Variable-Precision model, the IM does not attribute the variability in precision to random fluctuations, but rather to the fact that in a random subset of $1/n$ trials the target happens to be represented both with lower precision outside, and with higher precision in the focus of attention, whereas in the remaining trials it is only represented with lower precision.

We asked whether adding random fluctuations in the nonfocal precision κ improved the fit of the IM. There was at best weak evidence that it did: Introducing variable precision into the IM through an additional parameter led to only a small improvement of AIC in Experiments 2 and 3, and none in Experiment 1 (see Table 5). This result suggests that the mixture of high-precision activation from the focus of attention and low-precision activation from outside the focus is sufficient to account for most of the variability in precision reflected in the data (Fougnie, Suchow, & Alvarez, 2012; van den Berg et al., 2012).

Particularly diagnostic data for testing this interpretation come from the invalid-cue conditions in Experiments 2 and 3, because in these conditions the target was never in the focus of attention, so the activation distribution is not a mixture of distributions with different precisions. There was no systematic deviation of the residuals from zero in the invalid-cue conditions in Figure 13. In particular, there was at best a hint of the Mexican-hat shape in the residuals that would be indicative of variable precision in the data but not captured by the model (van den Berg et al., 2012). We conclude that when the focus of attention is controlled, there is little evidence in the data for variability of precision. This is not to say that there is no fluctuation in precision at all—this would be implausible because neural systems are fundamentally noisy (Fougnie et al., 2012). Moreover, there is compelling evidence that colors differ systematically in how precisely they are represented in perception and memory (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014). Any random or stimulus-specific variability in precision, however, appears to be too small to have much impact on the present data. The lion's share of precision variability evident in our data can be explained as a mixture of trials in which the target is in the focus of attention and trials in which it is not.

An Interference Model With Guessing

The Slot-Averaging model explains the response distribution as a mixture of trials on which people respond on the basis of information in memory and trials on which they guess. This assumption is also incorporated in the popular mixture model of continuous-reproduction data (Bays et al., 2009; Zhang & Luck, 2008). The guessing component of the mixture is modeled as a uniform distribution that accounts for the fat tails of the error distributions (see Figure 3).

So far we did not incorporate guessing into the IM; rather, the fat tails of the error distribution are attributed to background noise. Because background noise is independent of the array stimuli, it is modeled as a uniform distribution of activation. Here we consider whether the assumption of background noise can be replaced by the assumption that, on a proportion of trials, participants do not use any information from memory and instead guess at random.

The decision to guess should be motivated by low confidence in the information in memory. Participants have remarkably good insight into how well they memorize visual information in WM on a trial-to-trial basis (Adam, Mance, Awh, & Vogel, 2013; Rademaker, Tredway, & Tong, 2012). We implemented a simple idea of how the decision to guess could be tied to a measure of the quality of information in visual WM: The circular variance of activation $A(x)$ at retrieval. The IM-guessing model is a mixture model in which people guess on a proportion g of trials, and respond on the basis of memory on the remaining trials. Guessing replaces background noise, so we

Table 6
Number of Participants for Whom Each Model (Rows) Fit Better Than Each Other Model (Columns) in Competitive Fitting of Experiments 1 to 3

Winning Model	SA	SA-B	VP	VP-B	IM	IM-VP	IM-g	IM-DC	IM-DR	IM-Phase
Experiment 1										
Slot-Averaging	0	3	2	0	0	0	0	0	0	0
Slot-Avg.-Binding	16	0	7	0	2	2	2	2	0	2
Variable Precision	17	12	0	0	2	2	3	2	1	1
VP-Binding	19	19	19	0	9	9	10	9	8	9
IM	19	17	17	10	0	18	10	13	3	2
IM-VP	19	17	17	10	1	0	11	2	1	2
IM-Guessing	19	17	16	9	9	8	0	9	7	5
IM-DC	19	17	17	10	6	17	10	0	2	2
IM-DR	19	19	18	11	16	18	12	17	0	10
IM-Phase	19	17	18	10	17	17	14	17	9	0
Experiment 2										
Slot-Averaging	0	6	0	1	0	0	0	0	0	0
Slot-Avg.-Binding	15	0	2	2	0	0	0	0	0	0
Variable Precision	21	19	0	14	2	3	3	2	9	6
VP-Binding	20	19	7	0	3	3	4	3	6	4
IM	21	21	19	18	0	14	9	3	20	16
IM-VP	21	21	18	18	7	0	12	4	20	16
IM-Guessing	21	21	18	17	12	9	0	3	20	18
IM-DC	21	21	19	18	18	17	18	0	21	17
IM-DR	21	21	12	15	1	1	1	0	0	11
IM-Phase	21	21	15	17	5	5	3	4	10	0
Experiment 3										
Slot-Averaging	0	3	1	—	0	0	0	0	0	0
Slot-Avg.-Binding	18	0	6	—	0	0	0	1	1	2
Variable Precision	20	15	0	—	1	1	0	6	5	6
IM	21	21	20	—	0	13	1	18	19	18
IM-VP	21	21	20	—	8	0	3	18	19	20
IM-Guessing	21	21	21	—	20	18	0	20	20	21
IM-DC	21	20	15	—	3	3	1	0	10	15
IM-DR	21	20	16	—	2	2	1	11	0	15
IM-Phase	21	19	15	—	3	1	0	6	6	0

Note. For Experiments 2 and 3 the results of Slot-Averaging, Resource, and Variable-Precision (VP) models are from model versions fit simultaneously across the three cueing conditions.

eliminated the A_b component of activation. Hence, Equation 2 simplifies to

$$A(x \mid L_\theta) = cA_c(x \mid L_\theta) + aA_a(x), \tag{16}$$

and likewise, A_b was dropped from Equation 15 for Experiments 2 and 3. The probability of guessing is a linear function of the variance of activation at retrieval, with g replacing b as a free parameter:

$$P(guess) = g\ \mathrm{var}(A(x \mid L_\theta)) \tag{17}$$

The probability of choosing each response x (Equation 1) becomes a mixture of memory-informed and uniformly distributed guessing responses:

$$P(x) = (1 - P(guess))\frac{A(x)}{\sum_{j=1}^{N} A(j)} + P(guess)\frac{1}{2\pi} \tag{18}$$

The IM-guessing model did about as well as the original IM (see Table 5). This is because the average variance of $A(x)$ increases nearly linearly with set size. Hence, in both model versions the uniform component of the predicted error distribution increases approximately linearly with set size, so that the two model versions mimic each other closely.

The Generalized Interference Model

In the IM we assume that the focus of attention is concentrated on exactly one item at any time. Here we relax this assumption and allow that the focus spreads out in space to incorporate multiple items. We still assume that the focus is centered on one item in location L_f, but it spreads to neighboring items. The content of the focus is the weighted sum of high-precision representations of several items. The weight W of each item i depends on its spatial distance from the focus center L_f:

$$W(i \mid L_f) = \exp[-s_f D(L_i, L_f)], \tag{19}$$

with s_f a free parameter governing the spatial spread of the focus. The focus contains information about each item's feature as well as its context, so the weights apply to the item's features, x_i, as well as their locations, L_i. A core assumption in our theoretical framework conceptualizing the focus of attention is that the focus has no mechanism of content-context bindings (Oberauer, 2009; Oberauer & Hein, 2012). Therefore, the feature representations of multiple items in the focus are not bound to their corresponding context representations in the focus. As a consequence, the feature information in the focus cannot be used selectively: When feature information of multiple items is mixed in the focus, there is no way

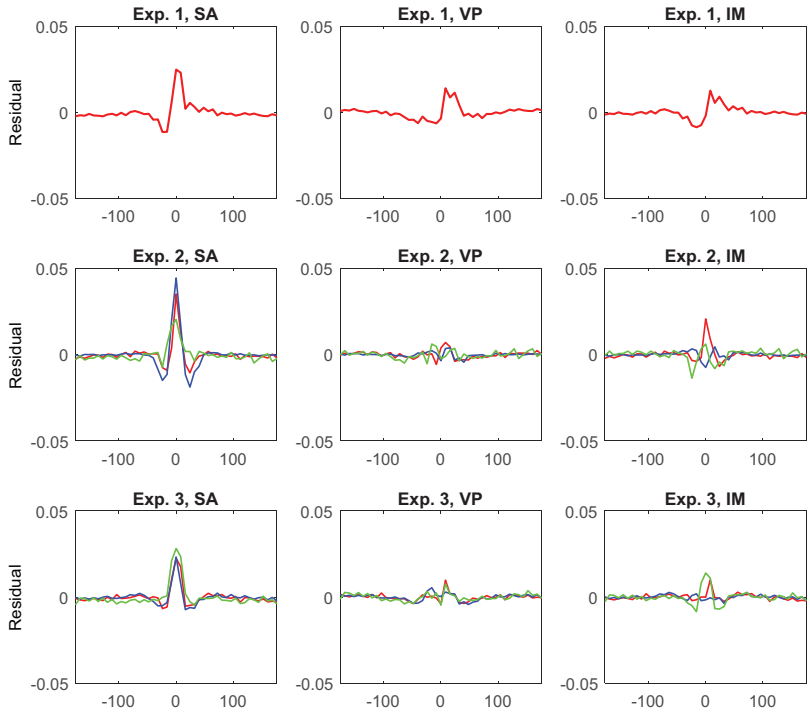

Figure 13. Residuals of response distributions from Experiments 1 through 3, centered on target, after subtracting predictions from the Slot-Averaging model (SA), the Variable-Precision model (VP), or the Interference Model (IM), averaged over all set sizes. For Experiments 2 and 3, red = neutral, blue = valid, green = invalid cueing condition. See the online article for the color version of this figure.

to read out the feature information of only one item without the feature information of the other items. To the extent that feature information from the focus is used, the entire mixture is used. However, the model can use the context information in the focus to decide to what degree it uses the feature information in the focus—this is important to avoid a strong bias toward recalling the feature of the item in the focus center even when an item in an entirely different location is tested.

We implemented the selectivity of using the focus information as follows: At test, the content of the focus $A_f(x)$ is added to the activation $A(x)$ to the extent that the tested context L_θ matches the mixture of context representations in the focus: The content of the focus is added to the activation with weight $W(L_\theta, L_f)$. When L_θ matches the focus center L_f, the content of the focus is added with weight 1; the further away L_θ is from the focus center L_f, the less weight is given to the content of the focus (see Equation 19). In this way, the content of the focus is used to the degree that it contains information about the tested feature. The IM with a fixed single-item focus is a special case of the GIM in which the focus-spread parameter s_f is fixed to a high value, such that only the item in the focus center is represented in the focus with a non-negligible weight. At test, the content of the focus is added to the activation only when the target location matches the focus center, that is, when the focused item is tested.

When fitting the GIM to the data of Experiments 1 to 3 we found that the focus-spread parameter s_f is estimated to a value so high (for instance Experiment 1: $M = 7.7$, $SD = 5.8$, median = 5.6) that the focus is effectively limited to a single item. The *−log(Likelihood)* of both models was virtually identical, and because the GIM has one more free parameter, it received a worse AIC than the IM because of its larger penalty term (see Table 5).

A Slot-Limited Interference Model

It could be argued that interference plays a role in visual WM but is not alone responsible for the capacity limit. To investigate the viability of that conjecture, we constructed a version of the IM that includes a discrete capacity limit: Whereas in the IM all items in the array are assumed to be encoded into WM, the discrete-capacity interference model (IM-DC) assumes that a person encodes only a maximum of K items. Any array with a set size exceeding K is treated as if the not-encoded items had not been presented. This means that the sums in Equations 3 and 5 run only over the K items that were encoded: Only encoded items have context-independent activation A_a, and only encoded items can receive activation A_c from cue-based retrieval; background noise A_b, in contrast, increases with every item presented. If a not-encoded item is tested, the location probe does not cue the target, but it cues those nontargets that were encoded according to their spatial distance to the target. In all other regards, the IM-DC operates exactly as the IM.

