TUTORIAL REVIEW

CrossMark

Visual memory, the long and the short of it: A review of visual working memory and long-term memory

Mark W. Schurgin[1]

Abstract

The majority of research on visual memory has taken a compartmentalized approach, focusing exclusively on memory over shorter or longer durations, that is, visual working memory (VWM) or visual episodic long-term memory (VLTM), respectively. This tutorial provides a review spanning the two areas, with readers in mind who may only be familiar with one or the other. The review is divided into six sections. It starts by distinguishing VWM and VLTM from one another, in terms of how they are generally defined and their relative functions. This is followed by a review of the major theories and methods guiding VLTM and VWM research. The final section is devoted toward identifying points of overlap and distinction across the two literatures to provide a synthesis that will inform future research in both fields. By more intimately relating methods and theories from VWM and VLTM to one another, new advances can be made that may shed light on the kinds of representational content and structure supporting human visual memory.

Keywords Visual working memory · Memory: Long-term memory · Memory: Visual working and short-term memory

Introduction

Human visual memory is impressive and fundamental to cognition and remains one of the core areas of human cognition that artificial intelligence (AI) systems have yet to be able to replicate (Andreopoulos & Tsotsos, 2013; DiCarlo, Zoccolan, & Rust, 2012; Pinto, Cox, & DiCarlo, 2008). We can recognize past objects we have seen when given incredibly brief exposures (Biederman & Ju, 1988; Potter, 1976; Potter & Levy, 1969), despite drastic image-level differences across encounters (through variability in factors such as orientation and lighting: Cox & Dicarlo, 2008; Cox, Meier, Oertelt, & DiCarlo, 2005; Dicarlo & Cox, 2007; Rust & Stocker, 2010; Wallis & Bülthoff, 2001), and even when we have not seen an object for months, years, or even decades (i.e., many people can distinctly remember what subway tokens or an Atari game system look like). The exact computational and neural explanations of this ability remain unknown and are likely key to

understanding disorders that impair visual memory and recognition, such as dementia.

The majority of research on visual memory has taken a compartmentalized approach, focusing exclusively on memory over shorter or longer durations, that is, on working or episodic long-term memory, respectively. The purpose of the present review is simply to supply a primer that spans the two areas, with readers in mind who may only be familiar with one or the other. Through this survey, I aim to identify points of overlap and distinction across the two literatures to provide a synthesis that will inform future research investigating the major questions remaining on human visual memory.

Distinguishing visual working memory and visual episodic long-term memory

Memories for visual information are typically distinguished as belonging to either visual working memory (VWM) or visual episodic long-term memory (VLTM). VWM is considered an online system that retains and manipulates information over the short term (Baddeley & Hitch, 1974; Cowan, 2008; Ma, Husain, & Bays, 2014; Vogel, Woodman, & Luck, 2006), whereas VLTM is typically defined as the passive storage of visual information over longer periods of time (Brady,

✉ Mark W. Schurgin
mschurgin@ucsd.edu

[1] Department of Psychology, University of California, San Diego, McGill Hall 5217, 9500 Gilman Drive #0109, La Jolla, CA 92093, USA

Konkle, & Alvarez, 2011; Cowan, 2008; Squire, 2004). Further distinguishing these two systems is that they appear to rely on distinct neural substrates. For example, VWM tends to be associated with activity in the occipital and parietal cortex (Harrison & Tong, 2009; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Todd & Marois, 2004), whereas VLTM is associated with the use of the medial temporal lobe and hippocampus (Squire, 2004; Yassa et al., 2011; Yonelinas, Aly, Wang, & Koen, 2010).

Broadly speaking, it is intuitive to distinguish different memory systems, especially VWM and VLTM, by the time scale over which the memory takes place. However, while time may constrain systems in different ways, it may not be sufficient to understand all possible differences and similarities between them. A more holistic approach may be to distinguish these systems in terms of their functions. If we know the function of a system, we can clearly identify the specific challenges facing such functions, even if we do not know the specific solutions to these challenges a system may choose to employ.

The function of VWM

When establishing the function of VWM, researchers early on defined it as a space divided between storage and other processing demands (Baddeley & Hitch, 1974). It is thought to be the combination of multiple processes, providing an interface between perception, short-term memory, and other mechanisms such as attention (Cowan, 2008). VWM's most notable characteristic is that it has a limited capacity (Awh, Barton, & Vogel, 2007; Vogel, Woodman, & Luck, 2001), and it is thought to be a core cognitive process underlying a wide range of behaviors (Baddeley, 2003; Ma et al., 2014). Given these descriptions, the function of VWM may be best described as supporting complex cognitive behaviors that require temporarily storing and manipulating information in order to produce actions.

VWM representations must therefore be in a format that is easily amendable and malleable, while remaining general enough to support a wide range of behaviors: it needs to output representations that can be taken as input by relevant processes. For example, an important role of VWM is to bridge gaps in our perception (both spatially and temporally) created by eye movements. Indeed, VWM has been shown to store features of saccade targets for subsequent comparisons (Hollingworth, Richard, & Luck, 2008), and the contents of VWM influence even simple saccades (Hollingworth, Matsukura, & Luck, 2013). In order to act as this bridge, representations in VWM must be content neutral (i.e., not specific to any one feature or class of objects) and also malleable (in order to integrate and compare information across time).

