A two-level hierarchical framework of visual short-term memory

Tal Yatziv

Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel ✉

Yoav Kessler

Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel 🏠 ✉

Over the last couple of decades, a vast amount of research has been dedicated to understanding the nature and the architecture of visual short-term memory (VSTM), the mechanism by which currently relevant visual information is maintained. According to discrete-capacity models, VSTM is constrained by a limited number of discrete representations held simultaneously. In contrast, shared-resource models regard VSTM as limited in resources, which can be distributed flexibly between varying numbers of representations; and a new interference model posits that capacity is limited by interference among items. In this article, we begin by reviewing benchmark findings regarding the debate over VSTM limitations, focusing on whether VSTM storage is all-or-none and on whether object complexity affects capacity. After that, we put forward a hybrid framework of VSTM architecture, arguing that this system is composed of a two-level hierarchy of memory stores, each containing a different set of representations: (1) perceptual memory, a resourcelike level containing analog automatically formed representations of visual stimuli in varying degrees of activation, and (2) visual working memory, in which a subset of three to four items from perceptual memory are bound to conceptual representations and to their locations, thus conveying discrete (digital/symbolic) information which appears quantized. While perceptual memory has a large capacity and is relatively nonselective, visual working memory is restricted in the number of items that can be maintained simultaneously, and its content is regulated by a gating mechanism.

Introduction

The ability to maintain and manipulate representations of currently relevant information is supported by a mental mechanism referred to as *working memory* (WM; e.g., Baddeley, 1996; Cowan, 1988). In the past two decades, much attention has been given to the operation of WM in the visual modality—that is, the maintenance and manipulation of visual information (e.g., Luck & Vogel, 2013). A central debate regards the nature of the constraints on maintenance of representations in visual WM. According to one side of the debate—discrete-capacity models of visual WM (e.g., Rouder et al., 2008; Vogel, Woodman, & Luck, 2001)—capacity is limited in the number of items that can be maintained simultaneously, which is typically three or four items. Under this account, the precision of memory representations is *quantized*, and therefore this architecture is commonly conceptualized as slots. The second stance in the debate is that precision of maintained representations in visual WM is *continuous* rather than discrete. According to one class of such models—shared-resource models (e.g., Bays & Husain, 2008; Wilken & Ma, 2004)—visual WM is limited in resources, which can be allocated and distributed flexibly among a varying number of representations. For example, a small number of representations can be held with high precision, or conversely, a large number of representations can be maintained with low precision. Recently, an interference account of WM has been applied to visual memoranda (Oberauer & Lin, 2017), arguing that visual WM is constrained by interference between maintained representations. According to this model, WM capacity is continuous (and not quantized), but restrictions are due to increasing cross talk between items and their bindings to contexts rather than a result of resources allocation.

In this article, we argue for a hybrid framework that can reconcile the two stances. We suggest that performance in so-called visual WM tasks relies on two separate sets of representations: perceptual memory

(PM), which holds a variable number of representations whose strength varies on a continuum, and visual WM itself, which—at least in many cases—appears quantized. We propose that memory over the short term does not necessarily depend on visual WM, but can rather rely on lower and less selective structures. Accordingly, it is important to distinguish between the two in order to correctly attribute the empirical findings in visual WM tasks to the cognitive construct that gives rise to them. We argue that such a distinction between PM and visual WM provides a simple yet powerful solution to the debate over whether there is an item limit to visual WM.

Our theoretical framework distinguishing between WM and PM is not specific to the visual domain but is rather a general distinction between automatically activated memory traces outside WM and selected representations inside WM (see Kessler, 2018; Rac-Lubashevsky & Kessler, 2016a, 2016b). The goal of the present article is to apply these ideas to the visual domain and to suggest a synthesis of the literature on this debate together with three other lines of research in the study of WM—the embedded-components model of WM (Cowan, 1999; Oberauer, 2002, 2009), the notion of fragile visual short-term memory (Sligte, Scholte, & Lamme, 2008), and gating models of selection to WM (Braver & Cohen, 2000; Hazy, Frank, & O'Reilly, 2006). We will begin by reviewing benchmark findings regarding capacity limitations of visual WM. Because a comprehensive review of this vast literature is beyond the scope of this article, we will focus only on the major findings and arguments supporting each class of model. We will especially highlight findings showing evidence for both quantized and continuous capacity allocation within the same task, which are mostly compatible with our framework. Then we will introduce our two-level framework and review evidence linking the two levels to discrete and continuous limitations.

A note on terminology is required to avoid confusion. Our notion of two levels of representation implies that not every manifestation of memory over the short term relies on working memory; memory also takes place outside it. However, the term *visual working memory* is frequently used in the literature to refer to any memory phenomenon that takes place in the visual domain across short time periods. Thus, we essentially argue that not all aspects of visual WM tasks actually rely on working memory. This, of course, might lead to an undesired confusion between our framework and previous work, and between the construct of working memory (denoting a selective and limited set of representations, as used, for example, in gating models; e.g., Braver & Cohen, 2000; Hazy et al., 2006) and the general ability to remember visual items over the short term (as measured by visual WM tasks). To avoid

confusion, we will use the term *visual short-term memory* (VSTM) to refer to the general ability to remember visual information over several seconds, which is measured by VSTM tasks. That is, VSTM denotes the system as a whole, including both PM and visual WM.

Is VSTM capacity quantized? Overview of current stances in the debate

Unlike our framework, performance in VSTM tasks under slot, shared-resource, and interference models is attributed to a single memory store, typically termed WM. To be compatible with our definitions, we will refer to it here as VSTM. According to discrete-capacity models, VSTM encoding can be conceptualized as an all-or-none, high-threshold process. VSTM is limited in the *number of representations* that can be maintained simultaneously, denoted K (Luck & Vogel, 2013). That is, VSTM has a capacity of K discrete quanta, conceptualized as slots that can be filled with representations of K visual items. In trying to maintain visual arrays exceeding K items, only K will be stored and retained; the rest will not be represented in short-term visual memory. In contrast, shared-resource models argue that VSTM is limited not in the number of memory items but in the amount of resources that can be distributed among them. These models suggest that the amount of resources devoted to an item is proportional to the precision with which the item is retained. Consequently, the capacity limitation is manifested in the *precision* of each memory representation (e.g., Bays & Husain, 2008; Ma, Husain, & Bays, 2014). For small set sizes, memory representations would be quite precise, whereas the representation of items within large memory arrays would be less accurate. Building on signal-detection-theory accounts, this property of VSTM is assumed to be a result of random noise throughout sensory, maintenance, and retrieval processing channels, all leading to an increase in noise with set size (Ma et al., 2014; Wilken & Ma, 2004). Very recently, Oberauer and Lin (2017) have introduced a new stance with regard to the debate over the architecture of VSTM, arguing that VSTM is limited by interference. According to this account, all to-be-remembered items are represented in VSTM, and performance at large set sizes is degraded because items interfere with one another. This interference leads to lowered precision and weak connections between items and their locations (item-to-context binding, a notion on which we will elaborate later). Furthermore, one privileged item among VSTM representations, held in
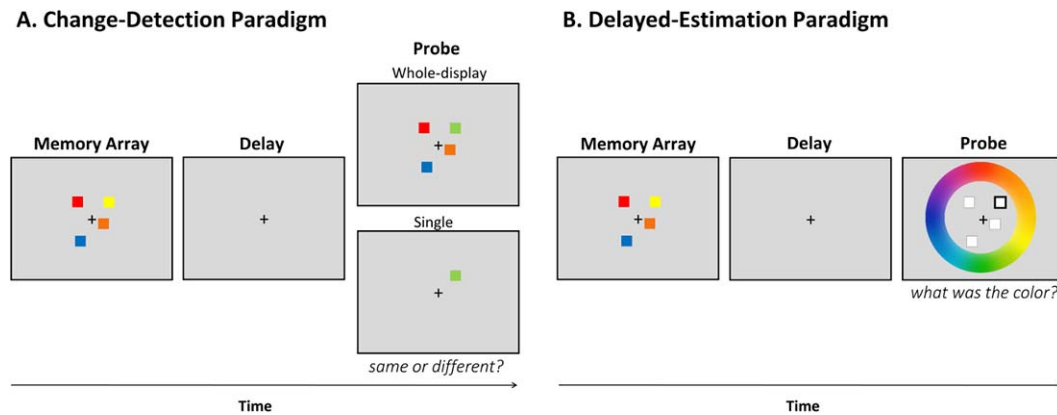
Figure 1. Examples of trials in (A) the change-detection paradigm and (B) the delayed-estimation paradigm.

the focus of attention, is maintained with higher precision compared to other items. Although different from shared-resource models, the interference model posits that VSTM precision is set on a continuum. In this section, we briefly outline benchmark findings regarding the question of whether VSTM storage is all-or-none. For more extensive reviews, see Fukuda, Awh, and Vogel (2010); Luck and Vogel (2013); and Ma et al. (2014).

Before we continue, it is important to consider the two common experimental paradigms utilized in the study of VSTM capacity limitations. The first is the change-detection task (e.g., Luck & Vogel, 1997; see Figure 1A). In each trial, participants are presented with a brief display containing several visual items and are asked to remember these items. The number of to-be-remembered items is commonly referred to as *set size*. After a retention interval, participants are presented with a test display and asked to indicate whether it is the same or different compared to the memory array. The test display could require either comparison of all items (whole-display probe) or a partial report regarding whether one item has changed (single probe; Vogel et al., 2001).

The second paradigm is the delayed-estimation task (Wilken & Ma, 2004; see Figure 1B), also commonly referred to as the continuous-report task. The difference between delayed estimation and change detection lies in the test probe: In the delayed-estimation task, participants are asked to adjust the feature of a cued item (e.g., choose its color on a color wheel, adjust the orientation of a bar) according to their memory of that item. This design enables estimation of recall probability (the probability that an item was represented in VSTM) and precision of memory representations (provided these items are represented), as well as swap errors in which items' features or locations are intermixed (but see Schurgin, Wixted, & Brady, 2018, for recent criticism regarding the validity of the linear

psychological interpretation of physical distances along the response wheel).

Precision

One way to approach the question of whether VSTM capacity is quantized is to test the effect of set size on precision, using the delayed-estimation paradigm. Zhang and Luck (2008) found that while recall probability decreased as set size increased, precision decreased between set sizes 1 and 3 but then reached a plateau for set sizes exceeding K. Their model suggests that there is a fixed upper limit of slots (K), and that when fewer than K items are to be remembered, one item can be represented in several slots (i.e., receive several capacity quanta, a model they named "slots + averaging"; see also Zhang & Luck, 2011). On the other hand, Bays and Husain (2008) found a decrease in precision as set size increased, with a large decline in precision between set sizes of 1 and 2. In addition, in that experiment there was no major decline in precision following the K limit of four items. Thus, Bays and Husain's results are more in line with the notion of a shared resource. According to Bays, Catalao, and Husain (2009), the discrepancy between Zhang and Luck's results and those of Bays and Husain stems from a misinterpretation of swap errors as random noise in Zhang and Luck's analysis. Conversely, according to Cowan and Rouder (2009), Zhang and Luck's model has a similar (and even slightly better) fit to Bays and Husain's data compared to the fit of a resource model. Recently, Adam, Vogel, and Awh (2017) asked participants to reproduce all memory items, and found that precision decreased systematically within a trial (i.e., the most precise representations were reported first). When set size was supracapacity (6), participants' responses were best fit by a model assuming that three responses were based on guesses. The researchers interpreted this result as indicating that only three items can be maintained in VSTM,

supporting discrete-capacity models (but for a criticism of this conclusion, see Bays, 2018). On the other hand, Bays (2018) examined whether the set size at which precision and recall probability plateau coincide. Contrary to predictions of discrete-capacity models, these two capacity estimations were not correlated across several studies, raising a new challenge for this class of models. At any rate, this discussion should be conducted while taking into account that the usefulness of plateau statistics in model comparison has been questioned by van den Berg and Ma (2014), who argued that plateau statistics are highly unreliable.