In applying the IM-DC to the data of Experiments 1 to 3 we made two technical assumptions. First, when only K out of n objects of an array are encoded, there is no way of knowing which of them are encoded. For Experiments 1 and 3 we assumed that each item had the same probability of being encoded, because

participants had to assume at encoding that each item was equally likely to be the target. Therefore, we computed predictions for each possible subset of K out of n items and averaged them with equal weights. For Experiment 2 we assumed that the precued item was always encoded, and the remaining $K - 1$ slots assigned with equal probability to the remaining items. Second, we allowed K to vary continuously as a free parameter in the same way as in the Slot-Averaging model.

As the relative fit indices in Table 5 and the pairwise competition results in Table 6 show, the IM-DC performed about as well as the IM across the three experiments. We conclude that the present data do not rule out a discrete capacity limit, but they do not mandate it either. If a discrete capacity limit is assumed, it must be augmented with the mechanisms of interference and the focus of attention, as in the IM-DC. As we have shown above, a bare-bones discrete-capacity model without interference and without focus of attention, such as the Slot-Averaging model, is not competitive. We also tested a version of IM-DC that keeps interference but eliminates the focus of attention (i.e., fixing parameter $\kappa_f = \kappa$, and the interference-reduction parameter $r = 1$). This model fit the data of Experiment 1 substantially worse than the IM (ΔAIC = 201) and also worse than the complete IM-DC (ΔAIC = 195). We did not apply this reduced version of IM-DC to Experiments 2 and 3 because it is unclear how it could explain the precue and retro-cue effects without a focus of attention. We conclude that a discrete capacity limit remains a viable theoretical assumption, but only in combination with interference and a high-precision focus of attention as implemented in the IM.

A Decay-Refreshing Model

Although decay models have not played a role in the discussion about visual WM, they have a venerable history in research on verbal WM (Ricker, Vergauwe, & Cowan, 2016). Therefore, we explored a version of the IM in which the strength of bindings decays, and the focus of attention cycles among the object in WM, refreshing their bindings in turn. This model is an application of the Time-Based Resource-Sharing theory (Barrouillet et al., 2004) to the continuous-reproduction paradigm, closely following the computational implementation of that theory by Oberauer and Lewandowsky (2011).

The IM with decay and refreshing (IM-DR) differs from the IM in two regards. First, the binding parameter c declines exponentially over time with decay rate D. The focus of attention takes R ms to refresh each object, restoring its binding to its original value, $c = 1$. At the time of test, the last-refreshed object has its full binding strength, the next-to-last refreshed object has suffered decay for R ms, the object refreshed before it has suffered decay for 2^*R ms, and so on. In general, the binding strengths of n objects in WM at the time of test are

$$c(i) = \exp(-D \cdot R(i-1)), \tag{20}$$

with i, the position in the refreshing sequence, counted backward from the time of test, ranging from 1 to n. Hence, decay imposes a recency gradient on memory strength as a function of each item's refreshing lag (cf. Donkin & Nosofsky, 2012). Because the participant does not know which object will be tested, the tested object has an equal chance of being in each position of the refreshing sequence, i. We therefore calculated the predicted response distribution for all values of i between 1 and the current trial's set size n, and took the unweighted mixture of these distributions as the model prediction. Because D and R cannot be identified independently, we fixed R to 50 ms (Vergauwe & Cowan, 2014).

A second difference from the IM is that, in the absence of a pre- or retro-cue, the focus of attention does not hold one object throughout the retention interval—rather, it cycles rapidly over all objects, except in the case of set size 1. Therefore, it cannot hold one object with increased precision; hence, parameter κ_f was relevant only for set size 1 in Experiments 1 and 3. When a precue or retro-cue is presented, the focus stays on the cued object, ensuring that its binding remains at full strength ($c = 1$), whereas the bindings of the remaining objects decay for the entire retention interval following the cue. In this way, IM-DR implements the idea that a precue or retro-cue protects the cued item from decay (Matsukura, Luck, & Vecera, 2007; Pertzov, Bays, Joseph, & Husain, 2013). In case of a precue (Experiment 2), the focus stays on one item since encoding, thereby also maintaining it with increased precision κ_f.

The IM-DR gave an excellent fit to the data of Experiment 1, surpassing all other models, but it fared considerably worse on Experiments 2 and 3 (see Tables 5 and 6). The main problem of the IM-DR, particularly noticeable in Experiment 2, was that it overpredicted the cost of an invalid cue: When the precue was invalid, the focus of attention stayed on the precued item, thereby maintaining it with high precision. This assumption enabled the model to explain the good performance on valid-cue trials, but it implies that all other memory representations decay throughout the retention interval, so that when one of them is tested, performance is predicted to be much poorer than it actually was.

Binding by Synchrony and Interference in Phase Space

Some theorists have argued that the capacity limit of WM arises naturally from a neural mechanism for binding by synchrony (Horn, Sagi, & Usher, 1991; Lisman & Idiart, 1995; Raffone & Wolters, 2001; Vogel, Woodman, & Luck, 2001). Applied to a typical visual WM task, the idea is the following: Each color in an array is bound to its location by synchronous firing of the neurons coding the color and the neurons coding the corresponding location. Different color-location couples are kept separate by firing out of sync. Assuming a constant rate of firing for all neurons, the average time between spikes of neurons belonging to different color-location couples decreases with increasing set size. Further assuming that spike timing is noisy, the probability that two neurons from different couples fire in synchrony by chance increases as their temporal separation decreases. Accidental synchronization leads to disruption of memory performance by binding errors. Models of binding by synchrony often assume that neurons fire according to an oscillating pattern, and therefore describe the average interval between two subsequent spikes of a neuron as its phase. The capacity of a binding-by-synchrony model is limited because only a limited number of desynchronized color-location couples can be allocated within a given phase. This limit has been argued to account for a discrete, slot-like capacity limit (Luck & Vogel, 2013).

The basic idea of a capacity limit arising from a limited phase space as described above implies a model that is formally equivalent to the IM (ignoring, for the time being, the focus of attention), replacing the circular context space by the equally circular phase space: In a binding-by-synchrony model, the binding between the retrieval cue and the target feature is mediated by their common location in phase space (see Figure 14). Retrieval of a target feature must start from the probe as retrieval cue, which serves to select the probe's location in phase space (i.e., the time at which neurons coding the probe fire) and read out the feature information synchronized with it. To determine which spikes of feature-coding neurons are synchronized with the probed location,

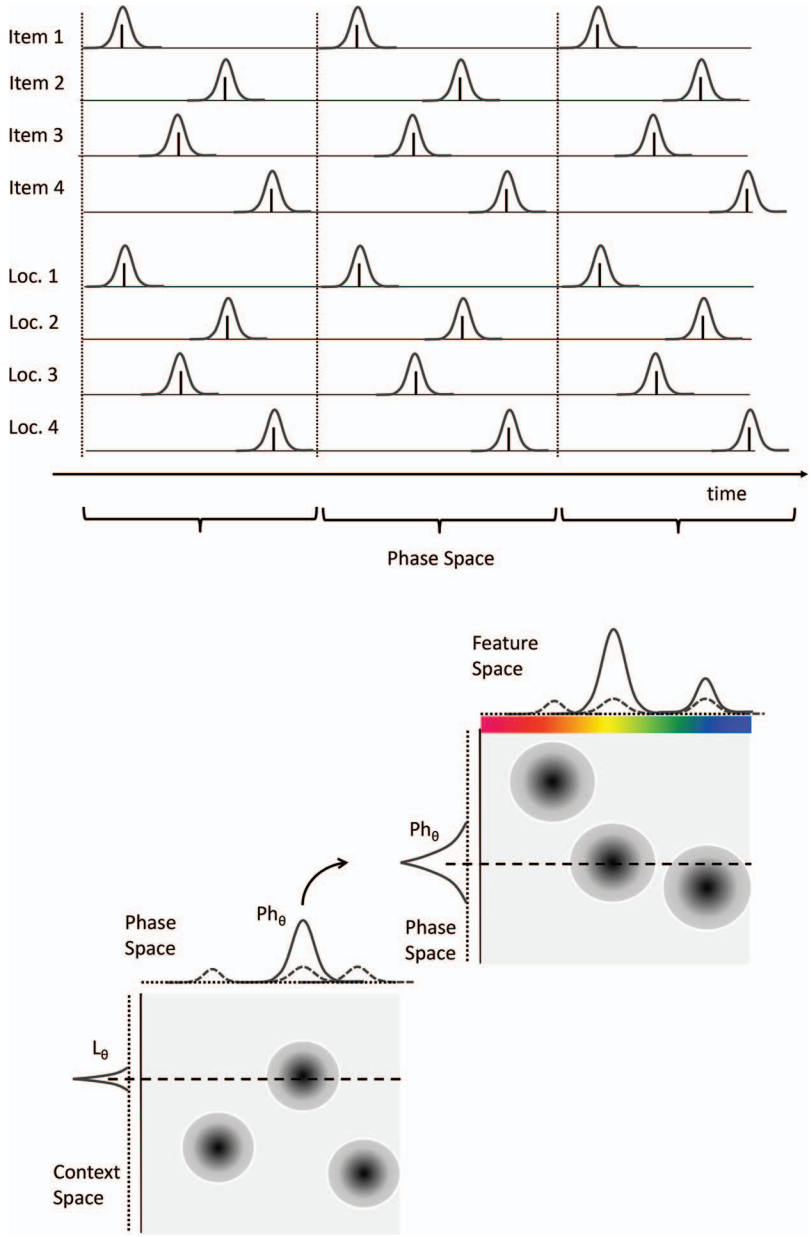

Figure 14. Binding by synchrony. The top panel shows spike trains over time by neurons representing four items, and neurons representing their corresponding locations. Items are bound to their locations by synchronous firing. Each spike is accompanied by a distribution showing its temporal uncertainty. The spike timing pattern repeats over successive phases. Dotted vertical boundaries demarcate the phase boundaries. The bottom panel shows the two-step process of cue-based retrieval. In the first step (bottom left) a location cue serves to retrieve the spike times of that location in phase space. In the second step (upper right) that spike time serves as retrieval cue to the item's color that is firing at roughly the same time in phase space. Both steps involve uncertainty—reflected in the dispersion of the distributions—due to spike time noise. See the online article for the color version of this figure.

the system must set a generalization gradient on time that is narrow enough to exclude most features from nontarget features but broad enough not to miss many target feature neurons that are no longer perfectly synchronized because of temporal noise. The target's distinctiveness therefore depends on its distance from nontargets in phase space, and the temporal precision of each neuron's firing. Hence, we can directly apply the IM to model the distinctiveness of representations bound by synchrony, mapping the circular context space of the IM onto phase space, and the contextual-gradient parameter s onto the temporal generalization gradient in phase space.

One implication of the equivalence between binding by synchrony and the IM is that binding by synchrony, without additional assumptions, does not lead to a slot-like discrete capacity limit, but rather to a gradual decline of performance with increasing set size n because of the gradual increase of interference: As the distance between target and nontarget in phase space decreases, the feature information returned for a retrieval cue includes a larger contribution from nontargets. Unless additional assumptions are made—for instance about mechanisms by which desynchronized neural assemblies annihilate each other as they get too close in phase space—there is no hard limit to how many content-context bindings can be encoded into a binding-by-synchrony model. The bare-bones binding-by-synchrony model is an interference model.