The function of VLTM

VLTM is typically described in terms of episodic recognition, which is commonly defined as the conscious recognition of visual events (Brady et al., 2011; Squire, 2004). This is a generally vague description that does not elucidate the function of VLTM. However, we can find insight into the function of VLTM from the study of object recognition. The goal of any object recognition system is to identify an object based on some previous input. Clearly, when VLTM studies utilize images of real-world objects in their tasks, the situation is analogous to an object recognition task. When referring to object recognition here, I mean so explicitly in the context of episodic long-term memories for objects, as opposed to other kinds of learning utilized by long-term memory, such as semantic or categorical information about those objects.

The fundamental challenge facing object recognition is the need for the system to demonstrate both tolerance and discrimination.[1] When we see an object again, we never encounter that same object exactly the same way twice. We see things briefly, in different contexts, with differences in lighting, viewpoint, or orientation that make direct comparisons of the perceptual input to our long-term memory representations next to impossible (Cox & Dicarlo, 2008; Cox et al., 2005; Dicarlo & Cox, 2007; Rust & Stocker, 2010; Wallis & Bulthoff, 1999). In order to recognize objects over the long term, our representations need to demonstrate tolerance. By tolerance, I am referring to our ability to recognize the same object across different encounters.

When perceiving or recognizing objects over the short term, your visual system has multiple cues to assist in acquiring tolerance. For example, if you toss a ball and watch it move through the air, the image hitting your retina transforms considerably over the ball's journey. However, despite drastic changes in the appearance of the ball from moment to moment, you are still able to recognize that the ball is the same at each point in time. This is because your visual system can exploit expectations of spatiotemporal continuity (where an object is in space and time) and as a consequence can demonstrate tolerance to these changes (Kahneman, Treisman, & Gibbs, 1992; Scholl & Flombaum, 2010; Schurgin & Flombaum, 2018). However, over the long term, spatiotemporal continuity and other cues used to facilitate tolerance are unavailable, and thus tolerance must be reflected within long-term memory representations themselves.

While tolerance allows us to identify previous experiences despite considerable changes across encounters, there is a potential danger of overtolerance. One could imagine that if a

---

[1] It is important to note that discrimination is sometimes referred to as "explicitness," and tolerance is sometimes referred to as "invariance" in the object recognition literature. For the purposes of this review, the terms are interchangeable.

representation is too tolerant, then it risks an observer mistaking every object she encounters as one that she's seen before. To address this problem, our representations must also demonstrate a certain level of discrimination. In contrast to tolerance, discrimination refers to our ability to distinguish similar but distinct inputs from one another.

For example, imagine one day my coworker and I are having lunch in a break room and we each have an apple. We place our lunches beside each other and then both exit the room to grab some napkins. When we return, there are two apples on the table that are quite similar in appearance, but one apple is mine and the other is my coworker's. In order to know which is mine, I must have a representation that is discriminating. Despite the similarities between both apples, I should be able to identify which apple belongs to me.

Similar to issues created by overtolerance, if a memory system is too discriminating then it would operate poorly since an observer would likely fail to recognize any previously seen input unless it was exactly the same as when it was first encountered. Given the variability in our visual inputs described earlier (e.g., due to differences in size, lighting, orientation, etc.), as well as the general assumption that our visual experience is noisy, this is simply an impossible threshold to meet. As a result, there is a natural tension that VLTM must manage between tolerance and discrimination. This is the function of VLTM—to manage this tension in order to recognize a previous visual experience given new input.

Core concepts of VLTM

As discussed previously, VLTM is typically defined as the passive storage system of visual episodic memory. The function of this system is to identify a past visual experience given some previous input, while balancing the need for both tolerance and discrimination. Having defined VLTM and identified its functions, we can expand our understanding of VLTM by establishing what core concepts researchers have identified in the field.

Familiarity and recollection

When trying to understand the nature of VLTM, researchers have focused on investigating the *kinds* of information VLTM may utilize. One particularly influential distinction in VLTM (and episodic long-term memory research generally) has been recollection and familiarity. Recollection refers to an observer accessing specific details about a previously experienced item. For example, let us say you happen to have gone to a friend's house for a party the other night and met a variety of new people. In particular, you happened to meet a man named Bob, with whom you talked at length about the Chicago

Cubs. The next day, you enter a coffee shop and as you get in line you happen to see that Bob is in front of you. Despite seeing Bob after a long delay, and in a completely new context, you remember all the specific details associated with him—his funky hairstyle, the color of his eyes, along with all the details of when you last met (i.e., your conversation about the Cubs, etc.).