Van den Berg, Awh, and Ma (2014) compared the fit of several models to performance on the delayed-estimation task across different studies, and found that shared-resource models in which precision is variable across items both within a given trial and across trials in a given set size (e.g., Fougnie, Suchow, & Alvarez, 2012; Sims, Jacobs, & Knill, 2012; van den Berg, Shin, Chou, George, & Ma, 2012) fit the accuracy data better than the slots + averaging model (Zhang & Luck, 2008) and the nonvariable version of shared-resource models (e.g., Wilken & Ma, 2004). Furthermore, Oberauer and Lin (2017) compared the fit of their interference model with several variations of discrete-state models and shared-resource models, including the variable-precision version of resource models. They found that their model had the best fit to the data, supporting the notion that VSTM is limited by interference and contains a single-item focus of attention. However, it should be noted that a model in which the interference model was combined with an additional discrete-capacity limit had a good fit as well, raising the possibility that VSTM is limited by a combination of interference and a single-item focus of attention, together with an additional discrete item limit.

Modeling receiver operating characteristics (ROC) curves

Another way to approach the question is to examine the shape of ROC curves of change-detection perfor-mance. Linear ROC curves (for nonstandardized hit and false-alarm rates) support high-threshold models, while curvilinear curves support signal-detection mod-els (e.g., Egan, 1975). Studies in which this methodol-ogy was used have yielded conflicting results. For example, Wilken and Ma (2004) used a variant of the change-detection task in which participants were asked to rate their confidence in the responses they made. The researchers found nonlinear ROC curves and deter-mined that detection-theory models with the assump-tion of an increase in noise with set size had a better fit compared to a high-threshold model, supporting the resource stance. On the other hand, Rouder et al. (2008; for a replication, see also Donkin, Tran, & Nosofsky, 2014, experiments 1, 2, and 4) manipulated

change probabilities in a change-detection task and found an opposite pattern of results, supporting slots models: ROC curves were linear for each set size, change probabilities affected guessing rates but not capacity estimates, and a high-threshold model had a better fit compared to a variable-capacity model and to a signal-detection model assuming variability in the strength of the representations.

According to Donkin, Kary, Tahir, and Taylor (2016), participants can use either slotlike or resource-like encoding strategies, depending on whether the experimental conditions enable them to distribute attention to all items in the memory array. For example, they demonstrated that a slot model fits ROC data when the set size varies (and is therefore unpredictable). However, when the set size is fixed, participants use resourcelike encoding on some of the trials. The researchers concluded that flexible resources can be allocated as slots under certain conditions, and suggested that this may explain why this method provides evidence for both stances.

Reaction time (RT)

Recently, Donkin, Nosofsky, and colleagues have modeled RTs, in addition to accuracy rates, in two versions of the change-detection task: a common single-probe version and a sequential-presentation version of the task in which the to-be-remembered items were presented one after the other (similar to rapid serial visual presentation; Donkin, Nosofsky, Gold, & Shiffrin, 2013; Nosofsky & Donkin, 2016). They have compared the fit of discrete-slots, shared-resource, and hybrid models to RTs, using modeling of evidence-accumulation processes. They suggest that while shared-resource models predict that RTs would be characterized by a single distribution, discrete-slots models predict that RTs would be characterized by a mixture of two distributions—one of a memory-based evidence-accumulation process (items stored in VSTM) and another of a guessing-based distribution (items not encoded to VSTM). In general, their results favor slots models, indicating that RTs are based on two processes: evidence accumulation and guessing. How-ever, in small set sizes or for short lags of presentation (in sequential presentation), a hybrid model of slots and shared-resource combination had a better fit to RTs.

Influence of task requirements on performance

One account of the mixed results is that task demands can modulate the number of maintained items as well as their precision (e.g., Dempere-Marco, Melcher, & Deco, 2012; Fougnie, Cormiea, Kanabar, & Alvarez, 2016; Knops, Piazza, Sengupta, Eger, & Melcher, 2014).

Melcher, Piazza, and colleagues have examined commonalities between VSTM capacity and subitizing range in enumeration, two capacity-limited phenomena that appear to share a limit of three to four items (Knops et al., 2014; Melcher & Piazza, 2011). They found that a saliency manipulation affected VSTM capacity estimates and subitizing ranges in a similar manner (Melcher & Piazza, 2011), and that performance on both tasks was associated with activity in the same brain region, the posterior parietal cortex (Knops et al., 2014). Subitizing range was consistently larger than VSTM capacity estimates, and this discrepancy was evident in posterior parietal activity as well (Knops et al., 2014). A computational attractor-network model based on their behavioral VSTM results has led them to identify two capacity limitations: shared resource which can be distributed among items based on saliency and task demands, and an upper limit on the number of retained items (Dempere-Marco et al., 2012). Knops et al. argued that these two limitations account for the mixed results regarding slots and resources, and that actual capacity in a given task changes in accordance with task requirements (up to the upper capacity limit), such that more items are represented when required precision is low (e.g., enumeration tasks) compared to high (e.g., change detection).

Fougnie et al. (2016) came to a similar conclusion, proposing that strategic considerations can modulate the mode in which representations are encoded and maintained (but see Zhang & Luck, 2011) by manipulating the number of probed items in a delayed-estimation task. They asked participants to perform two variations of the task—one in which they had to report the color of one item (standard task) and another in which they had to report the colors of all the items in the display ("get-them-all" task). Results indicated that when participants were informed ahead of time that the probe would require production of all items (either by arranging single-task blocks or by using a task cue during the retention interval in mixed-task blocks), they showed lower guessing rates but also worse precision compared to the standard task. Performance was not modulated by task when participants knew how many items to reproduce only when the probe array was presented. This indicates that VSTM is flexible, and that under certain conditions one can control a trade-off between the number of maintained representations and their precision.

Capacity for complex objects

The question of whether storage is all-or-none is inherently related to another open question regarding the architecture of VSTM: Is capacity limited by the number of to-be-remembered objects, regardless of their complexity, or do more complex objects consume more capacity? One definition of complexity refers to the number of features (namely, dimensions) per object (e.g., Luck & Vogel, 1997; Wheeler & Treisman, 2002). Vogel et al. (2001; see also Luck & Vogel, 1997) found that performance on a change-detection task for conjunctions of features (e.g., color and orientation of bars, or colors of bicolored squares) is similar to that for single features. These findings have led them to conclude that VSTM is limited by number of objects, such that VSTM can hold representations of K integrated objects regardless of complexity. However, attempts to replicate this seminal finding have yielded conflicting results (e.g., Olson & Jiang, 2002; Wheeler & Treisman, 2002).

Fougnie, Asplund, and Marois (2010) used the delayed-estimation task to examine the effect of the distribution of features between objects on recall probability and precision. Their results indicated that when the number of simple objects increased, precision and recall probability both decreased. However, when the number of features within an item increased, the probability that the item would be encoded remained the same but the resolution of its representation decreased. This led the researchers to conclude that VSTM is limited in both the number of objects and their precision. Supporting this conclusion are findings indicating that although manipulation of task requirements of either integrated objects or separate features may affect the type of representation used, both separate features and integrated objects are represented (Cowan, Blume, & Saults, 2013; Geigerman, Verhaeghen, & Cerella, 2016; Oberauer & Eichenberger, 2013; Vergauwe & Cowan, 2015).

Alvarez and Cavanagh (2004) have suggested defining object complexity as visual informational load, or visual details contained in an object, measured based on processing speed in a visual-search task. They used a change-detection task in which to-be-remembered items consisted of objects with varying informational load—colors, polygons, shaded cubes, English letters, and Chinese characters. They found that capacity estimates decreased monotonically as informational load increased, indicating that VSTM is limited by the amount of information. Moreover, for objects containing very little information, capacity was estimated at between four and five items, indicating that there is also an upper limit to the number of objects. Thus, they concluded that VSTM capacity can maintain up to four to five objects, depending on the amount of information the objects contain.

The notion that quality and quantity of representations in VSTM are dissociable to a certain degree (e.g., Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Fougnie et al., 2010) has gained further support from findings indicating that retention precision of items can be controlled according to task demands, but

only when set size is below capacity (Machizawa, Goh, & Driver, 2012), and that in a given set size, an increase in the number of features affects precision but not recall probability (Fougnie et al., 2010). Moreover, using functional magnetic resonance imaging (fMRI), Y. Xu and Chun (2006) found that neural activity in the inferior interparietal sulcus (IPS) is sensitive to the number of objects, whereas activity in the superior IPS and lateral occipital complex is sensitive to complexity. They concluded that the superior IPS and lateral occipital complex represent object identities (and some spatial information), while a fixed number of items are automatically encoded and maintained in the inferior IPS based on spatial information (even when irrelevant to the task at hand). Furthermore, they suggested that reduced capacity for complex items does not stem from perceptual load or higher difficulty in retrieval but rather from encoding and maintenance. Despite this support for limits on both the number of items and their precision, it should be noted that recent studies have posed challenges for this view, suggesting that complexity affects only the number of maintained items and that this effect does not arise from comparison errors during the test phase (Brady & Alvarez, 2015; Taylor, Thomson, Sutton, & Donkin, 2017).

Two-level hierarchical framework

Both discrete- and continuous-capacity models attempt to explain VSTM capacity as resulting from a single set of representations (WM) and aim to specify its features and constraints. In this article we propose an alternative framework, in which performance arises from two sets of representations with different characteristics. The novelty of this framework is in its mechanistic account of the seemingly contradictory findings of both discrete and continuous representational states in VSTM, which current theories of VSTM do not account for. Furthermore, this framework relies on several supported predecessors by integrating the VSTM literature with several separate lines of research in the study of WM over the last couple of decades: the embedded-processes model (Cowan, 1999; Oberauer, 2002, 2009), gating models (e.g., the prefrontal cortex and basal ganglia model of WM gating; Hazy et al., 2006), and research on the retro-cue effect (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Souza & Oberauer, 2016).

Two representational states in VSTM

We identify two sets of representations that enable memory storage over several seconds and are both potentially active—in both encoding and retrieval—during performance on VSTM tasks (see Figure 2). The higher level is visual WM. There is a broad agreement that WM (regardless of modality) is highly selective, and that updating of its content is goal directed and subject to control (e.g., Braver & Cohen, 2000; Conway, Cowan, & Bunting, 2001; Engle & Kane, 2004; Vogel, McCollough, & Machizawa, 2005). The control of input selection provides WM with the ability to maintain goal-relevant information that is required for the task at hand. Such a mechanism is necessary in order to make an efficient use of its limited capacity, by enabling goal-relevant information to enter while keeping irrelevant information out. The gating metaphor is typically used to describe input selection: Opening and closing the gate to WM enables the selection of targets from a stream of information that unfolds in time (e.g., Braver & Cohen, 2000; D'Ardenne et al., 2012; Frank, Loughry, & O'Reilly, 2001; Kessler, 2018; Kessler & Oberauer, 2014, 2015; Olivers & Meeter, 2008; Rac-Lubashevsky & Kessler, 2016a, 2016b). O'Reilly and colleagues (e.g., Frank, et al., 2001; Hazy et al., 2006; O'Reilly, 2006; O'Reilly, Braver, & Cohen, 1999) suggested that the gate to WM separates anatomically posterior, perceptual representations (akin to our concept of PM; see later) and prefrontal-based representations, namely WM. In a nutshell, their computational, physiologically based prefrontal cortex and basal ganglia model of WM gating asserts that gating is implemented by alternating between two states: a tonic inhibition of the substantia nigra on the excitatory connections between the prefrontal cortex and the thalamus, which serves as a default closed gate, and a transient phasic disinhibition of this circuit, carried out by the dorsal striatum, which enables transient gate opening. Based on these assumptions of two representational states (nonselective and selective), O'Reilly and colleagues were able to model performance on various WM and executive function tasks (e.g., Chatham et al., 2011; Herd et al., 2014; Kriete, Noelle, Cohen, & O'Reilly, 2013).