We therefore considered a version of the IM (including the focus of attention) that incorporates the idea of binding by synchrony. In this version, the degree to which nontarget information is superimposed on target information at retrieval does not depend on their proximity on the retrieval-relevant context dimension (i.e., spatial location or, in Experiment 4, color space) but rather on their proximity in phase space. We assumed that the n items are allocated equidistantly in phase space, thereby optimizing their distance. Because we do not know which item is assigned which location in phase space, we calculated predictions for all permutations of the n items and obtained the final prediction as the unweighted average across the permutations. Whereas in each permutation the nontargets temporally close to the target are weighted more strongly by the gradient in phase space—governed by parameter s—on average each nontarget receives the same weight. This makes parameter a redundant, and therefore we removed it from the model to increase its parsimony.

Tables 5 and 6 show that IM-Phase competed well with the standard IM on Experiment 1, but did considerably more poorly on Experiments 2 and 3. The reason for its poor performance is that it cannot capture the effect of spatial distance between target and nontargets (see Figure 10), because in IM-Phase, interference depends on distance in phase space, not in context space.

A Slot-Averaging Model With Binding Errors

The original Slot-Averaging model does not incorporate the possibility of binding errors that result in recall of nontargets. Here we consider two ways in which binding errors could arise in a slot model. First, Donkin, Tran, and Le Pelley (2015) present a slot model in which binding errors occur at encoding: Each item encoded into a slot is encoded with its correct context (e.g., its correct spatial location) with probability B, and with another context with probability $1 - B$. Second, Souza, Rerko, Lin, et al. (2014) describe a version of the Slot-Averaging model in which binding errors occur at retrieval: Each slot contains an item's feature together with its correct location, but both have limited precision. A retrieval cue in the location of one item matches to some degree the imprecise location representation of other nearby items, so that sometimes the wrong slot is identified as containing the item in the tested location. We fit two versions of a Slot-Averaging model with binding errors to the data of Experiment 1, a model inspired by Donkin et al. (2015) with binding errors at encoding, and the model of Souza et al. (2014) with binding errors at retrieval. The model with binding errors at encoding fit much better ($\Delta AIC = 915$), so for simplicity we only report results for that model.

In the Slot-Averaging-Binding model, up to K feature-context conjunctions (e.g., conjunctions of a color with a spatial location) are encoded in K slots. The probability of encoding each feature x_i in conjunction with a given context L_j is a decreasing function of the spatial distance between L_j and the location of feature x_i in context space:

$$P(x_i \mid L_j) = \frac{\exp[-sD(L_i, L_j)]}{\sum_{k=1}^{n} \exp[-sD(L_k, L_j)]}, \tag{21}$$

with n for the number of items in the array. At test, a retrieval cue is presented in the target context L_θ. This context is represented in a slot with probability $\min(1, K/n)$. If the tested context is in a slot, the feature in that slot is reported; otherwise the person guesses at random. Hence, the probabilities of recalling the target feature x_θ, of recalling a nontarget feature x_i ($i \neq \theta$), and of guessing are:

$$\begin{aligned} P(x_\theta \mid L_\theta) &= \min(1, K/n) \frac{1}{\sum_{k=1}^{n} \exp[-sD(L_k, L_\theta)]}, \\ P(x_{i(i\neq\theta)} \mid L_\theta) &= \min(1, K/n) \sum_{i\neq\theta}^{n} \frac{\exp[-sD(L_i, L_\theta)]}{\sum_{k=1}^{n} \exp[-sD(L_k, L_\theta)]}, \\ P(guess) &= 1 - \min(1, K/n) \end{aligned} \tag{22}$$

The predicted distribution of responses is a mixture of target and nontarget responses from a von-Mises distribution with shared precision parameter κ, and uniform guesses. This model fit the data of Experiments 1 to 3 better than the Slot-Averaging model, but still much worse than the IM (see Tables 5 and 6).

A Resource Model for Bindings

The Variable Precision model gives a good account of many aspects of the data from the present experiments, as well as other data from the continuous-reproduction paradigm (van den Berg et al., 2014). Its only discernible disadvantage compared with the IM is that it has no mechanism for explaining binding errors, as reflected in the intrusion of nontarget features in recall. Van den Berg and colleagues added nontarget intrusions to the model and found that this improves the model's fit (van den Berg et al., 2014). We replicated this observation: We fit the Variable Precision model with nontarget intrusions (model VP-A-NT in Van den Berg et al., 2014) and found that it outperformed all other models (ΔAIC relative to the IM = −130, −153, and −427 for Experiments 1 to 3, respec-

tively).[9] This model, however, is not a fully explanatory model because it provides no explanation for the nontarget intrusions and their increase with set size: Nontarget intrusions are simply assumed to occur by an unknown mechanism with a probability that, for an unknown reason, increases linearly with set size. The model does not incorporate assumptions about how the resource distribution over memory representations affects nontarget intrusions (e.g., are nontargets that receive a larger resource share more likely to intrude? Are nontarget responses more likely when the target has a low resource share?). The addition of nontarget intrusions compromises the model's explanation of the set-size effect because the set-size effect now arises from two sources, the need to share a constant resource, and the increase of nontarget intrusions with set size—the first source has a mechanistic interpretation but the second does not. When we find that the model reproduces the set-size effect well, we do not know to what extent the set-size effect is due to resource sharing and to what extent it is attributable to the unknown mechanism that generates nontarget intrusions. Because no mechanism generating nontarget responses is specified, the model offers no guidance on how to apply it to new experimental situations, such as the precue or the retro-cue paradigm. For these reasons, we do not consider the Variable-Precision model with nontarget errors as an explanatory model, but rather as a tool that helped to discover what needs to be done to improve the original Variable-Precision model: It needs to be augmented by a mechanism that generates nontarget intrusions.

Here we investigate one way in which a resource model—such as the Variable-Precision model—can explain binding errors. We assume that the resource determines not the precision of feature representations but the precision of bindings. The idea is illustrated in Figure 15. As set size increases, bindings become less precise because of resource dilution. This has two consequences: On the one hand it reduces the precision of the retrieved feature information; on the other hand, it leads to more overlap of nontargets with the target location in context space, resulting in a larger contribution of nontargets to the retrieved information. Thereby, the Variable-Precision Binding model can account with a single mechanism for the effect of set size on target precision and on the prevalence of nontarget intrusions.

We implemented this model with two precision parameters for the two-dimensional binding distributions, one for the feature dimension and one for the context dimension. Precision on the feature dimension, κ_x, was determined by the corresponding Fisher information J_x, and precision on the (spatial) context dimension, κ_s, was determined by J_s. Precision was variable on both dimensions. To avoid needing two variability parameters, we first drew J_x from a Gamma distribution with mean $\bar{J}_x$ and scale τ, and then multiplied J_x with a scaling parameter to obtain J_s. In this way, variability of binding precision was perfectly correlated, in keeping with the assumption that the fluctuation of a single resource determines binding precision on both dimensions. At retrieval, the probed location L_θ feeds through binding space to generate an activation distribution in feature space. The strength of activation of each feature x depends on how strongly it is bound to L_θ across all individual feature-location bindings. The binding strength from each of n feature-location pairs is the density of the two-dimensional binding distribution for that pair at the intersection of

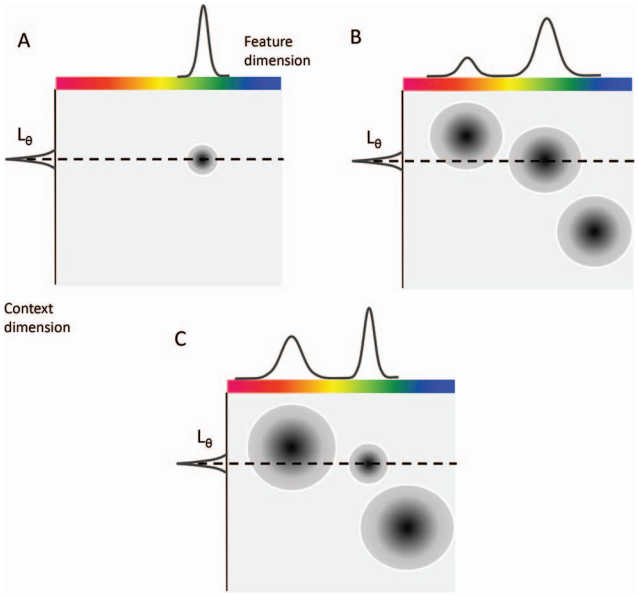

Figure 15. A resource limit on feature-context bindings. A: At set size 1, the entire resource can be concentrated on the one binding to be maintained, allowing for high precision in binding space. B: At set size 3, the resource must be shared among three bindings, so that they are represented with lower precision. C: A precued target receives a larger resource share at the expense of the two nontargets. Thereby the target is bound to its context with high precision, but the bindings of nontargets to their contexts have very low precision, so that they overlap substantially with the probe in context space. As a consequence, the model predicts that more nontarget information is retrieved when the target is precued (panel C) than when it is not (panel B). See the online article for the color version of this figure.

x with L_θ; this density is the product of the two von-Mises distributions describing binding precision in context space and feature space, respectively:

$$A(x \mid L_\theta) = \sum_{i=1}^{n} vMises(L_i, L_\theta, \kappa_s) \cdot vMises(x, x_i, \kappa_x) \qquad (23)$$

Table 5 shows that the Variable-Precision Binding model provides an excellent fit to the data of Experiment 1. The model fares poorly, however, with Experiment 2. The problem that causes the model's poor performance with the precue paradigm of Experiment 2 is illustrated in panel C of Figure 15: When the target is validly precued, the model concentrates a larger resource share on it, thereby increasing the precision of the target's binding to the probed location. This is beneficial insofar as it increases the precision with which the target feature is retrieved, but it comes at the expense of allocating less of the resource to nontarget bindings. As a consequence, the nontarget bindings overlap even more with the probed location than they do in the absence of a precue. Therefore, the model inevitably predicts that more nontarget information contributes to the retrieved activation in the presence of a valid precue than in its absence. This prediction is clearly contradicted by the data, which show virtually no nontarget intru-

[9] We fit the model separately to each cueing condition in Experiments 2 and 3.

sions in validly precued trials (see Figure 11). We did not fit the Variable-Precision Binding model to Experiment 3 because it is not clear how shifting resources to the postcued item after encoding can increase the precision of binding of the cued item.

General Discussion

In this article we proposed an interference-based model of WM for information in continuous feature spaces, such as the features of visual objects. The IM successfully accounted for several findings from experiments with the continuous-reproduction paradigm, which measures memory accuracy on a continuous scale. In particular, the IM gave a good quantitative account for six phenomena: The set-size effect on reproduction accuracy, the shape of the response distribution around targets, the frequency of nontarget responses, the contextual gradient on the prevalence of nontarget responses for both spatial and color probes, the beneficial effect of similarity between targets and nontargets, and the effects of precues and retro-cues, including the differences between these two kinds of cues. Of these six phenomena, the first two are also explained by the other two published explanatory models that we considered in the first round of competitive model fitting (see Table 1), whereas the last four are novel predictions of the IM.

The set-size effect is a benchmark finding that every theory of (visual) WM must explain, and all models do so, at least to a good approximation. All three models predict a slight deceleration in the rise of mean deviation over set size that was not present in the data (see Figure 3). The reason for this discrepancy is not clear—perhaps some participants simply give up on some trials with large sets and guess randomly although they could have provided a somewhat better response from memory.

The shape of the response distributions around targets is reasonably well explained by all three models, but as shown by the distribution of residuals (see Figure 13), only the Variable-Precision model and the IM provide a satisfactory fit. Both models achieve this through a mixture of von-Mises distributions with different precisions, though motivated by different assumptions. Whereas the Variable-Precision model incorporates substantial variability of precision as a fundamental feature, in the IM this variability arises from a mixture of responses that can and responses that cannot draw on target information in the focus of attention. When we controlled this variable in Experiments 2 and 3, and analyzed responses from invalidly cued trials, which can never draw on information in the focus of attention, we found little evidence for variability of precision in the shape of the response distribution.