In contrast to recollection, familiarity refers to an observer knowing an item is old or new, without having specific details associated with that memory—"familiarity is the process of recognizing an item on the basis of its perceived memory strength but without retrieval of any specific details about the study episode" (Diana, Yonelinas, & Ranganath, 2007, p. 379). This notion of familiarity is quite common, and perhaps something we have all experienced. Given the example above, consider an alternate universe where you see Bob the next day in the coffee shop. While you recognize him as someone you know, you cannot remember any of the specific details related to this knowledge, including where you met him or what context you know him from. This is familiarity.

Dual-process signal-detection model One model used to explain recollection and familiarity processes is the dual-process signal-detection (DPSD) model (Yonelinas et al., 2010). The DPSD model specifies that familiarity is primarily a signal-detection process of discriminating between two Gaussian distributions of memory-match strength that differ on average between old and new items, where familiarity occurs when a signal exceeds a decision criterion (see Fig. 1). Therefore, your ability to be familiar with something you have seen before depends on the strength of the memory signal and an individual's decision criterion (how liberal or conservative that individual is being with their decisions). Experiences that are very different (i.e., new) should produce very little to no overlap, so familiarity should not occur. But if given visual inputs that are highly similar (even if you have not necessarily encountered them before), this will likely result in a familiarity signal.

The DPSD model defines recollection as a threshold-based process. In order to recollect an item, an observer must collect enough information. Once enough information is collected, it has passed the "threshold," and specific details will then be recollected. These classifications are exclusive and are an "all-or-nothing" process. It does not matter if an observer collects more information as long as the threshold has been passed. However, if an observer does not accumulate enough information, the threshold is not passed and recollection will not occur. Figure 1 schematizes this process. For example, let us imagine you have a memory associated with your favorite coffee mug, which has a certain recollection threshold. If you see your coffee mug clearly on your desk, then you receive a strong memory signal (Stimulus A), so the threshold is passed and your memory is recollected. The next day you see the coffee mug on your desk, but it is in a different place and is
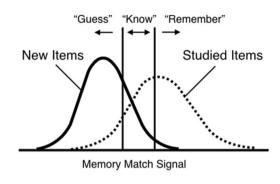
DPSD Model

CDP Model

Fig. 1 Illustration of the dual-process signal-detection (DPSD) and continuous dual-process (CDP) models. In the DPSD model, familiarity and recollection are two distinct but parallel processes. Familiarity is a signal-detection process of discriminating between two Gaussian distributions between old and new items, where familiarity occurs when the signal exceeds a decision criterion (i.e., participants "know" they saw the item). Recollection is a threshold-based process where signal strength passes a certain threshold and is either recollected or not. In the diagram, both Stimuli A and B pass the threshold and would thus be recollected with the same amount of detail, regardless that each stimulus may illicit

different amounts of memory-match signal strength. Stimuli C does not pass the threshold and would thus not be recollected. In the CDP model, both familiarity and recollection vary continuously and operate using signal-detection-based processes. These processes are interactive and are combined during decision-making, resulting in a single distribution for studied items. In the simplest version of the model, familiarity occurs when the memory-match signal exceeds a lower decision criterion (i.e., "know"), whereas recollection occurs when the signal exceeds a higher decision criterion (i.e., participants "remember" the details of the item)

partially occluded (Stimulus B). While the perceptual information and signal provided is not as strong, you still gain enough information that it passes the threshold and your memory is recollected with the same amount of detail as the day before. However, later in the week, you see your coffee mug from very far away (Stimulus C). While there is some overlapping perceptual information here and you do receive some memory signal, it is not enough to pass your threshold and so you do not recollect the memory.

It is important to note that the example described above is provided to illustrate how a threshold process might operate (specifically within one potential visual context). While sensory overlap may drive recollection automatically in some cases, it can operate in many other circumstances as well. Indeed, in many studies of recollection, subjects are instructed to attempt recall, with varying levels of success. In these studies, a subject's success or failure to recall information may depend on higher or lower thresholds relative to the memories being recalled.

The DPSD model is based on the claim that familiarity and recollection are two distinct but parallel processes. Despite their distinction from one another, humans likely utilize both processes in a variety of long-term memory tests and behaviors. Familiarity or recollection would both be sufficient to recognize something you have seen before. For example, in order to recognize someone you have previously met, you

could rely on familiarity (i.e., a feeling that you have seen that person before) or recollection (i.e., specific details related to that person). Along similar lines, many common long-term memory tests ask observers to classify images as ones they saw during encoding (*old*) or as completely new images. An observer could use either familiarity or recollection to correctly classify an image as *old*.

Continuous dual-process model Numerous studies exist in support of the DPSD, extending past humans (Bowles et al., 2010; Yonelinas et al., 2010) to research involving rodents (Fortin, Wright, & Eichenbaum, 2004) and monkeys (Miyamoto et al., 2014). However, this is not the only computational approach to understanding the possible distinction between familiarity and recollection. For example, Dede, Squire, and Wixted (2014) suggested these processes are better described computationally by the continuous dual-process (CDP) model. In this model, familiarity operates exactly as defined in the DPSD model. However, unlike the DPSD, the CDP model proposes that recollection can vary continuously using the same signal-detection process as familiarity. This means recollection is not an all-or-nothing process, and that different memories may illicit different memory-match signals that will vary in detail. Furthermore, during memory decision-making, familiarity and recollection are combined and are thus interactive. Since both familiarity and recollection can vary in

confidence and accuracy, observers can combine both sources of information to inform subsequent performance. As a result, the CDP model has only a single distribution for studied items, reflecting variable information from both familiarity and recollection. In the simplest version of this model, familiarity occurs when the memory-match signal exceeds a lower decision criterion (i.e., participants "know" they saw an item), whereas recollection occurs when the signal exceeds a higher decision criterion (i.e., participants "remember" specific details of the item; Dunn, 2004; Wixted & Stretch, 2004; see Fig. 1).