This architecture has two implications that are relevant for our discussion. First, a default closed gate enables WM to maintain information that is no longer available perceptually, by shielding it from new perceptual input. Complemented by a self-excitation mechanism within the prefrontal cortex (but for an alternative account, see Postle, 2006), active maintenance within WM enables maintained items to be kept in a highly accessible state that counteracts decay and interference. Second, while gating regulates the input to WM, it does not prevent perceptual input from forming memory traces outside of it. Hence, a lower-level set of representations, hereafter termed *perceptual memory* (PM), coexists in parallel to WM. As the term suggests, these are memories that are created automatically, by
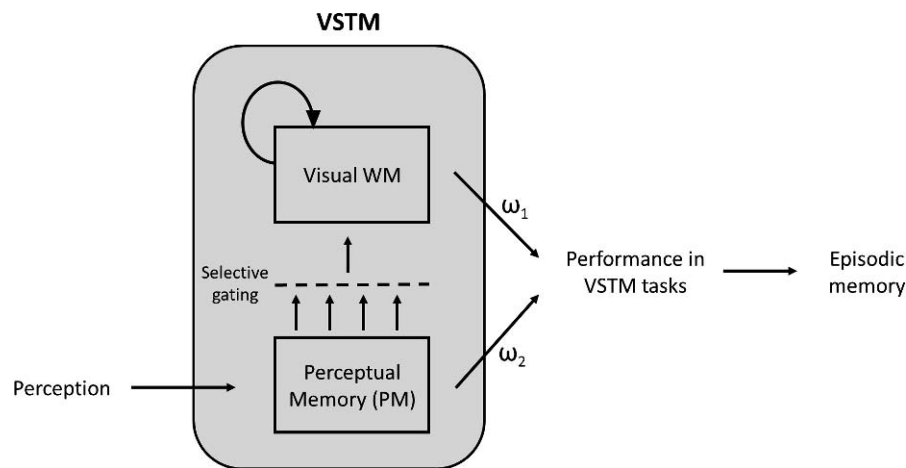
Figure 2. A schematic description of the two-level hierarchical framework of visual short-term memory (VSTM). VSTM is composed of two levels of representation: perceptual memory (PM), storing analog representations of visual stimuli in varying activation levels, and visual WM, storing digital/conceptual representations of a subset of three or four items. PM is the outcome of visual perceptual processing, and the most activated PM items are selected by a gating mechanism to be represented in visual WM, where these perceptual representations are bound to their corresponding conceptual representations in semantic long-term memory, creating structured representations. Performance on VSTM tasks is affected by representations in both levels, but to different degrees, depending on the task requirements (denoted as ω_1 and ω_2, which represent weights). Specifically, performance on the change-detection task is likely to tap mainly the discrete aspect of these structures, while the delayed-estimation task is likely to tap mostly the continuous/analog aspects of these structures.

the virtue of perceiving information, and with no or minimal selective attention or intention. However, the term PM does not imply that these memories are necessarily short-lived, akin to sensory or iconic memory. On the contrary, we suggest that once created, PM representations can lead to creation of long-term episodic memories. The idea of PM can be traced back to Craik and Lockhart's (1972) levels-of-processing framework, regarding (episodic) memory as "a by-product of perceptual analysis" (p. 671). Unlike the controlled nature of WM, PM representations are created in an automatic, obligatory manner, as an emergent property of perception (Graham, Barense, & Lee, 2010).

The notion that PM representations are formed and maintained outside WM implies that the same item can be represented in both PM and WM in parallel, although different aspects might be represented at each level. That is, gating an item into visual WM does not transfer the information, but rather forms a new copy of that item within visual WM in addition to its representation in PM. This implies, among other things, that probing an item, as typically done in VSTM tasks, serves as a retrieval cue for information held both in visual WM and in PM. This is the source-impurity problem in VSTM tasks. However, because visual WM representations are an *attended* subset of PM representations, visual WM items are easily accessible and receive further processing, while unattended PM representations are harder to retrieve (i.e., less accessible) and are more susceptible to interference.

Before we continue to elaborate on each level, we should explicitly discuss the reliance of our framework on Cowan's and Oberauer's conceptualizations of the embedded-components model of WM (Cowan, 1988, 1999; Oberauer, 2002, 2009). In this model, long-term memory (LTM) consists of many representations in varying degrees of activation; a subset of these representations are activated enough that they can be retrieved readily and reliably. The three or four most accessible of these are attended and enter awareness, and thus are also reportable (Cowan, 1999). Activated representations that do not reach awareness are considered *activated LTM* (aLTM), while the few representations that receive complete activation are considered the (broad) *focus of attention* (FoA; but for the idea that the FoA may include one to three items under certain conditions, see Cowan, Saults, & Blume, 2014; Oberauer, 2013). The proposal that representations of short-term memoranda can undergo different states of activation, yielding different representational properties, has received support from neuroscientific studies as well (for reviews see, e.g., LaRocque, Lewis-Peacock, & Postle, 2014; Nee & Jonides, 2011).

Our characterization of the two levels of VSTM draws on the embedded-components model's distinction between the aLTM and the broad FoA. Similar to the aLTM, PM can be considered a form of unattended automatic memory. We regard PM as representations of perceptually based short-term memoranda—that is, traces of visual properties of perceived stimuli (for a similar argument see Hasson, Chen, & Honey, 2015).

This idea echoes the suggestion by Cowan et al. (2014) that aLTM could be viewed as a peripheral, modality-specific form of short-term memory storage, in which information is stored automatically. We embrace and extend this approach, and regard PM as a form of perceptual aLTM.

As mentioned, Oberauer and Lin (2017) recently applied the interference model (e.g., Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012), which relies on the embedded-components framework, to delayed-estimation performance. In terms of the embedded-components framework, in the interference model all items to be remembered are represented in the broad FoA/WM, and performance at large set sizes is degraded because items inside the FoA interfere with one another. Other than that, previous discussions of the embedded-processes framework with regard to VSTM have focused mainly on the distinction between the broad and the narrow FoA (e.g., Cowan, 2011; Rerko & Oberauer, 2013; Souza, Rerko, & Oberauer, 2014). Furthermore, studies regarding the effects of LTM on performance in VSTM tasks have examined the effects of prior knowledge or previously learned associations (e.g., Brady, Konkle, & Alvarez, 2009; Oberauer, Awh, & Sutterer, 2017). In contrast, our notion of PM is inspired by a single-store view of LTM, in which PM representations are formed and retrieved instantaneously, similar to the notion of the aLTM in the embedded-components model.

Visual properties of the two representational states

The debate over the architecture of VSTM is about discrete representational states (quanta) versus continuous representational states. All sides seem to agree that the information encoded in VSTM is analog (or metric), representing shades of colors, orientation in degrees, and so forth. That is, the different models seem to assume that *visual representations themselves* are set on a continuum, drawn from continuous, rather than discrete, distributions. A question to be raised here is whether encoding of visual information is indeed solely analog, without involvement of symbolic codes. This question has only recently begun receiving attention in modeling of performance on delayed-estimation tasks (Bae, Olkkonen, Allred, & Flombaum, 2015; Hardman, Vergauwe, & Ricker, 2017).

According to the visual-perception literature, complete visual processing along the ventral stream is geared toward object recognition. Perception therefore results not only in extraction of a visual representation of the perceived object but also in identification of the object by assigning it to a conceptual, abstract category. For example, in Marr's (1982) seminal analysis of visual-perceptual processing, object recog-

nition is achieved through three levels: (1) a primal sketch level where blobs, bars, and edges are encoded; (2) a 2.5-D sketch level in which a viewpoint-specific representation of an object is obtained; and (3) a 3-D model level in which a viewpoint-invariant representation of an abstract object is achieved. Thus, according to Marr, the final stage of visual processing involves assignment of the perceived object to a category, relating it to conceptual information. This notion is supported by the finding that continuous (metric) and categorical color information are related to different brain regions: While the visual cortex is sensitive to metric (i.e., continuous) differences between hues, the medial prefrontal cortex is sensitive to differences between color categories (Bird, Berens, Horner, & Franklin, 2014). More recent models of visual perception regard visual processing not as occurring in a serial, bottom-up process such as described by Marr (and other classic models) but as an iterative process that involves interactions between bottom-up and top-down processes, manifested in coactivation and interactions between ventral-path visual areas and higher prefrontal and medial-temporal areas, via combined feed-forward and feedback connections (e.g., Bar, 2004; Hochstein & Ahissar, 2002). These models also locate specific categorization for object recognition as the end goal of visual perception. According to these theories, while the gist of the scene, gross categories, and some contextual features are extracted at early stages of processing, exact categorization is achieved at final stages of processing, as a result of coactivation of high and low areas (e.g., Bar, 2004; Hochstein & Ahissar, 2002; Lamme, 2010; Schendan & Stern, 2008).

These characterizations of visual processing have implications for understanding the properties of representations stored in PM and visual WM (see Table 1 for a summary). PM representations are early, partial, and relatively involuntary. They are created en passant, in an automatic and obligatory manner, as part of ongoing perception. These representations include the perceived object's specific features, such as hue, shape, and orientation, all in metric/continuous space—namely, in analog representations. In addition, complete visual processing ends in the assignment of objects to categories. Indeed, studies suggest that visual representations entail automatic encoding of categories (Bird et al., 2014) or verbal tags in memory as well (Postle, 2006; Potter, 1993), although these may be weak when unattended (Cowan, 1999). Thus, perception of visual stimuli would result in *episodic-perceptual* traces of the visual properties of these stimuli in PM (e.g., a particular shade of green), and consequently in activation of *semantic* traces of conceptual tags related to these properties (e.g., the color tag "green") in semantic memory.

Characteristic	PM	Visual WM
Content	Analog/metric information: Representations of specific perceptual features	Digital/symbolic information: Structures including perceptual and conceptual dimensions
Attention	Unattended	Attended
Item-to-context binding	Weak	Strong
Encoding	Automatic	Controlled
Retrieval	Signal-detection process	Threshold process

Table 1. Characteristics of representations in perceptual memory (PM) and visual working memory (WM).

Not all visual input reaches visual WM eventually—only a subset of the perceived information. However, this optional level of processing, when utilized, is the final station of attending to products of perceptual or response processing (e.g., Chun, 2011; Postle, 2006). Visual WM thus holds information that is already completely processed in lower-level perceptual modules, including feature extraction, feature binding, and categorization. Thus, representations in visual WM are *conceptual* (at least whenever any categorization can take place) and include the assignment of the represented items to their categories, such as color tags. In this sense, then, visual WM stores digital information, which although not necessarily held in discrete states of strength may appear as if it is discrete (i.e., may appear slotlike, or as if capacity is quantized) due to the digital information its representations convey. This is in sharp contrast to PM, which comprises partial, incomplete, and nonunitized pieces of information that result from earlier stages of perceptual analysis.