The prevalence of nontarget responses has been observed before, but the IM is currently the only model explaining it. Van den Berg et al. (2014) included versions of slot models and resource models with a free parameter for nontarget responses in their factorial model comparison, but the occurrence of nontarget responses is not explained by the theoretical mechanisms incorporated in those models. The models say nothing about how the availability of slots or resources affects bindings, and therefore they cannot predict how binding errors—which result in nontarget intrusions—respond to experimental manipulations, such as variations of set size, of precueing and retro-cueing, or of the proximity of nontargets to the target in context space.

In contrast, the basic mechanisms of the IM entail the possibility of binding errors, and in addition uniquely predict the contextual gradient on them: Nontargets influence responses to the extent to which they are close to the target on the retrieval-relevant context dimension, that is, the dimension on which the probe distinguishes the target from the nontargets. Experiment 4 confirmed that the contextual gradient arises not only over spatial locations but also over the color dimension when color serves as the retrieval cue. This finding confirms the generality of the predicted contextual gradient, and in addition rules out the possibility that this gradient arises from perceptual competition between stimuli in close spatial proximity (Emrich & Ferber, 2012). The IM further predicts that reproduction accuracy increases as the average similarity of nontargets to the target increases. This prediction has been confirmed by Experiments 1 to 3. The prediction of a similarity benefit arises directly from the notion of superposition in the IM—the same mechanism that also gives rise to nontarget related responses. Other models incorporating nontarget responses—such as our Slot-Averaging-Binding and Variable-Precision-Binding models—also predict the similarity benefit, at least qualitatively, whereas models without a mechanism for nontarget intrusions do not predict the similarity benefit.

Finally, the present Experiments 2 and 3 are the first to test the novel predictions of the IM that precueing substantially reduces the set-size effect for validly cued targets, whereas retro-cueing, although still generating a benefit, reduces the set-size effect on accuracy only very little. This dissociation corroborates our assumption about the two roles of the focus of attention: When the target is focused already at encoding, a highly precise representation of it can be maintained in the focus. When the target representation is focused only after encoding, increasing its precision is no longer possible, but the focus can still shield it to some extent from interference arising from other information in WM, and from additional visual input (Makovski, Sussman, & Jiang, 2008; Souza et al., 2016). The latter mechanism can explain why a retro-cue improves memory even compared to a neutral condition with a shorter retention interval (Murray et al., 2013; Souza, Rerko, Lin, et al., 2014). The IM is currently the only computational model offering an explanation for this retro-cue benefit and its interaction with set size.

The two features of the IM that are key to its success are, first, a mechanism for bindings that explains how binding errors arise, and second, a focus of attention that contributes to explaining the set-size effect and accounts for the effects of pre- and retro-cues. In a second round of competitive model fitting we investigated the importance of these features. First, we augmented the Slot-Averaging model and the Variable-Precision model by a mechanism for bindings (models Slot-Averaging-Binding and VP-Binding). This improved their fit and enabled them to account for additional phenomena such as nontarget intrusions and the similarity benefit, but they still provided a poorer quantitative fit to the data across Experiments 1 to 3 than the IM. We also explored an alternative binding mechanism, binding by synchrony (model IM-Phase). This model failed to capture the finding that nontarget intrusions arise more often from nontargets close to the target in context space, and therefore provided a poorer fit than the IM to Experiments 2 and 3; for the same reason the IM-Phase model also cannot account for the contextual gradient effects in Experiment 4. Finally, we asked whether the IM could be com-

bined with a discrete capacity limit. This model, IM-DC, accounted for the data about as well as the IM across Experiments 1 to 3. Hence, our findings are compatible with a discrete capacity limit in visual WM, but not as an alternative to the IM but rather as an additional source of the WM capacity limit within the IM.

Our investigation of alternative model variants confirms that the success of the IM does not arise from a single assumption in the model that could be transplanted into other models, and that none of the mechanisms we implemented in the IM are easily replaced by alternatives discussed in the literature.

Is Information in Working Memory Discrete or Continuous?

One key assumption in the IM is that representations in WM are best described as varying on a number of continuous dimensions, including their precision and their strength of feature-context binding. In this regard the IM differs from discrete-state models of WM, such as slot models. In discrete-state models it is assumed that each item's representation is in one of a limited number of possible states. For instance, in a simple slot model, each visual object can either be represented in one slot, or not represented in WM at all. Therefore, if that object is tested, the memory system can be in one of two states: Either it has information about the object with a degree of precision provided by a slot, or it has no information about the object. Support for discrete state models comes primarily from two lines of research on visual WM, which we address in turn.

On plateaus. The first line of evidence supporting a discrete capacity limit in WM revealed that neural correlates of WM set size, such as the contralateral delay activity (CDA) and the BOLD signal in the inferior intraparietal sulcus (IPS), reach plateau at the set size at which people's capacity limit is reached (Kuo, Stokes, & Nobre, 2012; Vogel, McCollough, & Machizawa, 2005; Xu & Chun, 2006). This is to be expected if any objects exceeding a person's discrete capacity are not represented in WM at all, so that they contribute nothing to the neural activity involved in maintenance. A related observation is that the precision of memory representations estimated from the two-parameter mixture model reaches a plateau at the estimated capacity limit (Zhang & Luck, 2008). This plateau in precision is predicted by a discrete-capacity model because objects are either represented in a slot with a certain degree of precision, or not at all, so that increasing set size beyond capacity simply leads to less objects being represented, but not to a further dilution of precision.

Whereas each of these findings in isolation provides evidence for predictions from discrete-capacity models of WM, jointly they do not: The plateaus on neurophysiological correlates of WM load and the plateau on precision are predicted by two different, mutually incompatible versions of slot theory. The rise of neural correlates of set size to a plateau set by capacity is predicted by a simple slot model: Each object in the memory array receives exactly one slot until capacity is exceeded, and any additional objects receive no slot. In this model, only one slot is used at set size 1, two slots are used at set size 2, and so on until set size reaches K, at which point the number of slots used remains constant at K. With the additional assumption that the CDA, the BOLD signal in inferior IPS, or any other load-sensitive neural signal reflects the number of slots used, the observed plateau is predicted by the simple slot model. In contrast, the finding that precision in the two-parameter mixture model drops with increasing set size up to K and then plateaus is predicted by the Slot-Averaging model: In this model, at set size 1 the single object receives all K slots, and therefore can be recalled with high precision, whereas at set size 2 each object receives only $K/2$ slots, and so on, until at set size K each object receives exactly one slot, and recall precision reaches its minimum at the precision of a single slot.

The problem for slot theorists is that the simple slot model does not predict the change of precision over set size up to the plateau: In the simple slot model, each object in WM receives exactly one slot, and hence precision is predicted to be constant across all levels of set size, contrary to the data. This is the reason why Zhang and Luck (2008) proposed the Slot-Averaging model. The Slot-Averaging model assumes that all available slots are used regardless of set size. Assuming that neural activity reflects the number of slots engaged, this model predicts that the neural activity invested for maintenance is constant across all set sizes, and hence it cannot predict the initial rise of the CDA, or other neural signals, with set size. It follows that there is currently no slot model that is compatible with the behavioral data—in particular the decrease of precision over small set sizes—and at the same time predicts the rise of a neural signal up to a plateau at capacity.

The IM does not predict a plateau in the precision parameter of the two-parameter mixture model, or in the SD_{mix} parameter implied by it. (We use SD_{mix} for the standard-deviation estimated from the two-parameter mixture model to distinguish it from the SD parameter in the IM). The IM rather predicts a decelerating increase of SD_{mix} with increasing set size. Figure 16 shows the result of a simulation with the IM, using the best-fitting parameters of Experiment 1 to generate synthetic data, which we then fit with the 2-parameter mixture model separately for each simulated subject. A rapidly decelerating increase in SD_{mix} is difficult to distinguish from a bilinear function that initially rises and then reaches a plateau. Figure 16 shows the averaged bilinear functions fitted to the synthetic SD_{mix} data; although the SD_{mix} do not strictly follow a bilinear function, the bilinear fit is very good. In general, it is difficult to determine what competing models predict for a parameter of the mixture model when these models do not assume that the response distribution is a mixture of responses from two different states. Therefore, the assumption of a plateau on SD_{mix} implied by the Slot-Averaging model is best tested by competitively fitting the Slot-Averaging model—which implies the plateau—and other models—not implying the plateau—to response distributions. This is the approach we took in the present article, and the IM provided a much better account of the data than the Slot-Averaging model. Therefore, the fact that the IM does not predict a plateau in SD_{mix} does not appear to be a serious problem for the model. This conclusion is in agreement with a simulation study by van den Berg and Ma (2014), which revealed that summary statistics reflecting the plateau are not diagnostic for distinguishing between competing models of visual WM.

We acknowledge that the correlation between estimates of K and the increase of the CDA with set size (Kuo et al., 2012; Vogel et al., 2005) provides support for the discrete-capacity notion. We add two notes of caution, however. First, these correlations are based on very small sample sizes compared to what is recommended for correla-

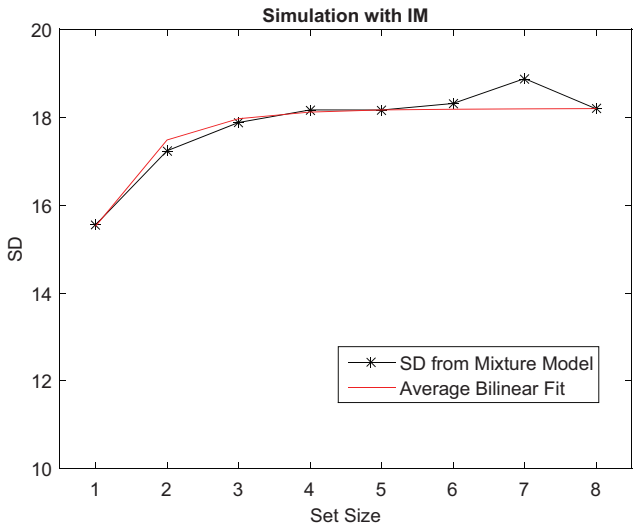

Figure 16. Mean estimated SD_{mix} from applying the two-parameter mixture model to synthetic data from a simulation with IM (black line), and mean of bilinear functions fitted to each subject's data (red line). The simulation generated 200 synthetic subjects, using the following parameter means (*SD* in parentheses) $b = 0.16$ (0.11), $a = 0.20$ (0.16), $s = 7.1$ (5.8), $\kappa = 7.8$ (1.7), $\kappa_f = 48$ (20), $r = 0$ (0). For each subject 300 trials were run per set size. The mean of the bilinear functions is not itself bilinear because it averages across bilinear functions that change slope at different values of set size. See the online article for the color version of this figure.

tional research. Second, correlations of any variable with *K* are meaningful only insofar as *K* is a meaningful parameter—which it is only in discrete-state models. The observed correlations of *K* with neural markers of visual WM confirm a prediction of the simple slot model, but interpreting this finding as evidence against alternative models would be circular: If discrete-state models are not true, there is no fixed capacity measured by *K*, and it is unclear what the correlation means. For instance, if the IM were true, people's performance on a continuous-reproduction task would be determined by their values of the IM parameters, none of which maps directly to *K*. To illustrate, we correlated the IM parameter values in our simulation with the *K* estimates; these correlations are listed in Table 7. Assuming that the IM is true, then what *K* measures is a combination of high precision of context representations (high parameter *s*) and low contributions of nontarget information (*a*) and noise (*b*) relative to the contribution of cue-based retrieval (*c*). The latter can be interpreted as the relative strength of the feature-context bindings that drive cue-based retrieval. Thus, the IM decomposes variance in *K* into two components: Variance in contextual precision and variance in the strength of bindings. A potentially interesting question for future research is whether correlates of *K* are correlates of one or the other or both of these sources of variance in the IM.