In contrast, the DPSD model assumes these two processes are independent but operate in parallel. As a result, they should not interact. For example, if recollection is successful, then an observer would not use any familiarity information at test, as recollection would provide all the evidence needed to make a decision. However, if recollection failed, then an observer would rely solely on the strength of the familiarity signal.

Both the CDP and DPSD models have similar predictions in terms of performance in recognition tasks, but vary according to the types of systematic errors they would produce. By modeling systematic errors, Dede et al. (2014) found support that human data was better fit by the CDP model than a DPSD equivalent. This demonstrates that the CDP model is consistent with previous research in support of the DPSD model, but may more accurately capture the specific computations used in recognition decision-making. It also alters the previous distinctions between familiarity and recollection processes by suggesting an interactive relationship, rather than one that is orthogonal.

Potential limitations It is unclear if researchers intend for the familiarity/recollection dichotomy to designate different types of representations, or different affordances from a single VLTM representation. In general, a feature of these models that leaves them difficult to interpret is that they are content neutral; the same models are applied in the same way regardless of what is to be remembered, and without any claims about how those things are described in symbolic or activation terms.

For example, it is possible that familiarity and recollection each reflect memories with different contents, something like a gist (familiarity) and a more detailed representation (recollection). Yonelinas et al. (2010) appear to endorse this kind of view, as they argue that recollection and familiarity are separate processes that make independent contributions to recognition memory. This is an intuitive conclusion under the DPSD model, as it specifies familiarity is a signal-detection-based process, whereas recollection is a threshold-based process, and thus likely differ in their content and format. This would suggest that individuals are able to make accurate familiarity-based judgments even when detailed representations fail to consolidate.

But under this view, we should then want specific accounts of the contents and formats in each representation. What goes

into a gist? One possibility is that it is simply categorical or semantic knowledge for something you have seen before. Such knowledge would be able to interact with more detailed representations, allowing for observers to utilize both types of information when making memory judgments. However, defining gist in this way seems problematic, since almost any information related to a category should therefore create some familiarity signal.

By defining recollection and familiarity as exclusive processes, this also creates potential limitations in the kinds of information gist and more detailed representations can provide. Can an observer have familiarity with some aspects of an object and recollect others? Given a DPSD framework, this is not possible. Familiarity is recognition without any specific details, and recollection is an all-or-none process. What happens then when an observer confuses a previously seen object with a similar-looking item? Does this mean that they incorrectly recollected that object? Or does it mean that they relied only on familiarity to make their judgment? It is not clear in the DPSD model what information an observer may be relying on when making such errors.

In contrast, the CDP model would be able to account for such differences, as recollection can vary continuously (and is thus not all-or-none) and is also integrated into familiarity-based information when making a decision. Given that both familiarity and recollection can vary continuously, this creates more flexibility for different kinds of memory performance. It is possible for observers to have a strong familiarity signal (and more confidently report they "know" they saw an item), or other experiences where recollection occurred but the signal was weak, resulting in an observer reporting they "know" they saw an item rather than they "remember" all the details related to that item (Ingram, Mickes, & Wixted, 2012). For example, an observer could recollect varying amounts of specific details related to an object, which is not possible under the DPSD model.

However, the CDP model has its own limitations. The CDP provides a framework to explain the nature of memory signals but does not explain their underlying representations. The information and structure contained in memory remains unspecified (both in the CDP and DPSD). Again, it would be useful to specify the contents and formats of the representation in some detail to understand how it produces the response patterns taken to indicate familiarity and recollection.

Pattern separation and pattern completion

Another way researchers have tried to understand the nature of VLTM is to take an approach informed by neuroanatomy. The idea is that certain brain areas related to memory have specific properties that may make them amenable to completing different types of processes. Pattern completion and pattern

separation is one such model seeking to explain distinctions in memory, but is also supported by brain localization properties.

Pattern completion is the process by which incomplete or degraded signals are filled in based on previously stored representations (Yassa & Stark, 2011). After you see an object, the next time you encounter that object the image hitting your retina is not going to be exactly the same (due to differences in viewpoint, orientation, lighting, etc.). Pattern completion is the process that would assist in VLTM being tolerant to variability in inputs related to the same object across encounters.

Pattern separation is the process that reduces the overlap between similar inputs in order to reduce interference at later recall (Yassa & Stark, 2011; see Fig. 2). This process supports the discriminatory ability of VLTM. It seeks to parse overlapping signals into distinct representations in order to assist in the individual recall of both items.