Selection of items from PM to visual WM

Following this logic of VSTM as an emergent property of visual perception, a plausible mechanism by which levels of activations within PM might be determined is the use of *priority maps* (e.g., Franconeri, Alvarez, & Patrick, 2013; Knops et al., 2014; Melcher & Piazza, 2011). Priority maps are spatial maps on which the distribution of visual attention across locations is represented by assigning weights to different locations based on visual saliency, reward values, and task relevance (e.g., Fecteau & Munoz, 2006; Serences & Yantis, 2006; Zelinsky & Bisley, 2015). Importantly, the history of past stimuli is also represented in these maps (for a recent review, see Failing & Theeuwes, 2017). In perception, items in locations receiving greater sums of activation in the priority maps are then selected, or prioritized, for further processing and visual awareness (Fecteau & Munoz, 2006; Serences & Yantis, 2006; Zelinsky & Bisley, 2015). The different weighting factors result in items having variable, relative strength, such that unselected items are still processed to certain degrees (Serences & Yantis, 2006).

A growing body of research suggests that VSTM shares a common priority map with perceptual attention (Franconeri et al., 2013; Hedge, Oberauer, & Leonards, 2015; Theeuwes, Belopolsky, & Olivers, 2009) and enumeration (Knops et al., 2014; Melcher & Piazza, 2011). As would be expected, performance in VSTM tasks has been found to be modulated by visual saliency (e.g., Gaspar, Christie, Prime, Jolicœur, & McDonald, 2016; Klink, Jeurissen, Theeuwes, Denys, & Roelfsema, 2017; Melcher & Piazza, 2011), reward-related history of stimuli (Gong & Li, 2014; Infanti, Hickey, & Turatto, 2015; Klink et al., 2017), and task relevance (e.g., Heuer, Crawford, & Schubö, 2017; Melcher & Piazza, 2011). Importantly, Melcher and Piazza (2011) found that top-down and bottom-up saliency had a cumulative effect on change-detection performance (though bottom-up saliency seems to be processed faster; see Klink et al., 2017), supporting the notion of continuous levels of activation in VSTM.

Following these findings, Melcher, Piazza, and colleagues (Knops et al., 2014; Melcher & Piazza, 2011) suggested that competition between items in priority maps could underlie VSTM capacity limitations (see also Franconeri et al., 2013). Taken together with another finding that VSTM capacity estimates and subitizing range are correlated only under low maintenance load, they concluded that visual memory of objects includes two processes: individuation of objects in priority maps (shared with other capacity-limited operations such as subitizing; see also Mazza & Caramazza, 2015) and an additional maintenance process unique to VSTM (Melcher & Piazza, 2011). In this account, priority maps determine which items would be maintained and in what resolution, up to an upper item limit (Dempere-Marco et al., 2012). This account is also supported by an attractor-networks computational model that was able to account for both behavioral performance and neural activity patterns (Dempere-Marco et al., 2012), indicating that the two processing stages (especially during encoding) can determine actual capacity.

Here we argue that VSTM relies on two sets of representations rather than one. While we agree that the use of priority maps determines capacity allocation, and that gating of items to visual WM is based on

activity in these maps, we contend that the selection process renders visual WM representations qualitatively distinct from PM representations. Thus, because PM activation is determined by priority maps, PM representations are continuous in two respects: Their representational states are continuous (due to variable summation of weights in the priority maps), and the information they convey is analog. On the basis of attentional weights on these maps, a gating mechanism can select items to be represented in visual WM, where their strength would be maintained. Thus, there is an item limit in visual WM but not in PM.

Perceptual memory as an automatically encoded nonselective form of VSTM

As already explained, PM is characterized as a storage mechanism that contains automatically encoded representations of metric features of perceived visual stimuli (e.g., hues and orientation in degrees). Levels of activation of these representations are governed by priority maps, whereby weights based on top-down and bottom-up attentional priorities are combined to determine allocation of attention. While items in visual WM are most accessible for retrieval by definition (e.g., Cowan, 1999), PM representations are less accessible for deliberate retrieval (namely, recollection). One manner in which retrieval from PM can be accomplished is through familiarity signals (e.g., Oberauer, 2009). Another option for retrieval of items from PM is via gating items to visual WM, for example by directing attention to a PM representation based on retro-cues. This process is likely to take longer than recollection from visual WM (as would be suggested by the latency of retro-cue benefits, which requires at least 300 ms; see Souza & Oberauer, 2016). However, items outside visual WM are more susceptible to interference, and therefore interference from other relevant stimuli (such as secondary tasks or test displays; e.g., Pinto, Sligte, Shapiro, & Lamme, 2013; Souza & Oberauer, 2016) may undermine the accessibility of PM representations. In this sense, PM representations are relatively short-lived in terms of being accessible for deliberate retrieval. Nonetheless, under low interference these representations could exist in parallel to visual WM and be made accessible for longer periods of time.

Although PM representations are less accessible for deliberate retrieval and are more susceptible to interference compared to visual WM representations, they affect performance within a given trial and also across trials. Within a given trial (e.g., attempting to remember a memory array), items that are represented solely in PM (and not in visual WM) affect performance in two general ways. First, they generate familiarity signals (see Oberauer, 2009), which can be used for retrieval, although with lowered precision and heightened susceptibility to swap errors compared to items retrieved from visual WM (which are retrieved via recollection). Second, these items give rise to encoding of summary statistics (such as mean size), which bias performance in that trial toward the ensemble statistic (Brady & Alvarez, 2011). Across trials, these representations continue to affect performance through accumulating history of past stimuli (e.g., as evident in sequential effects), reflecting automatic, trial-by-trial updating of the content of PM (Failing & Theeuwes, 2017). It should be noted that by "updating" we do not necessarily mean removal of stimuli (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Nadel, Hupbach, Gomez, Newman-Smith, 2012); rather, automatic updating can be conceptualized as automatic modifications in weights in the priority maps based on the accumulation of the last several trials.

To summarize, we characterize PM as a large-capacity storage mechanism in which analog representations of visual items are maintained in varying degrees of activation. These items are maintained in parallel to visual WM but are less accessible for controlled retrieval; nonetheless, they affect performance within a given trial and between trials. It should be made explicit that although this description may appear similar to the notion of iconic memory, we do not consider PM as a temporary station before visual WM, holding extremely short-lived representations. Rather, we argue that PM representations can be maintained in parallel to visual WM representations, and continue to affect performance even after masking, after relatively long periods of time since stimulus offset, or in subsequent trials.

In this section we provide evidence for such an automatic form of VSTM that does not have an item limit. First, we provide evidence indicating that VSTM performance is affected by automatic processes. Next, we address findings indicating that under some conditions suitable for reliance on PM representations, VSTM capacity is less restricted than when visual WM is recruited. Finally, we review evidence that PM representations have weak item-to-context bindings.

Automatic effects

Automatic processes are processes that take place spontaneously, even if they are not part of task requirements (Tzelgov, 1997). Due to the nonselective nature of PM, we suggest that PM representations can be formed and activated automatically. Moreover, as demonstrated later, retrieval from PM is automatic because it gives rise to effects that take place implicitly

or without intention. Being automatic, these effects could bias responses or even impair performance.

Several effects of automatic processing are observed in VSTM task performance. First, Brady et al. (2009) found that participants implicitly learn regularities in color pairings of items over trials in a change-detection task and use this learning to improve their performance in trials where regularities appear. Further evidence for automatic processing comes from sequential effects in VSTM performance (e.g., Huang & Sekuler, 2010; Wilken & Ma, 2004). For example, Huang and Sekuler (2010) asked participants to perform a delayed-estimation task with Gabor patches and found that estimations of spatial frequencies were biased toward the mean frequency of the stimuli presented in the memory arrays of the last three to five trials. Both implicit learning of statistical properties and sequential effects are clearly memory phenomena, as they rely on internal representations of prior (but recent) stimuli. However, they occur automatically (e.g., Brady & Oliva, 2008; Fiser & Aslin, 2001; Kessler, 2018; Kim, Seitz, Feenstra, & Shams, 2009; Rac-Lubashevsky & Kessler, 2016b) and are not part of the task requirements, since the information presented in previous trials is no longer relevant.

The effects of automatic processing can also be demonstrated within a given trial, not only across trials. This point is important to our discussion, given that we regard PM as containing perceptual short-term representations activated within the course of a single trial. Indeed, recent studies have shown that delayed-estimation performance is affected by ensemble statistics, such as mean size (Brady & Alvarez, 2011) or mean spatial frequency (Huang & Sekuler, 2010; Wilken & Ma, 2004) of items within a memory array, as well as by regularities within a display (Brady et al., 2009; Brady & Alvarez, 2015; Victor & Conte, 2004). For example, Brady and Alvarez (2011) asked participants to estimate the size of an item in a delayed-estimation task and found that estimation was biased toward the mean size of the group of objects the probe belonged to (e.g., circles from a certain color). It is important to note that representing mean size (as well as averaging across other dimensions) of groups of objects is automatic, in that it takes place implicitly, without conscious intent, or without being part of task requirements (Brady & Alvarez, 2011; Turk-Browne, Jungé, & Scholl, 2005; for automatic averaging of orientation, see Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Furthermore, perception of mean size is accurate even after short exposure durations (e.g., 50 ms) and is hardly affected by the number of items to be summarized (Ariely, 2001; Chong & Treisman, 2003). However, research suggests that although this is an automatic process, it is executed only with regard to task-relevant dimensions (Brady & Alvarez, 2011;

Turk-Browne et al., 2005). Brady and Alvarez argue that this effect demonstrates that WM is susceptible to the same effects that LTM is susceptible to. We interpret these results as reflecting the involvement of PM in within-trial performance on VSTM tasks, due to representing information in parallel to visual WM.

These findings indicate that summary statistics are represented automatically, as a by-product of perception. However, they do not necessarily indicate that items outside visual WM were represented individually, other than a summary representation. Bronfman, Brezis, Jacobson, and Usher (2014) found evidence for a summary statistic that relies on differentiating colors of separate representations, indicating that items were represented separately. They asked participants to perform a Sperling task: to remember colored letters arranged in four rows of six items. Afterward, a retro-cue indicating the row to be probed appeared, and was followed by a question mark in one location in the retro-cued row; participants were to report the identity of the letter in the marked location. In addition, the researchers orthogonally manipulated the color diversity of the letters in the cued and noncued rows, such that letter colors were either nondiverse (e.g., different shades of blue and purple; low diversity) or diverse (e.g., yellow, red, green, etc.; high diversity). In separate blocks, after reporting the probed letter's identity, participants were asked to judge whether the color diversity of either the cued row or the three noncued rows was high or low. Participants were able to judge the color diversity of the noncued rows above chance level. Moreover, memory performance for letter identities was not affected by whether participants were asked to judge diversity or not, nor by whether diversity was judged for the cued or the noncued rows. Furthermore, when the researchers introduced a mask and asked participants to report only the visibility and the color diversity (without performing the Sperling task itself), participants were able to report diversity above chance level—but only if the array was not presented subliminally. Because at least part of the items in the cued row (containing six items) are in visual WM and items in noncued rows are outside visual WM (containing at least 18 items), the findings indicate that participants were able to use a summary statistic that requires differentiated representations maintained outside visual WM, within a single trial, without intent, after masking, and without a capacity cost. This provides further support for the notion of larger-capacity, PM-based automatic effects of summary statistics.

Recently, Lorenc, Sreenivasan, Nee, Vandenbroucke, and D'Esposito (2018) used an fMRI inverted encoding model in order to examine the effect of distractor interference on VSTM representations in a delayed-estimation task. They found evidence for two

distinct representations of memory items: one in early visual areas, and the other in the IPS. While IPS representations were stable at the face of distractors, early visual areas were more susceptible to interference, as indicated by a distractor bias. These findings support both the distinction between PM and visual WM representations and the automatic nonselective nature of PM representations.