On guessing. A second source of evidence for discrete-state models of WM comes from stimulus-independent responses that are consistent with guessing. For instance, the extended fat tails of the error distribution of continuous reports have been interpreted as reflecting the subset of trials on which participants guess at random (Zhang & Luck, 2008). If people sometimes provide a response informed by memory, and sometimes guess, then their responses arise from two qualitatively different mental states at test, a memory state and a guessing state. Evidence for two distinct states at test comes also from short-term recognition tests of visual WM. Data from several recent studies were better explained by high-threshold models of recognition than by signal-detection models (Donkin, Nosofsky, Gold, & Shiffrin, 2013; Donkin, Tran, & Nosofsky, 2014; Rouder et al., 2008). Whereas high-threshold models assume that recognition decisions rely on discrete states—a memory-informed state and a guessing state—signal-detection models assume that they rely on a memory signal varying continuously in strength.

Discrete memory states imply discrete states at test—in a slot model, if a person has a representation of the target in a slot, they can use that information, but if they do not they can only guess. The reverse implication does not hold: If people guess on a subset of trials, it does not follow that their memory states are discrete (Kellen & Klauer, 2015). Assuming that memory representations vary continuously in strength or precision, as in the Variable-Precision model or the IM, people could still resort to guessing on some trials. The IM-guess model incorporates this assumption and fit the data about as well as the original IM. In the IM there is never a memory state without any information about the target, hence guessing would entail throwing away information. This would be suboptimal if people tried to maximize accuracy, but people's response strategies are not necessarily geared toward maximizing only accuracy. Retrieving weak information is arguably a slow process (for a formal model relating strength of memory for continuous features to retrieval speed see Smith, 2016), so for participants aiming for a balance of accuracy and speed it might be optimal to guess when the state of memory signals very weak information.

To conclude, we find the interpretations of the fat tails in the error distribution as reflecting background noise or as reflecting guessing equally tenable. One way to empirically adjudicate between the two interpretations could be to offer participants a "guessing button" in the continuous-reproduction task. If the fat tails of the error distribution reflect guessing, and if people know when they guess, then they should choose the guessing button instead of selecting a response at random when they are in the guessing state, and the fat tails should disappear. In contrast, in the background-noise version of the IM there is no discrete guessing state, and people do not know when their responses are driven by background noise, so that providing a guessing button should not eliminate the fat tails of the error distribution.

Neuronal Mechanisms

A good cognitive model should assume mechanisms that can be implemented in the brain. Therefore, we next discuss possible neu-

Table 7
Correlates of Interference-Model Parameters With Capacity Estimate K and Precision Estimate Kappa From the Two-Parameter Mixture Model

IM parameter	Mixture model capacity (K)	Mixture model precision at set size 1	Mixture model precision at set size 8
b	−.59	.02	−.16
a	−.47	.01	−.21
s	.32	−.13	−.06
κ	−.01	−.74	−.50
κ_f	−.03	−.43	−.13

ronal realizations of the mechanisms incorporated in the IM, and implications for potential neural signatures of those mechanisms.

In the IM, memory rests on two components, the focus of attention and a binding mechanism. Whereas representation of the object in the focus of attention can be maintained through persistent firing of neurons selectively coding that object's features (e.g., its color and its location), bindings cannot be implemented by persistent firing of feature-coding neurons alone. Several solutions have been proposed for representing bindings in neural networks, and at least two of them are compatible with our proposal to model bindings as distributions in a two-dimensional binding space. One is to code feature bindings through conjunction-coding neurons, that is, neurons selectively responding to the conjunction of two features (e.g., a particular color in a particular location). The second version of the "palimpsest model" of Matthey et al. (2015), which has much in common with the IM, uses this solution. Alternatively, bindings could be encoded through rapid synaptic changes (Mongillo, Barak, & Tsodyks, 2008; Postle, 2015; Stokes, 2015).

Based on these considerations we can make the following predictions concerning the neural signatures of visual WM. First, it should be possible to read out information about the object in the focus of attention from the neural signal throughout the retention interval using multivariate pattern analysis (Riggall & Postle, 2012) or inverted encoding modeling (Ester, Sprague, & Serences, 2015). This information should be more precise after a precue than after a retro-cue. This information should reflect features, not their conjunctions with contexts. For instance, the same inverted-encoding model should be able to reconstruct the orientations of stimuli in all locations and regardless of their color. A further plausible expectation is that the representation in the focus of attention is selectively vulnerable to interventions perturbing neural firing patterns. In support of this hypothesis, Zokaei, Manohar, Husain, and Feredoes (2014) found that TMS over motion-sensitive visual cortex abolished the effect of a retro-cue to one of two motion stimuli held in WM.

In contrast, information from objects outside the focus of attention (i.e., not-cued objects) should not be detectable in the same way in neural activity during the retention interval, because that information is maintained primarily through bindings between the target feature and the retrieval-relevant context. If these bindings are maintained through firing of conjunction-coding neurons, it should be possible to read out conjunction information by tuning the decoding or inverted-encoding algorithm to conjunctions between target features and their contexts. If bindings are maintained through synaptic changes, readout of information on noncued items from neural activity should not be possible at all during the retention interval. There is emerging evidence for representations in WM that are "activity-silent" during the retention interval (Stokes, 2015) in particular during periods when other information in WM is cued as relevant (LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012), lending plausibility to the idea that bindings are represented by rapid synaptic changes.

Toward a Unification of Theories of Working Memory for Verbal, Spatial, and Visual Contents

The IM is more complex than models based on discrete slots or resources—in its basic form it has more free parameters, and its behavior rests on more different mechanisms than those competitors. When focusing only on the literature on visual WM one might find that the gain in accuracy and explanatory power is not worth the added complexity. Taking a broader perspective that includes the literature on WM for other materials and studied with other paradigms, the IM offers a simplification of the theoretical landscape because it lays the ground for conceptual unification. In theorizing about WM phenomena beyond the visual domain slot models or resource models play a minor role at best, because they provide no means for explaining well-established benchmark findings such as serial-position curves, transposition gradients, similarity effects, and many others, which are well explained by interference models (Lewandowsky & Farrell, 2008).

Research on WM with verbal and with spatial materials has revealed remarkably similar patterns of results (Hurlstone, Hitch, & Baddeley, 2013), strongly suggesting that WM operates according to the same principles regardless of material. Moreover, capacity measures of visual WM are substantially correlated with measures for verbal and spatial WM (Cowan et al., 2005; Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014). Therefore, we find it unlikely that visual WM should be fundamentally different from WM for other materials. Adopting fundamentally different models for visual WM and for nonvisual WM creates the intellectually uncomfortable situation of applying incompatible theoretical ideas to the same construct when studied with different materials and experimental paradigms. Adopting the IM paves the way for a more coherent theoretical framework for WM. The prospect of theoretical unification should far outweigh the modest increase in complexity of the IM compared to slot models and resource models of visual WM.

The IM was motivated by successful theories of working memory for discrete—mostly verbal—information. Its conceptual building blocks were drawn from two lines of theoretical work, computational models of serial recall, and the three-component theory of working memory. We next revisit the relation of the IM to these precursors.

The mechanisms of interference. The IM incorporates principles of interference-based computational models of working memory (Botvinick & Watanabe, 2007; Brown et al., 2000; Burgess & Hitch, 2006; Henson, 1998; Lewandowsky & Farrell, 2008; Oberauer, Lewandowsky, et al., 2012). Contents are bound to contexts through temporary bindings, and access to contents depends on cue-based retrieval, using the appropriate context representation as a cue. Interference arises when cue-based retrieval returns not only information from the target but also from other items currently in WM.

In interference models for discrete items such as digits or words, two kinds of interference can be distinguished, interference by superposition and interference by confusion of items (Oberauer, Lewandowsky, et al., 2012). Both kinds of interference also arise in our interference model for continuous features, albeit with different consequences.

Interference by *superposition* arises when representations of multiple items that need to be kept separate are added (i.e., superimposed) in the same representational space, resulting in their mutual distortion. In the IM, interference by superposition arises from two sources: Superposition of context-independent item representations is reflected in the summation across items in the equation for A_a; superposition of content-context bindings is reflected in the summation of activation from cue-based retrieval in the equation for A_c. The latter component is graded by the simi-

larity between contexts, such that features of nontargets bound to contexts similar to that of the target contribute more strongly to the mixture of activation in feature space. This effect of context similarity is captured by the contextual gradient parameter *s*.

In tests of WM using discrete items a second kind of interference is observed, *similarity-based confusion*. Lists of similar sounding letters or words are recalled worse than lists of dissimilar verbal items, in part because similar items tend to be confused with each other more often (Conrad, 1964; Poirier & Saint-Aubin, 1996; Tehan, Hendry, & Kocinski, 2001). In an array of continuously varying uni-dimensional features such as colors or orientations, the similarity between items is their proximity in feature space. Because of limited precision of feature memory—governed by the precision parameters κ and κ_f—features are likely to be confused with neighboring features in feature space, analogous to confusion by similarity in the verbal domain. However, whereas confusions of discrete items are typically regarded as errors, confusing a color with a neighboring color in color space is regarded as a highly accurate reproduction in continuous-reproduction tasks. As a consequence, high similarity between items in a memory set is not detrimental in a continuous-reproduction paradigm. In fact, it is beneficial, because it mitigates the distorting effect of superposition (see Figures 5 and 6). Therefore, the same mechanisms of cue-based retrieval and interference correctly predict the opposite effects of interitem similarity in tests of WM for discrete items and for continuous features.

Interference models of (working) memory often assume a role for proactive interference between trials. In fact, the demonstration of proactive interference has once motivated the development of interference theories of short-term memory (STM) as alternatives to decay theories (Keppel & Underwood, 1962). Proactive interference arises naturally in models that assign temporal context an important role for cueing retrieval of specific events (Brown et al., 2007). Including a temporal context to distinguish between successive trials in an experiment would be a natural extension of the IM. So far we did not include temporal context because we were unable to find unambiguous evidence for proactive interference in the continuous-reproduction task. In the experiments reported here we were unable to find an above-chance tendency for reporting the target of the previous trial, or the color presented in the previous trial at the current trial's target location. In a separate series of experiments designed for revealing proactive interference in continuous reproduction, Oberauer, Awh, and Sutterer (in press) failed to find any evidence for intrusions of past information. That said, temporal distinctiveness between trials does affect performance in continuous reproduction (Souza & Oberauer, 2015). This leaves open the possibility that proactive interference contributes to performance in this paradigm, but not through intrusions of specific information from previous trials, but rather in a more general fashion, for instance by raising the chronic activation of all colors presented recently—raising the background noise component of the IM—or in an increased uncertainty of whether information retrieved from memory comes from the current trial—potentially raising people's tendency to guess in the IM-guess variant. At present, we know too little about whether and how proactive interference affects visual WM for simultaneously presented arrays for incorporating it into the IM with any confidence.

Three embedded components of working memory. The IM incorporates ideas from our broader theoretical framework of working memory, the three-embedded-components theory (Oberauer, 2009). The context-independent activation component A_a reflects the contribution of activated LTM, responsible for holding representations likely to be relevant in a state of heightened retrievability. The context-dependent activation component A_c reflects the contributions of the direct-access region, responsible for maintaining temporary bindings. The focus of attention, responsible for selecting an item for (cognitive) action, shields that item to some extent from interference, and maintains it with higher precision than items outside the focus.