This model was in part designed to address the issue that we want to be able to remember certain events as related but other events as distinct (i.e., tolerance and discrimination). For example, even when an observer is given different stimuli, considerable overlap might exist between the inputs into the system, causing different items to be incorrectly identified as related to one another (i.e., the problem of saturation). One possible way to address this issue would be to make the input representation very sparse, with a large and distributed network of neurons creating multiple potential pathways of activation for inputs. The hippocampus, which previous examples have shown is critical to episodic and recognition memory, has a region with such properties—the dentate gyrus (DG). This is consistent with pattern separation, which seeks to separate overlapping representations as distinct from one another.
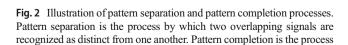
Another issue this model sought to address is to explain how we remember certain events as being related to one another. Specifically, what is the process that recognizes two partially overlapping signals as arising from the same source? One possible solution would be to have a set of

pathways that feeds back into itself to fill in potential missing information, so that partial or incomplete signals can be recognized. Another area of the hippocampus, CA3, has such properties, through recurrent collaterals that form a feedback loop (i.e., the axons of neurons in the circuit circle back to the inputs [dendrites] of neighboring axons), consistent with pattern completion computations (Yassa & Stark, 2011).

Even though there can be many different computations in support of distinctions in memory, the framework of pattern separation and pattern completion is made more viable by taking into account the specific properties of brain areas that may implement such processes. In fact, recent evidence using rats in a spatial maze has shown that the CA3 demonstrated coherent responses during conflict (when the global cues of the maze were rotated but local information remained the same), consistent with pattern completion, whereas the DG demonstrated disrupted responses during such conflict, consistent with pattern separation (Neunuebel & Knierim, 2014). While the contents of these processes are not specified, they provide a useful example for how a distinction supported by brain localization properties may inform our understanding of memory.
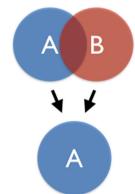
Potential limitations Unfortunately, many aspects of pattern separation and pattern completion remain unspecified. This results in a host of potential limitations. Often, this framework is referred to as a "neurocomputational" approach, but the computations of these processes are never defined. There are many possible ways one could design a system to parse or complete the same overlapping inputs that would result in vastly different outputs. In addition, the content and format of these inputs are not specified. This information would greatly affect the kinds of computations one might use to address these types of processes. While the specific representational content and format of pattern separation and completion remains elusive, researchers have begun using new approaches to further define these processes. For example,

Fig. 2 Illustration of pattern separation and pattern completion processes. Pattern separation is the process by which two overlapping signals are recognized as distinct from one another. Pattern completion is the process

by which two overlapping signals are recognized as being from the same source and are combined into a single representation

Schapiro, Turk-Browne, Botvinick, and Norman (2017) used hippocampal neural network modeling to make more specific assumptions for how pattern separation and completion may be instantiated. Although these assumptions remain primarily in neural terms, further specificity in the neural pathways and areas involved in these processes may help inform their potential representational content and format.

Pattern separation and pattern completion are also limited by their focus to explain memory-based processes exclusively on properties of the hippocampus. Pattern completion shares similarities with familiarity, as both processes could be used to explain how an observer might recognize a previously seen image with degraded input. But pattern completion is specified to occur in the CA3 region of the hippocampus. Thus, it is unable to explain why amnesic patients with brain damage to the hippocampus can still demonstrate unimpaired associative (or familiarity-based) memory performance (Gabrieli, Fleischman, Keane, Reminger, & Morrell, 1995).

Theoretical overlap and conclusion

Theories of familiarity and recollection bear some resemblance to pattern separation and completion. For example, familiarity refers to some sort of gist knowledge, when an observer knows they have seen something before but without specific details associated with that memory. And pattern completion is the process through which two overlapping signals are recognized as arising from the same source. One way the gist knowledge of familiarity could be expressed is through a mechanism such as pattern separation—recognizing overlap between signals to recognize an item as previously seen, without necessarily extracting specific details associated with a previous memory.

Along similar lines, recollection refers to an observer accessing specific details about a previously experienced item. This process is critical for our discriminatory ability to distinguish similar-looking items from ones we have seen before. Indeed, in order to assess recollection-based memory many times researchers give participants lures (very similar-looking items) at test and assess their ability to discriminate these lures. Pattern separation describes exactly the process through which a lure may be distinguished as being distinct from a previously studied item. Thus, while pattern separation has distinct functions and neural predictions in comparison to recollection, there is clearly overlap between these two theories of memory.

Regardless of the potential similarities between these models, and despite disagreements in the field as to which better explains VLTM, there is a critical consensus among these models that there are different *kinds* of information VLTM may rely on. This generally involves some sort of gist information associated with a previous item (familiarity, pattern completion), as well as more specific visual details related

to that same item (recollection, pattern separation). Both kinds of information likely affect decision-making in a variety of VLTM tasks. As a result, it is important to keep in mind that some tasks may rely on one process or kind of information more than another.