Large capacity for visual memoranda

Endress and Potter (2014) examined forgetting in VSTM using a paradigm of rapid serial vision presentation with real-world objects as stimuli. Participants were presented with a sequence of pictures and then asked to indicate whether a test item was part of the sequence or not. Importantly, proactive interference was manipulated by either repeating items between trials or using unique stimuli. The results revealed that while capacity estimates were low in the proactive-interference condition, they increased dramatically to up to 30 items in the low-interference condition, with no evidence for a fixed upper limit in this condition. Furthermore, a manipulation of the sequence-test lag duration revealed that representations in the low-interference condition were stable but decayed within several seconds. Endress and Potter's conclusion was that their results support the existence of an "unconsolidated form of LTM that functions as a temporary memory store" (p. 561). This description maps nicely onto our concept of PM.

Endress and Potter's results and conclusions coincide with findings regarding memory of complex visual scenes over periods of several minutes (e.g., Konkle, Brady, Alvarez, & Oliva, 2010; Melcher, 2001, 2006), indicating large visual capacity independent of WM. In the case of complex scenes, participants are able to discriminate between studied and novel scenes (e.g., Konkle et al., 2010) or objects presented within scenes (e.g., Hollingworth & Henderson, 2002; Melcher, 2001, 2006) with high precision, even after studying large numbers of visual scenes. The fidelity of these representations continues to increase with repeated exposure (Melcher, 2001), even if other to-be-remembered scenes or an additional task are introduced between these viewings (Melcher, 2006), indicating that these representations are not held in WM. However, these representations do not seem to be consolidated to LTM for more extended periods of times (e.g., days; Melcher, 2006). Furthermore, the fidelity of memory of scenes decreases when more exemplars from the same categories are encoded, indicating that between-items interference may reduce capacity for complex scenes (Konkle et al., 2010). These findings have led Melcher (2001, 2006) to suggest a "medium-short" memory mechanism, a large-capacity storage that maintains

visual representations over several minutes which does not consolidate into LTM and is independent of visual WM. Endress and Potter's (2014) study likely taps the same kind of "medium-short" memory mechanism, which is parallel to WM. These representations seem to be accessible for retrieval under low interitem interference.

Further support for the involvement of PM in VSTM tasks, indicating task impurity, comes from a line of research regarding the *retro-cue effect*. This well-established finding is that a cue which appears during the retention interval in a change-detection task (i.e., after the presentation of a memory array but before the appearance of the probe) and indicates the to-be-tested item, facilitates performance considerably, sometimes up to capacity estimates of 15 items (e.g., Griffin & Nobre, 2003; Landman et al., 2003; Sligte et al., 2008; for a review, see Souza & Oberauer, 2016). In delayed-estimation tasks, retro-cues facilitate performance mainly by increasing recall probability and lowering guessing rates (e.g., Makovski & Pertzov, 2015; Murray, Nobre, Clark, Cravo, & Stokes, 2013; Pertzov, Bays, Joseph, & Husain, 2013; Souza et al., 2016; Souza, Rerko, Lin, & Oberauer, 2014; Thibault, van den Berg, Cavanagh, & Sergent, 2016; van Moorselaar, Gunseli, Theeuwes, & Olivers, 2015). Importantly, the retro-cue effect reveals that the capacity of VSTM is underestimated (e.g., Makovski, Sussman, & Jiang, 2008; Souza & Oberauer, 2016).

Sligte and colleagues (Pinto et al., 2013; Sligte et al., 2008; Vandenbroucke, Sligte, & Lamme, 2011) attribute the retro-cue effect to an additional short-term memory store before WM, which is less restricted in capacity: fragile VSTM. That is, according to their account of the retro-cue effect, the increase in capacity results from information being stored outside visual WM, namely in fragile VSTM. They argue that three to four items are stored in robust visual WM, while several additional items are represented in the less accessible fragile VSTM. Given enough time, retro-cues allow one to direct attention to fragile VSTM representations and access and retrieve them by getting these representations to visual WM.

Although the fragile VSTM account of the retro-cue effect may be appealing, there is no consensus over whether this effect indeed indicates the existence of multiple memory stores, with some arguing that it can be accounted for under a single VSTM store (e.g., Makovski, 2012). The mechanisms underlying the retro-cue effect are not fully clear yet (for an evaluation of possible mechanistic accounts of the retro-cue, see Souza & Oberauer, 2016), but it is unanimous that the operation of retro-cues involves allocating attention to the cued item (e.g., Dell'Acqua, Sessa, Toffanin, Luria, & Jolicoeur, 2010; Lepsien, Thornton, & Nobre, 2011; Makovski et al., 2008; Oberauer & Hein, 2012; Souza &

Oberauer, 2016). However, any single-store account of the retro-cue effect must address how this cue facilitates performance—be it in accuracy rate or increased precision—without that information being stored elsewhere (e.g., Souza, Rerko, Lin, & Oberauer, 2014).

In our framework, we embrace the interpretation of the retro-cue effect as evidence for (at least) two *representational states* within a nested system (e.g. LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013; Rerko, Souza, & Oberauer, 2014; Souza & Oberauer, 2016; Zokaei, Ning, Manohar, Feredoes, & Husain, 2014). We suggest that the retro-cue effect reflects the distinction between representations in two hierarchical levels—PM and visual WM—where visual WM items are an attended subset of PM representations. If PM representations receive attention, as when probed by a retro-cue, they are gated into visual WM and become more accessible. This account is consistent with Endress and Potter's (2014) findings of increased capacity under low-interference conditions, indicating that capacity is less limited under low attentional demands. Furthermore, it is also consistent with findings indicating that retrieval of cued and noncued items is associated with activation in different brain regions (e.g., Schneider, Mertes, & Wascher, 2016; Sligte, Wokke, Tesselaar, Scholte, & Lamme, 2011; Vandenbroucke, Sligte, de Vries, Cohen, & Lamme, 2015).

Experimental findings support the notion that noncued items are represented in PM. PM representations are less accessible than visual WM items but can be retrieved when attention is directed toward them. First, the retro-cue benefit is observed only when there is sufficient time between the appearance of the cue and the test display (at least 300 ms; for a review, see Souza & Oberauer, 2016). Studies in which invalid retro-cues were used found lower performance on invalid trials compared to neutral or valid cues (*retro-cue cost*; e.g., Griffin & Nobre, 2003; LaRocque et al., 2015; Li & Saiki, 2014; Pertzov et al., 2013; van Moorselaar, Olivers, Theeuwes, Lamme, & Sligte, 2015), indicating that noncued items are indeed less accessible and providing evidence that some of the benefit of the retro-cue is at least partially due to removal of noncued items (for an analysis of the removal hypothesis, see Souza & Oberauer, 2016). However, these items are not entirely forgotten, and they can still be accessed later. Studies that used a double-cueing paradigm found that items not cued by a first retro-cue but subsequently cued by a second cue could be retrieved with high accuracy (Heuer & Schubö, 2016; Landman et al., 2003; Li & Saiki, 2014; Rerko & Oberauer, 2013; van Moorselaar, Olivers, et al., 2015; for supporting evidence using a different paradigm, see Zokaei et al., 2014).

Taken together, there is evidence for a large-capacity mechanism which operates in parallel to WM and

includes representations maintained for a short to medium period of time: PM. Representations maintained in PM can be recollected under certain circumstances—that is, when interitem interference is low or when top-down cues can be used to gate PM items into WM (provided that there is sufficient time for retrieval, as indicated by the time course of the retro-cue effect). In other cases, these representations affect performance via automatic effects, such as statistical biases and sequential effects between trials. These representations may continue to exist in parallel to WM for several seconds, and perhaps minutes, as evident in the double-cueing effect, sequential effects across trials, and memory for complex scenes. The longevity of PM representations is likely determined by several factors, such as the amount of interitem interference and each item's weight in the priority map (based on history, saliency, reward, etc.); future research should further examine the time course of these representations.

Weak item-to-context binding

Another difference between PM and visual WM representations lies in the strength of item-to-context bindings: PM representations have weak item-to-context bindings, whereas items in visual WM have strong ones. This should have an important implication for swap errors, which denote mistakenly reporting features of wrong items. Indeed, retro-cuing has been associated with reduced swap errors in studies incorporating retro-cues in delayed-estimation tasks (Makovski & Pertzov, 2015; Souza et al., 2016; Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015). Furthermore, noncued items probed in invalid cue trials are associated with an increase in swap errors (Gunseli, van Moorselaar, Meeter, & Olivers 2015).

The need for item-to-context binding may be especially important in holding representations of complex objects. Building on the embedded-process model, complex objects are understood in our framework as constructs in which PM representations of their features in different dimensions (e.g., colors, shapes) are bound to a single location (context). For items in visual WM, feature-location (item-to-context) bindings are amplified. Locations act as potent contextual cues for retrieval (Oberauer, 2009; see also Pertzov & Husain, 2014), and therefore location retro-cues should be more beneficial than feature retro-cues (color, shape, etc.). This is supported by studies showing better accuracy in change detection with complex objects on a location retro-cue condition compared to a color retro-cue condition (Li & Saiki, 2015; but see Kalogeropoulou, Jagadeesh, Ohl, & Rolfs, 2017), and lower performance when participants are asked not to encode locations compared to being

asked not to remember other features (Kondo & Saiki, 2012). Additional support for this notion comes from the finding of Cowan et al. (2014) that visual peripheral capacity, attributed to aLTM, is more restricted when binding information, rather than simple features, is to be retained (for additional supporting evidence, see also Fougnie & Marois, 2011).

Visual WM as a seemingly quantized, selective, and conceptual form of VSTM

As explained earlier, visual WM holds items for which the perceptual processing is complete. In these items, features such as hue and orientation are bound together, as well as to their context (e.g., location) and to a conceptual tag (e.g., "green," "tilted to the right"). The major property rendering visual WM representations slotlike, we argue, is the conceptual nature of these items.

Conceptual tags in visual WM

We suggest that the major factor that renders visual WM representations discrete is their conceptual, or categorical, nature. Categorical representations contain symbolic information ("green," "red"). Unlike analog representations, which are set on one continuum and therefore may elicit high levels of interference, symbolic representations can assist maintenance by helping distinguish between maintained items and thus reducing interitem interference (e.g., Brady, Störmer, & Alvarez, 2016). This feature of symbolic-categorical representations can lead to quantized-like effects in VSTM tasks.

Studies suggest that VSTM capacity estimates increase as memory items are easier to categorize. Olsson and Poom (2005) provided initial evidence for the involvement of categorization in change-detection performance by comparing VSTM capacity for items from distinct categories (e.g., discrete colors or different shapes) and for items that do not fall into discrete categories (e.g., ovals in different proportions). In accordance with our framework, they found that estimated capacity was higher for items from distinct categories. Furthermore, VSTM capacity was estimated as being only about one memory item in the noncategorical condition, indicating that conceptual representations assist in maintenance and accessibility of more items, namely three to four items. Brady et al. (2016) also found a higher capacity estimate as well as larger amplitude of contralateral delay activity for meaningful real-world items compared to simple colored squares. Because contralateral delay activity is associated with visual WM maintenance and not with LTM consolidation (Brady et al., 2016; Carlisle, Arita,

Pardo, & Woodman, 2011), these findings provide additional support for the benefit of concepts in visual WM maintenance.