The distinction between context-independent memory (i.e., memory for the items in the current set without information about their context within the set) and context-dependent memory (i.e., memory for bindings between items and their context) has a long tradition in research on WM. It features in studies of list recall as the distinction between item memory and order memory (Healy, 1974; Shiffrin & Cook, 1978). It features in studies of recognition as the distinction between familiarity and recollection (Atkinson, Herrmann, & Wescourt, 1974; Monsell, 1978; Oberauer, 2008). Therefore, the IM is well poised for an extension to recognition paradigms. The prospect of applying the IM to recognition engenders the expectation that the distinction between familiarity and recollection, which has been useful to account for short-term recognition results with verbal materials (Berman, Jonides, & Lewis, 2009; Göthe & Oberauer, 2008; Oberauer, 2008), will also be applicable to visual materials. In light of the findings discussed above, suggesting that single-process high-threshold models are best suited for explaining visual short-term recognition, we regard recognition paradigms of visual WM as an important and challenging test case for further work with the IM.

The concept of a focus of attention in the three-embedded-components framework plays a key role in the IM. The main reason for assuming a focus of attention within WM is theoretical: Although WM is often studied with tests that require little more than maintaining given information over a retention interval, there is broad consensus that the main function of WM is to hold information available for processing. Processing requires selective access to individual elements in a memory set held in WM. Selective access implies an attentional mechanism that selects the content required for the upcoming cognitive operation over other contents in WM (Oberauer, 2009). Selection usually requires unambiguously singling out one element from the current memory set. Therefore, the focus of attention typically holds one item or chunk, although it is not structurally limited to a single element (Oberauer, 2013; Oberauer & Hein, 2012). There is converging evidence from several experimental paradigms supporting the concept of a focus of attention in WM that meets these requirements (Oberauer & Hein, 2012; Olivers et al., 2011).

In the context of the IM the focus of attention plays two roles. One is to shield the item in the focus from extraneous information. This shielding effect follows directly from the function of the focus as a selection device within WM, and therefore we assume that it applies whenever the focus of attention is directed to an item in WM, whether this happens during or after encoding. The second role is to represent an item with especially high precision. We believe that this capability arises when the focus of attention in WM is aligned with the focus of perceptual attention—in the case of visual arrays: visual-spatial attention—at the time of encoding. Perceptual attention enables

encoding of one object with high precision, and if the focus of attention in WM is directed to that object from the start, its high-precision representation can be maintained. One way in which this could be accomplished is by maintaining a high-resolution pattern of neural firing throughout the retention interval.

Different from the context-independent activation of features of all objects in an array, the content of the focus of attention is linked to a specific context: The focus represents not only the object's features but also its location in context space. This is why in the IM we assume that the high-precision representation of the focused object contributes to the activation distribution only if that object is the target: The feature representation in the focus is read out only if the retrieval cue matches the context representation in the focus. For the same reason, a representation consisting of high-resolution firing patterns can contain only one feature-context conjunction: Holding two (or more) features together with two (or more) contexts by neural firing patterns would raise a binding problem: Which feature goes with which context? In the IM, bindings between features and contexts are held—with lower resolution—outside of the focus; there is no binding mechanism within the focus.

Conclusion

Memory is governed by the principles of cue-based retrieval and interference. These principles are incorporated in the best contemporary models of working memory for sets of discrete items. Here we show that the same principles can also be applied to the recall of continuously varying visual features. A computational model based on these principles gives a better quantitative account of fine-grained distributions of recall responses than slot or resource models. These results pave the way for a unified theory of working memory for verbal, spatial, and visual information.

References

Adam, K. C. S., Mance, I., Awh, E., & Vogel, E. K. (2013). Whole-report procedures reveal bimodal distribution of visual memory precision within a single array. *Journal of Vision, 13,* 1362–1362. http://dx.doi.org/10.1167/13.9.1362

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

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience, 31,* 1128–1138. http://dx.doi.org/10.1523/JNEUROSCI.4125-10.2011

Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 101–146). Potomac, MD: Erlbaum.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning & Verbal Behavior, 14,* 575–589. http://dx.doi.org/10.1016/S0022-5371(75)80045-4

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

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers, 28,* 584–589. http://dx.doi.org/10.3758/BF03200546

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133,* 83–100. http://dx.doi.org/10.1037/0096-3445.133.1.83

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,* 7. http://dx.doi.org/10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision, 11,* 6. http://dx.doi.org/10.1167/11.10.6

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854. http://dx.doi.org/10.1126/science.1158023

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia, 49,* 1622–1631. http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.023

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 317–333. http://dx.doi.org/10.1037/a0014873

Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: Behavioral and neural dissociations. *PLoS ONE, 9,* e90808. http://dx.doi.org/10.1371/journal.pone.0090808

Botvinick, M., & Watanabe, T. (2007). From numerosity to ordinal rank: A gain-field model of serial order representation in cortical working memory. *The Journal of Neuroscience, 27,* 8636–8642. http://dx.doi.org/10.1523/JNEUROSCI.2110-07.2007

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24,* 981–990. http://dx.doi.org/10.1177/0956797612465439

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review, 120,* 85–109. http://dx.doi.org/10.1037/a0030779

Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 77–102. http://dx.doi.org/10.1037/0278-7393.26.1.77

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181. http://dx.doi.org/10.1037/0033-295X.107.1.127

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. http://dx.doi.org/10.1016/j.jml.2006.08.005

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84. http://dx.doi.org/10.1111/j.2044-8295.1964.tb00899.x

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press. http://dx.doi.org/10.4324/9780203342398

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 731–747. http://dx.doi.org/10.1037/a0029687

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive

aptitudes. *Cognitive Psychology, 51,* 42–100. http://dx.doi.org/10.1016/j.cogpsych.2004.12.001

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. http://dx.doi.org/10.1177/0956797611430961

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

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review, 22,* 170–178. http://dx.doi.org/10.3758/s13423-014-0675-5

Donkin, C., Tran, S. C., & Le Pelley, M. (2015). Location-based errors in change detection: A challenge for the slots model of visual working memory. *Memory & Cognition, 43,* 421–431. http://dx.doi.org/10.3758/s13421-014-0487-x

Donkin, C., Tran, S. C., & Nosofsky, R. (2014). Landscaping analyses of the ROC predictions of discrete-slots and signal-detection models of visual working memory. *Attention, Perception, & Psychophysics, 76,* 2103–2116. http://dx.doi.org/10.3758/s13414-013-0561-7

Emrich, S. M., & Ferber, S. (2012). Competition increases binding errors in visual working memory. *Journal of Vision, 12,* 1–16. http://dx.doi.org/10.1167/12.4.12

Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *The Journal of Neuroscience, 33,* 6516–6523. http://dx.doi.org/10.1523/JNEUROSCI.5732-12.2013

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron, 87,* 893–905. http://dx.doi.org/10.1016/j.neuron.2015.07.013

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51,* 115–135. http://dx.doi.org/10.1016/j.jml.2004.03.007

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications, 3,* 1229. http://dx.doi.org/10.1038/ncomms2237

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17,* 134–141. http://dx.doi.org/10.1016/j.tics.2013.01.010

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology, 20,* 177–182. http://dx.doi.org/10.1016/j.conb.2010.03.005

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience, 31,* 8502–8511. http://dx.doi.org/10.1523/JNEUROSCI.0208-11.2011

Göthe, K., & Oberauer, K. (2008). The integration of familiarity and recollection information in short-term recognition: Modeling speed-accuracy trade-off functions. *Psychological Research, 72,* 289–303. http://dx.doi.org/10.1007/s00426-007-0111-9

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience, 15,* 1176–1194. http://dx.doi.org/10.1162/089892903322598139

Gunseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. L. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin & Review, 22,* 1474–1474. http://dx.doi.org/10.3758/s13423-015-0914-4

Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. H. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience, 1,* 235–241. http://dx.doi.org/10.1038/681

Hale, S., Myerson, J., Emery, L. J., Lawrence, B. M., & Dufault, C. (2007). Variations in working memory across the life span. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 194–224). New York, NY: Oxford University Press.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning & Verbal Behavior, 13,* 644–655. http://dx.doi.org/10.1016/S0022-5371(74)80052-6

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36,* 73–137. http://dx.doi.org/10.1006/cogp.1998.0685

Horn, D., Sagi, D., & Usher, M. (1991). Segmentation, binding, and illusory conjunctions. *Neural Computation, 3,* 510–525. http://dx.doi.org/10.1162/neco.1991.3.4.510

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10,* 1–27. http://dx.doi.org/10.1167/10.2.24

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. http://dx.doi.org/10.1037/a0034221

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193–224. http://dx.doi.org/10.1146/annurev.psych.59.103006.093615

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. http://dx.doi.org/10.1037/0033-295X.99.1.122

Kalogeropoulou, Z., Jagadeesh, A. V., Ohl, S., & Rolfs, M. (2016). Setting and changing feature priorities in visual short-term memory. *Psychonomic Bulletin & Review*. Advance online publication. http://dx.doi.org/10.3758/s13423-016-1094-6

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217. http://dx.doi.org/10.1037/0096-3445.133.2.189

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review, 122,* 542–557. http://dx.doi.org/10.1037/a0039251

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning & Verbal Behavior, 1,* 153–161. http://dx.doi.org/10.1016/S0022-5371(62)80023-1

Kruijne, W., Brascamp, J. W., Kristjánsson, Á., & Meeter, M. (2015). Can a single short-term mechanism account for priming of pop-out? *Vision Research, 115,* 17–22. http://dx.doi.org/10.1016/j.visres.2015.03.011

Kuo, B.-C., Stokes, M. G., & Nobre, A. C. (2012). Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience, 24,* 51–60. http://dx.doi.org/10.1162/jocn_a_00087

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43,* 149–164. http://dx.doi.org/10.1016/S0042-6989(02)00402-9

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience, 25,* 127–142. http://dx.doi.org/10.1162/jocn_a_00305

Lepsien, J., Thornton, I., & Nobre, A. C. (2011). Modulation of working-memory maintenance by directed attention. *Neuropsychologia, 49,* 1569–1577. http://dx.doi.org/10.1016/j.neuropsychologia.2011.03.011

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. http://dx.doi.org/10.1016/S0079-7421(08)00001-7

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience, 24,* 61–79. http://dx.doi.org/10.1162/jocn_a_00140

Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition, 17,* 356–372. http://dx.doi.org/10.1080/13506280701766313

Lindsay, P. H., Taylor, M. M., & Forbes, S. M. (1968). Attention and multidimensional discrimination. *Perception & Psychophysics, 4,* 113–117. http://dx.doi.org/10.3758/BF03209520

Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science, 267,* 1512–1515. http://dx.doi.org/10.1126/science.7878473

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17,* 391–400. http://dx.doi.org/10.1016/j.tics.2013.06.006

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research, 62,* 81–92. http://dx.doi.org/10.1007/s004260050043

Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 369–380. http://dx.doi.org/10.1037/0278-7393.34.2.369

Makovski, T., Watson, L. M., Koutstaal, W., & Jiang, Y. V. (2010). Method matters: Systematic effects of testing procedure on visual working memory sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1466–1479. http://dx.doi.org/10.1037/a0020851

Matsukura, M., Luck, S. J., & Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics, 69,* 1422–1434. http://dx.doi.org/10.3758/BF03192957

Matthey, L., Bays, P. M., & Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLoS Computational Biology, 11,* e1004003. http://dx.doi.org/10.1371/journal.pcbi.1004003

Mazyar, H., van den Berg, R., & Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision, 12,* 10. http://dx.doi.org/10.1167/12.6.10

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science, 319,* 1543–1546. http://dx.doi.org/10.1126/science.1150769

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology, 10,* 465–501. http://dx.doi.org/10.1016/0010-0285(78)90008-7

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience, 17,* 1410–1417. http://dx.doi.org/10.1038/nn.3807

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1021–1035. http://dx.doi.org/10.1037/0096-1523.29.5.1021