Common methods in VLTM

Old/similar/new judgment

One way to measure the discriminatory ability of VLTM is to utilize a paradigm often referred to as the old/similar/new judgment. Participants first view and encode images of objects, typically while doing an incidental encoding cover task (e.g., would this object fit in a shoe box?). At test, participants serially view objects that were present in the encoding task (old images), objects similar but not identical to ones in the encoding task (similar images), and completely new objects that were not present during encoding (new images). Participants then judge whether the images they see are old, similar, or new (see Fig. 3).

The primary purpose of this task is to evaluate the discriminatory ability of VLTM, and how memories may acquire different levels of discriminability. For example, correctly classifying a similar item likely requires an observer to have more specific details in memory (i.e., recollection-based) than correctly identifying an old item, which could be accomplished using a gist-based or familiarity-based process (Kensinger, Garoff-Eaton, & Schacter, 2006; Kim & Yassa, 2013; Schurgin & Flombaum, 2017; Schurgin, Reagh, Yassa, & Flombaum, 2013; Stark, Yassa, Lacy, & Stark, 2013).

A potential limitation of old/similar/new judgments is that it is not clear what constitutes a false alarm for similar items. In contrast, when using a simpler old/new procedure researchers can obtain an unbiased measure of memory discriminability (d') by taking the difference between the normalized proportion of hits (correctly classifying an old item is *old*) and the normalized proportion of false alarms (incorrectly classifying a new item as *old*; Green & Swets, 1966). However, what constitutes a false alarm for a similar judgment? Is it classifying an old item as similar? Or a new item as similar? Or a similar item as old? Without a clear understanding of what constitutes a false alarm, it is difficult to normalize responses given to similar items for potential biases. While different analyses have attempted to address this issue, research suggests analyzing old/similar/new responses using a signal-detection-based framework (d_a) may provide an accurate, unbiased measure of memory performance for similar items (see Loiotile & Courtney, 2015).

How does emotion affect visual memory? One study that utilized this method sought to investigate how negative emotional context may affect the likelihood of remembering an item's
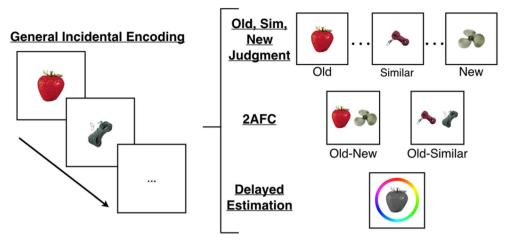
Fig. 3 Illustration of typical VLTM methods. In a general incidental encoding task, participants see a stream of images of real-world objects and make some judgment about those objects (i.e., indoor/outdoor, does this fit in a shoe box?). After, they may be tested using several methods. Using old/similar/new judgment, participants are shown objects that were exactly the same as encoding (old), similar but not identical to images at encoding (similar) and completely new images (new), and are asked to classify them accordingly. In two-alternative forced-choice tests, they are shown two images, one they have seen before and either a completely new (old–new comparison) or similar looking object (old–similar comparison) and are asked to judge which of the two images they have previously seen. In delayed-estimation tasks, participants are initially shown a grayscale image of a previously seen object and are asked to report its color using a color wheel. (Color figure online)

specific visual details. Participants first completed an incidental encoding task, where they were exposed to hundreds of images of real-world objects (for 250 ms or 500 ms) and had to judge whether each object would fit in a shoebox. Half of the images were rated as being negative and arousing, and the other objects were rated as neutral. Two days after incidental encoding, participants were then given a surprise test. Participants viewed old images (exactly the same as encoding), similar images (similar but not identical), and completely new images relative to encoding. They were told to classify them accordingly (old/similar/new; Kensinger et al., 2006).

At test, it was observed that for old items, negative emotional context led to an increase in correct classification. This was true for items presented both for 250 ms and 500 ms, but was stronger as encoding time increased. However, for similar items there was no main effect of emotional context or encoding time. As a result of the main effect of emotional context for old images, the researchers concluded that negatively arousing content increased the likelihood that visual details of an object would be remembered (Kensinger et al., 2006). However, given the lack of an effect for similar images, it would be more accurate to conclude that emotion may have enhanced certain aspects of visual memory (i.e., for old images).

How do familiarity and recollection contribute to responses?
This method has also been used to investigate the relative contributions that familiarity and recollection processes may have in the behavioral responses of old, similar, and new judgments. In the study, participants completed a two-stage recognition test. In the first phase, participants saw 128 images of real-world objects on a computer screen for 2 seconds each,

and were asked to report whether an object was an "indoor" or "outdoor" object. In the second phase, participants were given a surprise test where they viewed old images (exactly the same as encoding), similar images (similar but not identical), and completely new images. They were told to classify the images as either old, similar, or new. Additionally, after indicating what category an image belonged to, participants were then instructed to indicate whether they "remember" seeing the same image in the study session or if they just "know" that they have seen the same image without any conscious recollection of its original presentation (Kim & Yassa, 2013). It is theorized that "remember" judgments reflect recollection-based processes, whereas "know" judgments reflect familiarity-based processes, although there exist several criticisms that these are not true indices of these processes but rather reflect subjective states of awareness or differences in confidence (Yonelinas, 2002).