According to our framework, visual WM representations involve bindings of continuous information (stored in PM) to symbolic concepts. This characteristic can be evaluated by using the delayed-estimation paradigm, where performance should be influenced both by the specific feature of the probed item (token) and by the category the estimated item belongs to (type). Indeed, Bae et al. (2015) have found evidence for the existence of such combined continuous-categorical representations. Their results revealed that color delayed estimation of one item was biased toward the exemplar of a color category, especially for probes at category boundaries (i.e., that were more distant from the category exemplar on the continuous space). Interestingly, this bias was present also in undelayed estimation—that is, when participants estimated probes' colors while they were still present on the screen—though to a lesser degree. Konkle and Oliva (2007) reported a similar finding of size bias in a change-detection task with real-world objects: Detection of change was lower when the size of the probe changed toward the normative size of the real-world object compared to changes that deviated away from the normative size. Thus, the results from Bae et al. and Konkle and Oliva reveal that biases toward category centers are evident in perceptual tasks (for neuroscientific evidence in perception, see also Bird et al., 2014) but appear to get larger in VSTM tasks. These findings indicate that perceptual representations are both analog/continuous and digital/categorical already at the level of PM (as an emergent property of perception); however, their categorical aspect is *amplified* in visual WM (but see Hardman et al., 2017).

Strong item-to-context binding in visual WM

According to the embedded-components theory, the major defining function of visual WM (the broad FoA) is to make distinct items accessible and enable integration as well as identification of similarities and differences between items. This is enabled by binding content items to a context set on a cognitive coordinate system (Oberauer, 2009). Visual WM is limited in the number of bindings that can be maintained effectively with minimal interitem interference (Oberauer, 2009; see also Oberauer et al., 2016; Oberauer, Farrell, Jarrold, Pasiecznik, & Greaves, 2012).

Evidence from VSTM studies supports these notions. First, as already described, evidence indicates that while item-to-context binding in PM is quite weak, visual WM strengthens these connections. Second, this difference in item-to-context binding between PM and visual WM can be implemented neurally. Swan and

Wyble (2014) proposed the "binding pool model," a computational neural-network model of VSTM. In that model, VSTM representations are links between types and token, the former being simple features and the latter being pointers used for binding of different features. The type layer is continuous and resource based, while the token layer is slots based. This conceptualization is, in general, quite similar to our distinction between continuous PM and discrete visual WM. That model yielded characteristic results obtained in VSTM experiments, such as the set-size effect in precision and swap errors, indicating the plausibility of representations based on two levels, a continuous one and a discrete one.

It should be noted that the manner in which item-to-context bindings limit capacity can be subject to different interpretations. In Swan and Wyble's (2014) model and in Oberauer and Lin's (2017; as well as in other interference-based models of nonvisual material; see Oberauer et al., 2012; Oberauer et al., 2016) all to-be-remembered items are represented in visual WM (though some memory items may not be represented at all). Performance at large set sizes is degraded because items inside visual WM interfere with one another. Under this interpretation, although visual WM representations themselves are discrete, binding links can be thought of as having varying degrees of activation. Oberauer and Lin's interference model explains the retro-cue effect as a result of shifts in the narrow FoA, which strengthens one memory item. This model outperformed other models, such as the slots + averaging model and the variable-precision resource model (Oberauer & Lin, 2017).

In contrast to the interference-based models, we suggest that in large set sizes only three or four items are represented in visual WM, whereas *all* items (including the former) are represented in PM. Accordingly, supraspan items are held only in PM. We explain the set-size effect as resulting from less accessibility of items represented solely in PM. While Oberauer and Lin's model is admittedly more parsimonious than our framework (having a single memory store), more experimental investigations as well as modeling effort should address the issue of whether memory traces exist outside visual WM (namely, the broad FoA).

However, it is plausible that the known three- to four-item limit (which is found across modalities and is not restricted to VSTM) is not structural but rather practical, and can be subject to top-down control. That is, it is possible that three to four items are maintained in visual WM in order to keep interference minimal (and thus ensure high performance for a subset of the memory items), but this number can be changed (while bearing a performance cost) according to task requirements (see "Influence of task requirements on performance" subsection ; Bengson & Luck, 2016).

Summary and conclusions

Returning to the discrete- versus continuous-capacity debate

In the beginning of this article, we reviewed benchmark findings regarding the debate over whether VSTM capacity is discrete or continuous. Overall, findings were intermixed, with both the discrete-capacity (quantized) and continuous-capacity stances receiving some supporting evidence. Furthermore, hybrid models appeared to account for data better than each class of models by itself.

Our framework provides a theoretical foundation for a mechanistic account of why both stances gain empirical support, and why hybrid models fit the data better. We argued and provided evidence for the view that VSTM is a hierarchical system in which one level is perceptual and contains automatically encoded analog representations without an item limit, and another level is conceptual and therefore may appear quantized. The most accessible items, those in visual WM, are structures relating continuous-perceptual information with discrete-conceptual information, and thus this mechanism appears both quantized and continuous, or limited in both the number of items and the precision with which they are retained. Thus, this framework provides an architecture fitting the recent suggestion that the number of maintained representations is quantized (as in slots models) while the precision with which these representations is retained may be determined by division of activation levels (as in shared-resource models; Fougnie et al., 2010).

The two-level distinction as a characteristic of visual processing

After having provided evidence for the two levels in the case of VSTM, a final note on how this framework combines with current theories of visual processing is in order. The idea that two parallel (but interacting) stages of analysis are based on different kinds of representations—analog and automatic or digital and controlled—seems to be a common thread among theories explaining phenomena in various domains of perceptual, and specifically visual, processing. One example is the interaction between object and gist-of-scene perception (e.g., Bar, 2004; Hochstein & Ahissar, 2002; Melcher & Colby, 2008). While the gist of the scene (and its effect on performance, for example in change-blindness tasks) is based on statistical properties and regularities within the whole display, the recognition of specific objects is more analog and is limited to only several items. A second example, which

bears a striking resemblance to the framework we have described, comes from the study of enumeration in numerical cognition.[1] Anobile, Cicchini, and Burr (2016) recently suggested that two parallel processes take part in enumeration (in displays with low to medium densities): an attention-demanding level of subitizing that is dominant in the range of one to four items, and a more automatic level of numerosity estimation which is not limited in range of numbers (i.e., the whole range) and is dominant for processing of large numerosities outside the subitizing range (i.e., more than four items). Evidence for the difference in levels of processing of high and low numerosities comes from studies showing that attention-demanding manipulations (e.g., dual tasks, attentional blink, WM load) lead to impaired performance in the subitizing range but do not affect numerosity estimation (e.g., Anobile, Turi, Cicchini, & Burr, 2012; Burr, Turi, & Anobile, 2010; Olivers & Watson, 2008; Piazza, Fumarola, Chinello, & Melcher, 2011; Vetter, Butterworth, & Bahrami, 2008; X. Xu & Liu, 2008). Thus, it appears likely that this line of research provides further support for an automatic, large capacity PM, as an emergent product of visual perception. It is likely that the two enumeration processes map onto the same levels we suggest to underlie VSTM storage.

Predictions and challenges

Several empirical predictions can be derived from the framework. We will focus here on predictions based on the premise that two levels of VSTM operate in parallel: a large-capacity automatically encoded continuous level (PM) and a selective symbolic level (visual WM). In general, when set size is supraspan, some memory items are encoded only in PM, while three or four are represented in visual WM as well. The likelihood that a probed item would be represented only in PM increases with set sizes. Therefore, markers of PM representations are expected to be stronger at large set sizes compared to small set sizes, while markers of visual WM are expected to be stronger at small versus large set sizes. If retro-cues are signals for gating items to visual WM, then performance should be based primarily on visual WM regardless of set size.

Automaticity

One marker of PM representations, we argue, is that they give rise to automatic effects. Therefore, we predict that effects of automaticity would interact with set size. For example, the effect of ensemble statistics would be small for at-capacity set sizes but would increase with set size. Moreover, this interaction is

expected to dissolve under retro-cue conditions. This prediction receives preliminary support from Wilken and Ma's (2004) finding of an interaction between set size and reproduction error in a delayed-estimation task with Gabor patches, where a bias toward the mean frequency of the display grew larger as set size increased (but it should be mentioned that they did not find a similar bias in delayed-estimation of colors or orientation).

Updating

Following the same line of thought, another prediction derived from this framework is that when the content of VSTM needs to be updated, two updating processes are expected to emerge: controlled updating of relevant information in visual WM and automatic-obligatory updating of PM. Rac-Lubashevsky and Kessler (2016b) recently found evidence for automatic and controlled updating processes using the reference-back task, a variant of the *n*-back paradigm in which two types of trials were introduced: reference trials and comparison trials. Participants were asked to indicate whether each stimulus matched the last reference stimulus. In reference trials, they had to update the content of WM in order to compare the next stimuli to this item. However, in comparison trials they were required only to compare the presented stimulus with the reference item, without replacing it. Findings revealed that trial history (four trials backward) affected performance in both comparison and reference trials, and that reference trials had an additional additive cost in performance, reflecting WM updating. These findings indicate that short-term representations of stimuli presented over the last trials continued to affect performance even when those items were not encoded into WM, and (following the logic of additive factors; compare Sternberg, 1969) this automatic updating occurred in a different level of processing than controlled updating (for additional evidence, see Kessler, 2018). Future studies should attempt to combine similar methods with VSTM paradigms in order to examine whether sequential effects are additive to controlled updating when updating visual short-term memoranda.

Combination of continuous and categorical representations

In addition, according to our framework two types of representations are maintained—analog/continuous (PM) and symbolic/categorical (visual WM). This characteristic leads to the prediction that when performance on the delayed-estimation task is modeled, a combination of continuous and categorical representations should be evident. As already argued,

this characteristic is evident in delayed-estimation performance in general. However, a more specific prediction can be made: Items in at-capacity sets should be represented by both types of representations, while in supracapacity sets some items should be only continuous (without a category-center bias, but also with less precision).

In a recent study, Hardman et al. (2017) asked participants to perform a sequential version of the delayed-estimation task in which memory items were presented one by one and participants were asked after the retention interval to reproduce the colors of all memory items. The researchers modeled participants' responses and found that a model in which items could be represented either categorically or continuously fit performance better than a model in which they could be represented both ways (i.e., as in Bae et al., 2015). Furthermore, they found that only one item was represented continuously and not categorically. These findings are not in agreement with our predictions, and are also contradictory to the results of Bae et al. (2015). It is possible that the sequential presentation and probing in the experiments by Hardman et al. may have affected their findings by eliciting sequential effects, such as primacy and recency, or eliciting a higher susceptibility to interference (which we suggest should primarily affect continuous PM-based representations). Thus, further studies and analyses based on serial order should be conducted in order to settle the differences between the findings of Hardman et al. and Bae et al., and to examine whether our prediction receives empirical support. In addition, response patterns associated with continuous and categorical representations should be examined in larger set sizes as well, in order to distinguish between PM and visual WM representations.

Challenges and open questions

Although ideal performance in VSTM tasks would be based on a combination of visual WM and PM representations, it is possible that under certain conditions one level is more dominant than the other. For example, the change-detection task usually requires making gross same/different judgments, and thus performance on this task is likely to tap mainly the discrete aspect of these structures. On the other hand, the delayed-estimation task requires reporting specific features' values, and thus is likely to tap mostly the continuous-analog aspects of these structures, although it is also affected by the discrete aspect (as evident in category bias). In the delayed-estimation task, it is likely that some responses would be WM based and some would be PM based, depending on whether the probed item was represented in WM or not. WM-based responses should be fast, relatively precise, but biased

toward category center. PM-based responses should be slower, biased toward summary statistics and stimuli history, and with increased rate of swap errors. However, because the two mechanisms act in parallel, there are several possibilities for their combined effect on performance, depending on whether they compete (and in what manner) or interact with one another. An empirical challenge would be to disentangle the contribution of each level to performance on both tasks, and reveal the relations between the two retrieval processes. One way to examine this question is to examine the fit of mixture models to accuracy data (including a swap-error parameter) separately for trials with fast and slow RTs. Another line of inquiry can involve adapting the system factorial technology (SFT) logic to VSTM tasks, in order to examine whether the two retrieval processes compete in horse-race, parallel exhaustive, or coactivation architectures.