Müller, H. J., Töllner, T., Zehetleitner, M., Geyer, T., Rangelov, D., & Krummenacher, J. (2010). Dimension-based attention modulates feed-forward visual processing. *Acta Psychologica, 135,* 117–122. http://dx.doi.org/10.1016/j.actpsy.2010.05.004

Murdock, B. B., Jr., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74,* 137–143. http://dx.doi.org/10.1037/h0024507

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention restores discrete items to visual short-term memory. *Psychological Science, 24,* 550–556. http://dx.doi.org/10.1177/0956797612457782

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269. http://dx.doi.org/10.3758/BF03213879

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1524–1539. http://dx.doi.org/10.1037/xlm0000015

Oberauer, K. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50,* 257–269. http://dx.doi.org/10.1026//1618-3169.50.4.257

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,* 439–459. http://dx.doi.org/10.1037/0278-7393.34.3.439

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation: Advances in Research and Theory, 51,* 45–100. http://dx.doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2013). The focus of attention in working memory - from metaphors to mechanisms. *Frontiers in Human Neuroscience, 7.* http://dx.doi.org/10.3389/fnhum.2013.00673

Oberauer, K., Awh, E., & Sutterer, D. W. (in press). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/xlm0000302

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012). Interference between maintenance and processing in working memory: The effect of item-distractor similarity in complex span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 665–685. http://dx.doi.org/10.1037/a0026337

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

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55,* 601–626. http://dx.doi.org/10.1016/j.jml.2006.08.009

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review, 18,* 10–45. http://dx.doi.org/10.3758/s13423-010-0020-6

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. http://dx.doi.org/10.3758/s13423-012-0272-4

Oberauer, K., Sß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York, NY: Oxford University Press.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences, 15,* 327–334. http://dx.doi.org/10.1016/j.tics.2011.05.004

Orhan, A. E., & Ma, W. J. (2015). Neural population coding of multiple stimuli. *The Journal of Neuroscience, 35,* 3825–3841. http://dx.doi.org/10.1523/JNEUROSCI.4097-14.2015

Orhan, A. E., Sims, C. R., Jacobs, R. A., & Knill, D. C. (2014). The adaptive nature of visual working memory. *Current Directions in Psychological Science, 23,* 164–170. http://dx.doi.org/10.1177/0963721414529144

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

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psy-*

chology: Human Perception and Performance, 39, 1224–1231. http://dx.doi.org/10.1037/a0030947

Pertzov, Y., Dong, M. Y., Peich, M.-C., & Husain, M. (2013). Forgetting what was where: The fragility of object-location binding. *PLoS ONE, 7,* e48214. http://dx.doi.org/10.1371/journal.pone.0048214

Pertzov, Y., & Husain, M. (2014). The privileged role of location in visual working memory. *Attention, Perception, & Psychophysics, 76,* 1914–1924. http://dx.doi.org/10.3758/s13414-013-0541-y

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review, 20,* 1282–1289. http://dx.doi.org/10.3758/s13423-013-0460-x

Pinto, Y., Sligte, I. G., Shapiro, K. L., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review, 20,* 732–739. http://dx.doi.org/10.3758/s13423-013-0393-4

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491. http://dx.doi.org/10.1037/0033-295X.109.3.472

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 50,* 408–412. http://dx.doi.org/10.1037/1196-1961.50.4.408

Pollmann, S., Weidner, R., Müller, H. J., & von Cramon, D. Y. (2006). Neural correlates of visual dimension weighting. *Visual Cognition, 14,* 877–897. http://dx.doi.org/10.1080/13506280500196142

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences, 1,* 40–46. http://dx.doi.org/10.1016/j.cobeha.2014.08.004

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision, 12,* 21. http://dx.doi.org/10.1167/12.13.21

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13,* 766–785. http://dx.doi.org/10.1162/08989290152541430

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1075–1096. http://dx.doi.org/10.1037/a0031172

Rerko, L., Oberauer, K., & Lin, H.-Y. (2014). Spatial transposition gradients in visual working memory. *The Quarterly Journal of Experimental Psychology, 67,* 3–15. http://dx.doi.org/10.1080/17470218.2013.789543

Ricker, T. J., Vergauwe, E., & Cowan, N. (2014). Decay theory of immediate memory: From Brown (1958) to today (2016). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 69,* 1969–1995.

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience, 32,* 12990–12998. http://dx.doi.org/10.1523/JNEUROSCI.1892-12.2012

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. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105,* 5975–5979. http://dx.doi.org/10.1073/pnas.0711295105

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912. http://dx.doi.org/10.1037/a0013396

Shiffrin, R. M., & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning & Verbal Behavior, 17,* 189–218. http://dx.doi.org/10.1016/S0022-5371(78)90146-9

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. E. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72,* 116–141. http://dx.doi.org/10.1016/j.jml.2014.01.004

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119,* 807–830. http://dx.doi.org/10.1037/a0029856

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review, 123,* 425–451. http://dx.doi.org/10.1037/rev0000023

Souza, A. S., & Oberauer, K. (2015). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin & Review, 22,* 156–162. http://dx.doi.org/10.3758/s13423-014-0652-z

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. http://dx.doi.org/10.3758/s13414-016-1108-5

Souza, A. S., Rerko, L., Lin, H.-Y., & Oberauer, K. (2014). Focused attention improves working memory: Implications for flexible-resource and discrete-capacity models. *Attention, Perception, & Psychophysics, 76,* 2080–2102. http://dx.doi.org/10.3758/s13414-014-0687-2

Souza, A. S., Rerko, L., & Oberauer, K. (2014). Unloading and reloading working memory: Attending to one item frees capacity. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1237–1256. http://dx.doi.org/10.1037/a0036331

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences, 1339,* 20–31. http://dx.doi.org/10.1111/nyas.12603

Souza, A. S., Rerko, L., & Oberauer, K. (2016). Getting more from visual working memory: Retro-cues enhance retrieval and protect from visual interference. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 890–910. http://dx.doi.org/10.1037/xhp0000192

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

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics, 76,* 2071–2079. http://dx.doi.org/10.3758/s13414-014-0690-7

Surprenant, A. M., & Neath, I. (2009). *Principles of memory*. New York, NY: Taylor & Francis.

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics, 76,* 2136–2157. http://dx.doi.org/10.3758/s13414-014-0633-3

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9,* 333–348. http://dx.doi.org/10.1080/09658210042000049

Thomson, D. R., & Milliken, B. (2013). Revisiting the time course of inter-trial feature priming in singleton search. *Psychological Research, 77,* 637–650. http://dx.doi.org/10.1007/s00426-012-0455-7

Töllner, T., Müller, H. J., & Zehetleitner, M. (2012). Top-down dimensional weight set determines the capture of visual attention: Evidence from the PCN component. *Cerebral Cortex, 22,* 1554–1563. http://dx.doi.org/10.1093/cercor/bhr231

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1–26. http://dx.doi.org/10.1016/j.cogpsych.2014.01.003

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149. http://dx.doi.org/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, 76,* 2117–2135. http://dx.doi.org/10.3758/s13414-013-0618-7

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. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 109,* 8780–8785. http://dx.doi.org/10.1073/pnas.1117465109

Vergauwe, E., & Cowan, N. (2014). A common short-term memory retrieval rate may describe many cognitive procedures. *Frontiers in Human Neuroscience, 8,* 126. http://dx.doi.org/10.3389/fnhum.2014.00126

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438,* 500–503. http://dx.doi.org/10.1038/nature04171

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. http://dx.doi.org/10.1037/0096-1523.27.1.92

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1436–1451. http://dx.doi.org/10.1037/0096-1523.32.6.1436

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* 1120–1135. http://dx.doi.org/10.1167/4.12.11

Williams, M., Hong, S. W., Kang, M.-S., Carlisle, N. B., & Woodman, G. F. (2013). The benefit of forgetting. *Psychonomic Bulletin & Review, 20,* 348–355. http://dx.doi.org/10.3758/s13423-012-0354-3

Woodman, G. F., & Vogel, E. K. (2008). Selective storage and maintenance of an object's features in visual working memory. *Psychonomic Bulletin & Review, 15,* 223–229. http://dx.doi.org/10.3758/PBR.15.1.223

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440,* 91–95. http://dx.doi.org/10.1038/nature04262

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

Zokaei, N., Manohar, S., Husain, M., & Feredoes, E. (2014). Causal evidence for a privileged working memory state in early visual cortex. *The Journal of Neuroscience, 34,* 158–162. http://dx.doi.org/10.1523/JNEUROSCI.2899-13.2014

Appendix A

A Mechanistic Explanation of Context-Independent Memory and Background Noise

Here we show that the contributions of background noise, A_b, and the contribution of context-independent feature memory, A_a, to the overall activation distribution A can be derived from one simple assumption: Encoding of each object entails uniform noise in context space with density $Y/2\pi$, and uniform noise in feature space with density $X/2\pi$. Moreover, we show how focusing attention on one location in context space after encoding can reduce the contribution of A_a and A_b.

Encoding of each object's feature means to bind that feature to its context through a representation in binding space. Formally, binding representations are distributions in binding space computed as the outer product of the activation distribution in feature space representing the bound feature with the activation distribution in context space representing the bound context. Because both the feature and the context representations include a component of uniform background noise, the binding representation includes a bivariate uniform noise distribution in binding space, which is the outer product of the uniform background noises in feature space and context space. The density of background noise in binding space, N_b, increases linearly with the number of feature-context bindings, n (i.e., with the set size of the array):

$$N_b = \left(\frac{Y}{2\pi} \cdot \frac{X}{2\pi}\right)n = \frac{YX}{4\pi^2}n \tag{A1.1}$$

When memory for the target is cued by a probe in location L_θ in context space, we approximate the probe representation as a combination of a point-precise peak of activation of magnitude 1 at L_θ, (reflecting the high-precision perceptual representation of the probe) together with uniform background noise with density $Y/2\pi$, referred to as B_y. This cue generates an activation distribution in feature space reflecting the sum of the (low-precision) feature-context bindings and the uniform background noise in binding space. The activation of any value x in feature space arising from the point-precise context activation at L_θ is:

$$A(x \mid L_\theta) = N_b + \sum_{i=1}^{n} \exp\left[-sD(L_i, L_\theta)\right] \cdot vMises(x, x_i, \kappa) \tag{A1.2}$$

The background-noise component B_y acts as additional cue of uniform strength at all context locations L. The activation at any value x in feature space generated by the background noise in any location L of context space is:

$$A(x \mid L) = \frac{Y}{2\pi} \cdot N_b + \frac{Y}{2\pi}\sum_{i=1}^{n} \exp\left[-sD(L_i, L)\right] \cdot vMises(x, x_i, \kappa) \tag{A1.3}$$

Integrating over all locations L that are cued by uniform background noise B_y:

$$\begin{aligned} A(x) &= \frac{Y}{2\pi} \cdot \int N_b + \sum_{i=1}^{n} \exp\left[-sD(L_i, L)\right] \cdot vMises(x, x_i, \kappa) dL \\ &= \frac{Y}{2\pi} \cdot \int N_b dL + \frac{Y}{2\pi}\sum_{i=1}^{n} \int \exp\left[-sD(L_i, L)\right] dL \cdot vMises(x, x_i, \kappa) \end{aligned} \tag{A1.4}$$

(*Appendices continue*)

The total activation distribution in feature space is the sum of the activation generated from location cue L_0 (A1.2) and the activation generated from background noise at test, B_y (A1.4). The contribution of uniform background noise to the activation distribution is the sum of the N_b components in Equations A1.2 and A1.4. We can therefore write the contribution of background noise to $A(x)$ as

$$\begin{aligned} bA_b(x) &= N_b + \frac{Y}{2\pi} \cdot \int N_b \, dL \\ &= \frac{YX}{4\pi^2} n + \frac{Y}{2\pi} \cdot \frac{YX}{2\pi} n \\ &= \frac{YX(1+Y)}{2\pi} \cdot \frac{n}{2\pi} \end{aligned} \tag{A1.5}$$