As expected, at test they found very different accuracy for classifying old (70% correct), similar (53% correct), and new (74% correct) images. When analyzing these responses according to whether a participant reported they "remembered" (recollection based) or "knew" (familiarity based), a few interesting patterns emerged. When judging old items correctly, observers made primarily "remember" responses, suggesting that correct classifications of old items was primarily driven by recollection. For similar items, there was a slight trend to report "remember" rather than "know" both when the item was judged as old or similar. This suggests that observers can classify similar items with or without recollection, and that incorrectly identifying similar items (i.e., misclassifying them as old) is not simply driven by familiarity signals (Kim & Yassa, 2013).

Source localization judgments

One approach to evaluate the strength of memory and further distinguish potential errors is to use a paradigm combining old/new judgments with source localization. At encoding, participants view images of objects typically presented in one of four quadrants in the display. Then, at test, participants are shown old, similar, and new images in the center of the screen they must classify as old (previously seen images) or new (similar or new images), and indicate which of the four quadrants the object originally appeared in (Cansino, Maquet, Dolan, & Rugg, 2002; Reagh & Yassa, 2014; see Fig. 4).

The underlying assumption of source localization manipulations is that more episodic information is retrieved on trials when the source judgment was successful than on trials when it was not. This is similar to the logic behind the remember/know procedure discussed previously. When an observer makes a correct classification and source judgment, the assumption is that this indicates a recollection-like memory, whereas if an observer makes a correct classification of an image but an incorrect source judgment, this may indicate a familiarity-based memory. However, unlike the remember/know procedure, source localization judgments do not rely on the observer's own introspection in order to classify memory quality. Thus, the aim of the paradigm is to evaluate what percentage of responses using old/new judgments may rely on memories that contain more or less information, and what brain areas may be involved in these processes.

What brain areas and behaviors are involved in source judgments?

Cansino et al. (2002) were interested in using this method to investigate what brain regions may be involved in different memory processes beyond simple item recognition tasks. To accomplish this, participants first viewed images of real-world objects and were asked to judge whether the objects were natural or artificial. Critically, each image was presented in one of four quadrants in the display. After completing the task, participants were then administered a surprise test where previously shown images (old images) were mixed with completely new images. These images were shown in the center of the screen. Participants had to judge whether each image was old or new. They were instructed to press a single key if an image was new, and if an image was old participants indicated which position the image was presented during encoding using one of four keys. If a participant did not know which quadrant an old image originated from, they were instructed to guess.

At test it was found that when classifying previously seen (old) items, observers correctly identified 87% of items presented. However, 60.7% of these responses contained correct source responses, and 26.3% of responses contained incorrect source responses. This suggests that even when classifying old and new objects in a typical retrieval task, memories for these items likely contain additional information beyond simply categorical or familiarity-based knowledge. Additionally, they observed via collected fMRI data that when recognizing an old object with a correct versus incorrect source judgment, there was greater activity observed in the right hippocampus and left prefrontal cortex (Cansino et al., 2002). This suggests that memories containing more information may elicit greater memory signals and decision-making coordination.

Potential neural correlates of "what" and "where" memory

Source localization judgments have also been used to explore possible dissociations between object (what) and spatial (where) memories and their potential neural correlates. In the experiment, participants first completed an encoding task where images of real-world objects were presented in one of
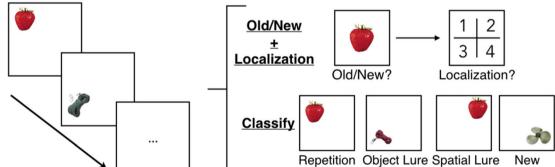

Fig. 4 Illustration of typical localization VLTM methods. In a general incidental encoding task, participants see a stream of images of real-world objects and make some judgment about those objects (i.e., Does this fit in a shoe box?). Critically, every image is presented in one of four (or more) possible quadrants. After, they may be tested using several methods. Using old/new and localization tasks, participants are shown objects that were exactly the same as encoding (old) and similar-looking or completely new images (new), and are asked to classify them accordingly. If they classify an object as old, they are then asked to indicate which of the four quadrants the image originally appeared in. In classification tasks, participants are shown either repetitions (old image, same location), object lures (similar images, same location), spatial lures (old image, different location), or new images. They are asked to classify the images accordingly

31 possible locations on the screen for 3 seconds each. They were instructed to first judge whether the object was an indoor or outdoor object, and then whether the object appeared on the left or right relative to the center of the screen. Afterward, participants were given a surprise test with four possible trial types: repeated images (old images in the same location), lure images (similar images in the original object's location), spatial lure images (old images in a slightly different location), or new images (not shown during encoding). Participants were instructed to indicate whether an image showed no change, object change, location change, or new (Reagh & Yassa, 2014).