Another empirical challenge is to understand how individual differences at each level of the hierarchy affect capacity estimates. Vogel et al. (2005) found that individuals with low visual WM capacity have poor control over the content of WM compared to high-capacity individuals, indicating that some of the individual differences in capacity stem not from the number of retained items but rather from regulation of the content of VSTM (see also Gaspar et al., 2016). In our framework, this could be manifested either in allocation of activation in PM (e.g., in top-down weights related to task relevance) or in inefficient selection of items based on these maps to enter visual WM.

Finally, another open question concerns how the two levels map onto LTM encoding and retrieval. Specifically, a question to be raised involves whether PM representations can be encoded to episodic LTM and in what form. One possibility is that PM representations can be retained as episodic traces that contain little contextual information, depending on each item's level of activation in the priority maps, while visual WM items are encoded with stronger binding to context. If PM items are retained in the form of contextless episodic traces, one would expect that PM representations would be associated with familiarity-based retrieval, while visual WM representations would be associated with recollection-based retrieval (e.g., Diana, Yonelinas, & Ranganath, 2007).

Conclusions

In this article we argued for a hierarchical architecture of VSTM, composed of two levels of representation: perceptual memory, storing analog representations of visual stimuli in varying activation levels, and visual WM, storing digital conceptual

representations of a subset of three or four items from PM and binding them to their context. While PM has a large capacity, is relatively nonselective, and gives rise to automatic effects, visual WM is restricted in the number of items that can be maintained simultaneously, and its content is regulated by a gating mechanism. Because items can be represented in both PM and visual WM, capacity allocation may appear as either discrete/quantized or continuous, depending on the task requirement and the exact method used to measure VSTM. Thus, our framework reconciles the mixed findings regarding whether VSTM is quantized or not.

Keywords: visual working memory, discrete capacity, shared-resource models, embedded components, conceptual representations

Footnote

[1] We thank David Melcher for this insight.

References

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

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

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number as a primary perceptual attribute: A review. *Perception*, *45*, 5–31, https://doi.org/10.1177/0301006615602599.

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2012). The effects of cross-sensory attentional demand on subitizing and on mapping number onto space. *Vision Research*, *74*, 102–109, https://doi.org/10.1016/j.visres.2012.06.005.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162, https://doi.org/10.1111/1467-9280.00327.

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

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*, 744–763, https://doi.org/10.1037/xge0000076.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences, USA*, *93*, 13468–13472.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629, https://doi.org/10.1038/nrn1476.

Bays, P. M. (2018). Failure of self-consistency in the discrete resource model of visual working memory. *Cognitive Psychology*, *105*, 1–8, https://doi.org/10.1016/j.cogpsych.2018.05.002.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11, https://doi.org/10.1167/9.10.7. [PubMed] [Article]

Bays, P. M., & Husain, M. (2008, August 8). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854, https://doi.org/10.1126/science.1158023.

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin & Review*, *23*, 265–270, https://doi.org/10.3758/s13423-015-0891-7.

Bird, C. M., Berens, S. C., Horner, A. J., & Franklin, A. (2014). Categorical encoding of color in the brain. *Proceedings of the National Academy of Sciences, USA*, *111*, 4590–4595, https://doi.org/10.1073/pnas.1315275111.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory ensemble statistics bias memory for individual items. *Psychological Science*, *22*, 384–392, https://doi.org/10.1177/0956797610397956.

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Spatial

ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 921–929, https://doi.org/10.1037/xlm0000075.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487–502, https://doi.org/10.1037/a0016797.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes extracting categorical regularities without conscious intent. *Psychological Science, 19*, 678685, https://doi.org/10.1111/j.1467-9280.2008.02142.x.

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences, USA, 113*, 7459–7464, https://doi.org/10.1073/pnas.1520027113.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 713–737). Cambridge, MA: MIT Press.

Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: "cost free" color phenomenality outside focal attention. *Psychological Science, 25*, 1394–1403, https://doi.org/10.1177/0956797614532656.

Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision, 10*(6):20, 1–10, https://doi.org/10.1167/10.6.20. [PubMed] [Article]

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *The Journal of Neuroscience, 31*, 9315–9322, https://doi.org/10.1523/JNEUROSCI.1097-11.2011.

Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: A computational model of the n-back task. *Journal of Cognitive Neuroscience, 23*, 3598–3619, https://doi.org/10.1162/jocn_a_00047.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404, https://doi.org/10.1016/S0042-6989(02)00596-5.

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*, 1407–1409, https://doi.org/10.1016/j.neuropsychologia.2011.01.029.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*, 331–335, https://doi.org/10.3758/BF03196169.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163–191, https://doi.org/10.1037/0033-2909.104.2.163.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.

Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia, 49*, 1401–1406, https://doi.org/10.1016/j.neuropsychologia.2011.01.035.

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, https://doi.org/10.1037/a0029687.

Cowan, N., & Rouder, J. N. (2009, August 8). Comment on "Dynamic shifts of limited working memory resources in human vision." *Science, 323*, 877, https://doi.org/10.1126/science.1166478.

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

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671–684, https://doi.org/10.1016/S0022-5371(72)80001-X.

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences, USA, 109*, 19900–19909, https://doi.org/10.1073/pnas.1116727109.

Dell'Acqua, R., Sessa, P., Toffanin, P., Luria, R., & Jolicœur, P. (2010). Orienting attention to objects in visual short-term memory. *Neuropsychologia, 48*, 419–428, https://doi.org/10.1016/j.neuropsychologia.2009.09.033.

Dempere-Marco, L., Melcher, D., & Deco, G. (2012) Effective visual working memory capacity: An emergent effect from the neural dynamics in an attractor network. *PLoS One*, 7, e42719, https://doi.org/10.1371/journal.pone.0042719.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, 11, 379–386, https://doi.org/10.1016/j.tics.2007.08.001.

Donkin, C., Kary, A., Tahir, F., & Taylor, R. (2016). Resources masquerading as slots: Flexible allocation of visual working memory. *Cognitive Psychology*, 85, 30–42, https://doi.org/10.1016/j.cogpsych.2016.01.002.

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, https://doi.org/10.1037/a0034247.

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, https://doi.org/10.3758/s13414-013-0561-7.

Ecker, U. K., Lewandowsky, S., Oberauer, K., & Chee, A. E. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 170–189, https://doi.org/10.1037/a0017891.

Egan, J. P. (1975). *Signal detection theory and ROC-analysis*. New York: Academic Press.

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General*, 143, 548–565, https://doi.org/10.1037/a0033934.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York: Academic Press.

Failing, M., & Theeuwes, J. (2017). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review*, 25, 514–538, https://doi.org/10.3758/s13423-017-1380-y.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, 10, 382–390, https://doi.org/10.1016/j.tics.2006.06.011.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504, https://doi.org/10.1111/1467-9280.00392.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12):27, 1–11, https://doi.org/10.1167/10.12.27. [PubMed] [Article]

Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quantity and quality in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1231–1240, https://doi.org/10.1037/xhp0000211.

Fougnie, D., & Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1329–1341, https://doi.org/10.1037/a0024834.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1–8, https://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, https://doi.org/10.1016/j.tics.2013.01.010.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 137–160, https://doi.org/10.3758/CABN.1.2.137.

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

Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicœur, P., & McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. *Proceedings of the National Academy of Sciences, USA*, 113, 3693–3698, https://doi.org/10.1073/pnas.1523471113.

Geigerman, S., Verhaeghen, P., & Cerella, J. (2016). To bind or not to bind, that's the wrong question: Features and objects coexist in visual short-term memory. *Acta Psychologica*, 167, 45–51, https://doi.org/10.1016/j.actpsy.2016.04.004.

Gong, M., & Li, S. (2014). Learned reward association improves visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 841–856, https://doi.org/10.1037/a0035131.

Graham, K. S., Barense, M. D., & Lee, A. C. (2010). Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia, 48,* 831–853, https://doi.org/10.1016/j.neuropsychologia.2010.01.001.

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

Gunseli, E., van Moorselaar, D., Meeter, M., & Olivers, C. N. (2015). The reliability of retro-cues determines the fate of noncued visual working memory representations. *Psychonomic Bulletin & Review, 22,* 1334–1341, https://doi.org/10.3758/s13423-014-0796-x.

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 30–54, https://doi.org/10.1037/xhp0000290.

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19,* 304–313, https://doi.org/10.1016/j.tics.2015.04.006.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience, 139,* 105–118, https://doi.org/10.1016/j.neuroscience.2005.04.067.

Hedge, C., Oberauer, K., & Leonards, U. (2015). Selection in spatial working memory is independent of perceptual selective attention, but they interact in a shared spatial priority map. *Attention, Perception, & Psychophysics, 77,* 2653–2668, https://doi.org/10.3758/s13414-015-0976-4.

Herd, S. A., O'Reilly, T. E., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia, 62,* 375–389, https://doi.org/10.1016/j.neuropsychologia.2014.04.014.

Heuer, A., Crawford, J. D., & Schubö, A. (2017). Action relevance induces an attentional weighting of representations in visual working memory. *Memory & Cognition, 45,* 413–427, https://doi.org/10.3758/s13421-016-0670-3.

Heuer, A., & Schubö, A. (2016). The focus of attention in visual working memory: Protection of focused representations and its individual variation. *PLoS One, 11,* 1–19, https://doi.org/10.1371/journal.pone.0154228.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791–804, https://doi.org/10.1016/S0896-6273(02)01091-7.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113–136, https://doi.org/10.1037/0096-1523.28.1.113.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*(2):24, 1–27, https://doi.org/10.1167/10.2.24. [PubMed] [Article]

Infanti, E., Hickey, C., & Turatto, M. (2015). Reward associations impact both iconic and visual working memory. *Vision Research, 107,* 22–29, https://doi.org/10.1016/j.visres.2014.11.008.

Kalogeropoulou, Z., Jagadeesh, A. V., Ohl, S., & Rolfs, M. (2017). Setting and changing feature priorities in visual short-term memory. *Psychonomic Bulletin & Review, 24,* 453–458, https://doi.org/10.3758/s13423-016-1094-6.

Kessler, Y. (2018). N-2 repetition leads to a cost within working memory and a benefit outside it. *Annals of the New York Academy of Sciences, 142,* 268–277, https://doi.org/10.1111/nyas.13607.

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 738–754, https://doi.org/10.1037/a0035545.

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin & Review, 22,* 1770–1776, https://doi.org/10.3758/s13423-015-0853-0.

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters, 461,* 145–149, https://doi.org/10.1016/j.neulet.2009.06.030.

Klink, P. C., Jeurissen, D., Theeuwes, J., Denys, D., & Roelfsema, P. R. (2017). Working memory accuracy for multiple targets is driven by reward expectation and stimulus contrast with different time-courses. *Scientific Reports, 7,* 9082, https://doi.org/10.1038/s41598-017-08608-4.

Knops, A., Piazza, M., Sengupta, R., Eger, E., & Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. *The Journal of Neuroscience, 34,* 9857–9866, https://doi.org/10.1523/JNEUROSCI.2758-13.2014.