It is clear from this development that the contribution of background noise to the activation distribution increases linearly with set size n. Because in the IM, the background-noise component to activation is

$$bA_b(x) = b \cdot \frac{n}{2\pi},$$

We can express b in terms of Y and X:

$$b = \frac{YX(1+Y)}{2\pi} \tag{A1.6}$$

The A_a component of the total activation is the unweighted sum of the von-Mises distributions over feature space representing all features in the array. These distributions are recreated by the background noise in context space, B_y, that accompanies the retrieval cue, and that effectively cues all features bound to any location in context space with equal strength, including the locations L_i of each array object i. This contribution to $A(x)$ is expressed in the second part of the sum in Equation A1.4:

$$aA_a(x \mid L_i) = \frac{Y}{2\pi} \sum_{i=1}^{n} \int \exp[-sD(L_i, L)] dL \cdot vMises(x, x_i, \kappa). \tag{A1.7}$$

We can therefore write $a(i)$ for each item i as:

$$a(i) = \frac{Y}{2\pi} \int \exp[-sD(L_i, L)] dL \tag{A1.8}$$

In a circular context space, as used in the present experiments, the integral of the exponential function over all possible locations L is independent of the location of the item, implying that parameter a is identical for all items:

$$a = \frac{Y}{2\pi} \int \exp[-sD(0, L)] dL \tag{A1.9}$$

Focusing attention on L_0 in response to a retro-cue can reduce the contribution of A_a and A_b by reducing the relative effect of background noise N_b in three ways: First, focusing on L_0 could reduce the magnitude of background noise in the context space B_y at all other locations but L_0 at the time of retrieval. Second, focusing on L_0 could lead to reduction of all bindings to locations other than L_0 by suppressing the binding space in all areas except for the line crossing at L_0 (i.e., the broken line in Figure 1). Third, focusing on L_0 could increase all bindings along the line crossing at L_0. All three mechanisms are mathematically equivalent in that they change the relative strength of bindings on the L_0 line relative to those in areas off the L_0 line, either by an increase of $A(x \mid L_0)$ as given in A1.2, or by a decrease of $A(x)$ as given in A1.4. Because calculation of response probabilities involves normalization of A, both effects are equivalent. A proportional decrease of $A(x)$ by factor r implies a partial reduction of A_b, and a complete reduction of A_a, by the same factor:

$$\begin{aligned} bA_b(x) &= N_b + r \cdot \frac{Y}{2\pi} \cdot \int N_b \, dL \\ &= \frac{YX}{4\pi^2} n + r \cdot \frac{Y}{2\pi} \cdot \frac{YX}{2\pi} n \\ &= \frac{YX(1+rY)}{2\pi} \cdot \frac{n}{2\pi} \end{aligned} \tag{A1.10}$$

$$aA_a(x \mid L_t) = r \frac{Y}{2\pi} \sum_{i=1}^{n} \int \exp[-sD(L_i, L)] dL \cdot vMises(x, x_i, \kappa). \tag{A1.11}$$

The reduction of A_b is only partial because the first component of the sum in Equation A1.5 reflects the contribution of background noise on the L_0 line, which cannot be reduced by focusing on L_0.

Appendix B

Experiments

Experiment 1

Participants. Twenty students were recruited from University of Zurich. One of the participants did not finish the second session and was excluded from analysis. Participants were remunerated with course credits or 30 Swiss Francs.

Stimuli. Items in the memory arrays were circular color patches with a diameter of 1.25° visual angle at a viewing distance of 50 cm. Each item's color was selected at random from 360 colors created from the CIE $L^*a^*b^*$ color model with luminance set to 70. The 360 colors were equidistant on a circle centered at (a = 20, b = 38) with radius = 60

Design. Experiment 1 consisted of two identical sessions carried out on different days. In each session, 400 trials were tested. The set-size of stimulus varied from 1 to 8 in random order, with 50 trials per set-size in each session.

(*Appendices continue*)

Procedure. At the beginning of each trial, a central fixation cross was displayed. After 0.5 s, one to eight color patches were displayed at locations selected at random from 13 locations. The 13 locations were equidistant on an invisible circle (diameter 5.1°) around the screen center. The color patches were erased after 0.1 s. We chose a short display duration (100 ms) because it is frequently used in the literature (primarily to minimize eye movements during encoding). A pilot experiment (not reported here) varying display duration (100 vs. 1000 ms) showed somewhat better performance at the longer display duration but no qualitative differences in the data patterns, or in the fit of the models under consideration.

After a 1-s retention interval, empty frames were displayed at the locations of the color patches, and the probed frame was thicker than others. Simultaneously with the frames a color wheel (diameter 17.2°) was displayed that consisted of the 360 possible colors used in the experiments. The wheel was rotated randomly from trial to trial. Participants were asked to select the color of the probed location by left-clicking one of the colors on the color wheel. A blank screen was displayed after the response, and then the next trial began after a 1 s inter-trial-interval. In all four experiments, participants were instructed to repeatedly say "bi bi bi" aloud during each trial to discourage verbal encoding of the stimuli.

Experiment 2

Participants. Twenty-one students were recruited from University of Zurich. They received course credit or 45 Swiss Francs for completing three 1-hr sessions.

Stimuli. The same as Experiment 1, except that each item's color was selected at random from 360 colors created from the HSL color model with saturation set to 1 and lightness set to .5. Hue was equally distributed between 1° to 360°.

Design. In two of the three sessions, valid-cue and invalid-cue conditions were mixed in random order; with 67% valid-cue trials and 33% invalid-cue trials (except for set size 1, where the cue was 100% valid). The third session consisted of trials from the neutral condition. The order of sessions was counterbalanced between participants. Each session consisted of 250 trials. The set-size varied over five steps: 1, 2, 4, 6, and 8 colors; trials with different set sizes were presented in random order.

Procedure. The procedure in Experiment 2 was similar to Experiment 1. Stimuli were presented in a random subset of 13 equidistant locations on a virtual circle of 7.7° diameter; the display duration was 0.1 second. A cue appeared at the center of the screen for 0.3 second before the stimuli were displayed. For the neutral condition, the cue was an asterisk, which does not point to any specific location. For valid-cue and invalid-cue conditions, the cue was an arrow that pointed to one of the locations in which a color patch was to appear. Participants were instructed that the cued location was more likely to be probed for response than not-cued locations.

Experiment 3

Participants. We enrolled 21 students from University of Zurich. They received course credit or 45 Swiss Francs for completing three sessions.

Stimuli. Stimuli were the same as for Experiment 2.

Design and procedure. Experiment 3 was exactly like Experiment 2, with two differences: First, we slightly increased the number of trials per session to 300. Second, the cue appeared after the 1 s retention interval. The cue was displayed for 0.3 s, after which it was immediately replaced by the test display with the frames in the locations of color patches, and the color wheel. In the neutral condition, no cue was displayed after the retention interval; rather, the test display was presented immediately after the 1 s retention interval. In this way, the retention interval of the cue conditions before the cue was the same as in the neutral condition, but the overall retention interval between encoding and test was 0.3 s longer. Therefore, any benefit from a valid cue compared to the neutral condition cannot be attributed to a shorter retention interval.

Experiment 4

Participants. Twenty-one students were recruited from University of Zurich. Participants were remunerated with course credits or 15 Swiss Francs.

Stimuli. Items were colored disks with a rectangular gap, the orientation of which was randomly selected from 1° to 360° (Anderson et al., 2011). The diameter of the disk was 1.9° visual angle at a viewing distance of 50 cm. The color of each disk was selected from 9 colors from equidistant points on the color circle that was used as the response space in Experiment 1. The color set was fixed for each participant but randomly shifted along the color circle between participants. Each memory array consisted of six visual objects displayed at locations selected at random from 13 equidistant locations on an invisible circle (diameter 5.1°) around the screen center.

Design. Experiment 4 consisted of 300 trials carried out in one session. The three types of probes (location, color, and double probes) occurred in random order, with 100 trials per condition.

Procedure. At the beginning of each trial, a central fixation cross was displayed. After 0.5 s, the memory array was displayed for 1 s, followed by a blank screen of 1 s. One of the six objects was probed for recall after the retention interval. In the color probe condition, a filled disk appeared at the center of the screen with the color of the probed item. In the location probe condition, a black disk was presented at the location of the probed item. In the double-probe condition, a colored disk was presented at the location of the probed item. The color and the location were always from the same object in the memory array. Participants were asked to reproduce the orientation of the probed disk. Participants could rotate the orientation of the probed disk's gap by moving the mouse cursor; the orientation always pointing toward the mouse cursor. Participants committed their response by clicking the left mouse button. A blank screen was displayed after the response, and then the next trial began after a 1 s inter-trial-interval.

(*Appendices continue*)

Appendix C

Model Recovery Simulations

We carried out a simulation study to investigate model recovery. We simulated data from the three explanatory models in our primary model competition: The Slot-Averaging (SA) model, the Variable-Precision (VP) model, and the Interference model (IM). In addition, we included the original Resource model (R) of Bays and Husain (2008) because it differs from the VP only in that the VP is more flexible. A fit indicator that penalizes flexibility adequately in the context of the models under consideration should distinguish well between the two resource model versions differing in flexibility. We simulated data from each model, and fit all four models to each simulated data set. The aim of this study was to gauge the chance of each model to win the competition against the other three models, given that it was the true model. In both simulation studies we calculated AIC and BIC as fit indicators and determined the winning model according to each fit indicator.

Each simulation generated data from 20 subjects for an experiment like Experiment 1, varying set-size from 1 to 8, with 80 trials per condition. We used one of four explanatory models (SA, R, VP, or IM) to generate the data. The parameter values of each subject were drawn from a Gamma distribution with mean and standard deviation informed by model fits to a pilot experiment (not reported here), which was similar to Experiment 1. Gamma distributions were chosen because all model parameters had a lower bound of zero, but no upper bound. Each simulated data set was fit by each of the four models (using a single run with the Simplex algorithm), and we determined which model won the competition according to AIC, and according to BIC. We planned to run 500 simulations for each generating model, but because of a technical failure the simulation process was interrupted prematurely, so that we obtained between 420 and 463 simulations per generating model. Table C1 shows for each generating model the proportion of simulations in which each of the four models won the competition. The main diagonal of the table shows the proportion of simulations in which the true model won the competition.

Table C1
Model Recovery Study

Generating model	SA	Resource	VP	IM
AIC				
SA	1	0	0	0
Resource	0	1	0	0
VP	0	0	1	0
IM	0	0	.55	.45
BIC				
SA	1	0	0	0
Resource	0	1	0	0
VP	1	0	0	0
IM	1	0	0	0

Note. Table entries are proportions of simulations in which the model (columns) won the competition when fit to the data of each generating model (rows) according to AIC (top) or BIC (bottom).

There are two conclusions to be drawn from the results. First, AIC provides much better model recovery than BIC; the latter overpenalizes complex models, so that simpler models (in particular the SA model) win the competition when a more complex model (VP or IM) generated the data. Second, even with AIC the VP model wins the competition in about half the cases in which the IM generated the data, whereas the converse never occurred. This means that the VP has a degree of flexibility not accounted for by the penalty term in the AIC, which enables it to fit data generated by the IM better than the IM itself. The source of this flexibility is certainly not an excess of free parameters in the VP, which has only three free parameters. Rather, its flexibility must arise from its functional form (for discussion see Pitt, Myung, & Zhang, 2002).

Received April 10, 2015
Revision received August 24, 2016
Accepted August 25, 2016 ■