Kondo, A., & Saiki, J. (2012). Feature-specific encoding flexibility in visual working memory. *PLoS One*, *7*, 1–8, https://doi.org/10.1371/journal.pone.0050962.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*, 1551–1556, https://doi.org/10.1177/0956797610385359.

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 407–413). Austin, TX: Cognitive Science Society.

Kriete, T., Noelle, D. C., Cohen, J. D., & O'Reilly, R. C. (2013). Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proceedings of the National Academy of Sciences, USA*, *110*, 16390–16395, https://doi.org/10.1073/pnas.1303547110.

Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*, 204–220, https://doi.org/10.1080/17588921003731586.

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

LaRocque, J. J., Eichenbaum, A. S., Starrett, M. J., Rose, N. S., Emrich, S. M., & Postle, B. R. (2015). The short- and long-term fates of memory items retained outside the focus of attention. *Memory & Cognition*, *43*, 453–468, https://doi.org/10.3758/s13421-014-0486-y.

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, https://doi.org/10.1162/jocn_a_00305.

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*, 5, https://doi.org/10.3389/fnhum.2014.00005.

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

Li, Q., & Saiki, J. (2014). The effects of sequential attention shifts within visual working memory. *Frontiers in Psychology*, *5*, 965, https://doi.org/10.3389/fpsyg.2014.00965.

Li, Q., & Saiki, J. (2015). Different effects of color-based and location-based selection on visual working memory. *Attention, Perception, & Psychophysics*, *77*, 450–463, https://doi.org/10.3758/s13414-014-0775-3.

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R., & D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. *The Journal of Neuroscience*, *38*, 5267–5276, https://doi.org/10.1523/JNEUROSCI.3061-17.2018.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281, https://doi.org/10.1038/36846.

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, https://doi.org/10.1016/j.tics.2013.06.006.

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

Machizawa, M. G., Goh, C. C., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *23*, 554–559, https://doi.org/10.1177/0956797611431988.

Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychonomic Bulletin & Review*, *19*, 470–476, https://doi.org/10.3758/s13423-012-0235-9.

Makovski, T., & Pertzov, Y. (2015). Attention and memory protection: Interactions between retrospective attention cueing and interference. *The Quarterly Journal of Experimental Psychology*, *68*, 1735–1743, https://doi.org/10.1080/17470218.2015.1049623.

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, https://doi.org/10.1037/0278-7393.34.2.369.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman.

Mazza, V., & Caramazza, A. (2015) Multiple object individuation and subitizing in enumeration: A view from electrophysiology. *Frontiers in Human*

Neuroscience, 9, 162, https://doi.org/10.3389/fnhum.2015.00162.

Melcher, D. (2001, July 26). Persistence of visual memory for scenes. *Nature, 412*, 401, https://doi.org/10.1038/35086646.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1): 2, 8–17, https://doi.org/10.1167/6.1.2. [PubMed] [Article]

Melcher, D., & Colby, C. L. (2008). Trans-saccadic perception. *Trends in Cognitive Sciences, 12*, 466–473, https://doi.org/10.1016/j.tics.2008.09.003.

Melcher, D. & Piazza, M. (2011) The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS One, 6*, e29296, https://doi.org/10.1371/journal.pone.0029296.

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, https://doi.org/10.1177/0956797612457782.

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews, 36*, 1640–1645, https://doi.org/10.1016/j.neubiorev.2012.03.001.

Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: Evidence for a 3-state model of memory. *NeuroImage, 54*, 1540–1548, https://doi.org/10.1016/j.neuroimage.2010.09.002.

Nosofsky, R. M., & Donkin, C. (2016). Response-time evidence for mixed memory states in a sequential-presentation change-detection task. *Cognitive Psychology, 84*, 31–62, https://doi.org/10.1016/j.cogpsych.2015.11.001.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411–421, https://doi.org/10.1037/0278-7393.28.3.411.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 51, pp. 45–100). San Diego, CA: Academic Press.

Oberauer, K. (2013). The focus of attention in working memory—from metaphors to mechanisms. *Frontiers in Human Neuroscience, 7*, 1–16, https://doi.org/10.3389/fnhum.2013.00673.

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). 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, 43*, 1–22, https://doi.org/10.1037/xlm0000302.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition, 41*, 1212–1227, https://doi.org/10.3758/s13421-013-0333-6.

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, https://doi.org/10.1037/a0026337.

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

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

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, https://doi.org/10.3758/s13423-012-0272-4.

Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review, 124*, 21–59, https://doi.org/10.1037/rev0000044.

Olivers, C. N., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review, 115*, 836–863, https://doi.org/10.1037/a0013395.

Olivers, C. N., & Watson, D. G. (2008). Subitizing requires attention. *Visual Cognition, 16*, 439–462, https://doi.org/10.1080/13506280701825861.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics, 64*, 1055–1067, https://doi.org/10.3758/BF03194756.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences, USA, 102*, 8776–8780, https://doi.org/10.1073/pnas.0500810102.

O'Reilly, R. C. (2006, October 6). Biologically based computational models of high-level cognition. *Science, 314*, 91–94, https://doi.org/10.1126/science.1127242.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375– 411). New York: Cambridge University Press.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744, https://doi.org/10.1038/89532.

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1224–1231, https://doi.org/10.1037/a0030947.

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

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*, 147–153, https://doi.org/10.1016/J.Cognition.2011.05.007.

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

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*, 23–38, https://doi.org/10.1016/j.neuroscience.2005.06.005.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, *21*, 156–161, https://doi.org/10.3758/BF03202727.

Rac-Lubashevsky, R., & Kessler, Y. (2016a). Decomposing the n-back task: An individual differences study using the reference-back paradigm. *Neuropsychologia*, *90*, 190–199, https://doi.org/10.1016/j.neuropsychologia.2016.07.013.

Rac-Lubashevsky, R., & Kessler, Y. (2016b). Dissociating working memory updating and automatic updating: The reference-back paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 951–969, https://doi.org/10.1037/xlm0000219.

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

Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory & Cognition*, *42*, 712–728, https://doi.org/10.3758/s13421-013-0392-8.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences, USA*, *105*, 5975–5979, https://doi.org/10.1073/pnas.0711295105.

Schendan, H. E., & Stern, C. E. (2008). Where vision meets memory: Prefrontal–posterior networks for visual object constancy during categorization and recognition. *Cerebral Cortex*, *18*, 1695–1711, https://doi.org/10.1093/cercor/bhm197.

Schneider, D., Mertes, C., & Wascher, E. (2016). The time course of visuo-spatial working memory updating revealed by a retro-cuing paradigm. *Scientific Reports*, *6*, 1–12, https://doi.org/10.1038/srep21442.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. B. (2018). Psychological scaling reveals a single parameter framework for visual working memory. *BioRxiv*, 325472, https://doi.org/10.1101/325472.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*, 38–45, https://doi.org/10.1016/j.tics.2005.11.008.

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

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS One*, *3*, 1–9, https://doi.org/10.1371/journal.pone.0001699.

Sligte, I. G., Wokke, M. E., Tesselaar, J. P., Scholte, H. S., & Lamme, V. A. (2011). Magnetic stimulation of the dorsolateral prefrontal cortex dissociates fragile visual short-term memory from visual working memory. *Neuropsychologia*, *49*, 1578–1588, https://doi.org/10.1016/j.neuropsychologia.2010.12.010.

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*, *7*, 1839–1860, https://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, https://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, https://doi.org/10.1037/a0036331.

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, https://doi.org/10.1037/xhp0000192.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276–315, https://doi.org/10.1016/0001-6918(69)90055-9.

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, https://doi.org/10.3758/s13414-014-0633-3.

Taylor, R., Thomson, H., Sutton, D., & Donkin, C. (2017). Does working memory have a single capacity limit? *Journal of Memory and Language, 93,* 67–81, https://doi.org/10.1016/j.jml.2016.09.004.

Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica, 132,* 106–114, https://doi.org/10.1016/j.actpsy.2009.01.005.

Thibault, L., Van den Berg, R., Cavanagh, P., & Sergent, C. (2016). Retrospective attention gates discrete conscious access to past sensory stimuli. *PLoS One, 11,* e0148504, https://doi.org/10.1371/journal.pone.0148504.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564, https://doi.org/10.1037/0096-3445.134.4.552.

Tzelgov, J. (1997). Specifying the relations between automaticity and consciousness: A theoretical note. *Consciousness and Cognition, 6,* 441–451, https://doi.org/10.1006/ccog.1997.0303.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149, https://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, 77,* 2117–2135, https://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. *Proceedings of the National Academy of Sciences, USA, 109,* 8780–8785, https://doi.org/10.1073/pnas.1117465109.

van Moorselaar, D., Gunseli, E., Theeuwes, J., & Olivers, C. N. L. (2015). The time course of protecting a visual memory representation from perceptual interference. *Frontiers in Human Neuroscience, 8,* 1–9, https://doi.org/10.3389/fnhum.2014.01053.

van Moorselaar, D., Olivers, C. N., Theeuwes, J., Lamme, V. A., & Sligte, I. G. (2015). Forgotten but not gone: Retro-cue costs and benefits in a double-cueing paradigm suggest multiple states in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1755–1763, https://doi.org/10.1037/xlm0000124.

Vandenbroucke, A. R., Sligte, I. G., de Vries, J. G., Cohen, M. X., & Lamme, V. A. (2015). Neural correlates of visual short-term memory dissociate between fragile and working memory representations. *Journal of Cognitive Neuroscience, 27,* 2477–2490, https://doi.org/10.1162/jocn_a_00870.

Vandenbroucke, A. R., Sligte, I. G., & Lamme, V. A. (2011). Manipulations of attention dissociate fragile visual short-term memory from visual working memory. *Neuropsychologia, 49,* 1559–1568, https://doi.org/10.1016/j.neuropsychologia.2010.12.044.

Vergauwe, E., & Cowan, N. (2015). Working memory units are all in your head: Factors that influence whether features or objects are the favored units. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1404–1416, https://doi.org/10.1037/xlm0000108.

Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS One, 3,* e3269, https://doi.org/10.1371/journal.pone.0003269.

Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research, 44,* 541–556, https://doi.org/10.1016/j.visres.2003.11.001.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005, November 24). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438,* 500–503, https://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, https://doi.org/10.1037/0096-1523.27.1.92.

Wallis, G., Stokes, M., Cousijn, H., Woolrich, M., & Nobre, A. C. (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. *Journal of Cognitive Neuro-science*, *27*, 2019–2034, https://doi.org/10.1162/jocn_a_00838.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64, https://doi.org/10.1037/0096-3445.131.1.48.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12):11, 1120–1135, https://doi.org/10.1167/4.12.11. [PubMed] [Article]

Xu, X., & Liu, C. (2008). Can subitizing survive the attentional blink? An ERP study. *Neuroscience Letters*, *440*, 140–144, https://doi.org/10.1016/j.neulet.2008.05.063.

Xu, Y., & Chun, M. M. (2006, March 2). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95, https://doi.org/10.1038/nature04262.

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*, *1339*, 154–164, https://doi.org/10.1111/nyas.12606.

Zhang, W., & Luck, S. J. (2008, May 8). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235, https://doi.org/10.1038/nature06860.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*, 1434–1441, https://doi.org/10.1177/0956797611417006.

Zokaei, N., Ning, S., Manohar, S., Feredoes, E., & Husain, M. (2014). Flexibility of representational states in working memory. *Frontiers in Human Neuroscience*, *8*, 1–12, https://doi.org/10.3389/fnhum.2014.00853.