Behaviorally, there was no difference in lure discrimination whether it was an object trial (i.e., similar image) or spatial trial (i.e., old image in slightly different location). This effect was consistent across both high-similarity and low-similarity stimuli. Neuroimaging data were also collected via fMRI and demonstrated unique differences based on lure type. It was observed that the lateral entorhinal cortex (LEC) was more engaged during object lure discrimination than during spatial lure discrimination, whereas the opposite pattern was observed in the medial entorhinal cortex (MEC). Additionally, the perirhinal cortex (PRC) was more active during correct rejections of object than spatial lures, whereas the parahippocampal cortex (PHC) was more active during correct rejections of spatial than object lures. Regardless of lure type, the dentate gyrus (DG) and subregion CA3 demonstrated greater activity during lure discrimination. Overall, this suggests two parallel but interacting networks in the hippocampus and related regions for managing object identity and spatial interference (Reagh & Yassa, 2014).

Two-alternative forced-choice test (2AFC)

A paradigm referred to as the two-alternative forced-choice (2AFC) test has been primarily used to study the capacity of visual episodic long-term memory. In a typical test, observers see two objects on the screen during test, one that they have seen before and another object they have not encountered previously. The other may be completely novel (old–new comparison) or a similarly related lure (old–similar comparison; Brady, Konkle, Alvarez, & Oliva, 2008; Brady, Konkle, Oliva, & Alvarez, 2009; Konkle, Brady, Alvarez, & Oliva, 2010b). The logic of this test is that it can tap into even "weak" memories that other methods may fail to reveal. It makes sense that a 2AFC judgment is easier than other kinds of responses because it is a binary response.

Typically, 2AFC is conceptualized using a signal-detection-theory framework (Green & Swets, 1966; Loiotile & Courtney, 2015). The logic is that observers have a memory representation that creates a normally distributed signal in a "memory strength" space. At test, when observers are shown an old item, that item elicits a normally distributed memory-match signal, which due to noise and other factors will vary in strength. If an observer was simply shown the old object, depending on the decision criteria this may result in incorrectly identifying an old object as *new*. However, by giving observers a foil in the 2AFC task, whether that foil is a completely new or similar-looking object, this gives observers a second normally distributed signal to assist in the comparison process. This second signal should be centered around a lower memory-match signal (i.e., zero) than the old image. As a result, observers can simply pick the max between the two items to correctly identify the old image (Macmillan & Creelman, 2004; see Fig. 5). This framework demonstrates from a modeling perspective why 2AFC tasks should be easier and are able to provide better performance for items that may otherwise fail to be remembered or classified correctly in other types of memory tasks—it is always easier to pick the max of two things.[2] Moreover, this performance is higher by a fixed amount, suggesting 2AFC taps into the same underlying memory signal as old/new testing procedures (Macmillan & Creelman, 2004).

In addition to traditional 2AFC tasks, which involve an old item paired with a new or similar-looking item, researchers have also expanded on the number of potential options, creating 3AFC and 4AFC-type tasks. Generally, these tasks involve the addition of multiple similar-looking lures at test, in order to further evaluate the ability to discriminate between different kinds of lures. A potential limitation of 3AFC or 4AFC tasks is that adding lures may create interference and increase task difficulty, such as through adding increased decision noise (Holdstock et al., 2002). Additionally, when varying the kinds of lures available at test (such as providing two similar lures, or a similar lure and a completely new image), this creates conditions where the information available to an observer is not equivalent (Guerin, Robbins, Gilmore, & Schacter, 2012). This means performance across different testing conditions cannot be directly compared with one another.

The capacity of VLTM Brady et al. (2008) sought to investigate the capacity of VLTM using a 2AFC method. Participants were presented 2,500 images of real-world objects for 3 seconds each and told to remember all the details of each image. After completing this study portion, participants were then given a 2AFC task where they saw two images on the screen. One was a previously encountered image from the previous session, whereas the other was either a novel image, an exemplar of an object they had previously encountered, or an image of an object they had previously encountered in a new state (i.e., changed orientation). Participants were instructed to indicate which of the two images they had previously encountered. Overall, performance was quite high, with significantly better accuracy for novel comparisons (92% correct), a replication of previous work by Standing (1973) showing

---

[2] Another common way of conceptualizing 2AFC performance is that participants compute a difference score between the two signals and base their decision on that variable (Macmillan & Creelman, 2004).
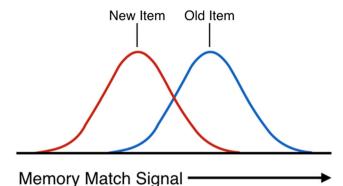
New Item   Old Item

Memory Match Signal ⟶

Fig. 5 Visualization of 2AFC logic. At test, observers are shown both an old and new image, which elicit their own normally distributed memory-match signals. While the old item signal may vary in the strength of its memory signal, it is more likely to be higher than the signal of the new item. By simply picking the item with the maximum signal, observers are likely to choose the old item. Thus, with the same underlying memory strength, 2AFC test procedures produce better memory performance compared to testing procedures where observers are only shown the old image

incredibly high performance for novel test comparisons even when encoding 10,000 items into VLTM (see also Shepard 1967). However, quite surprisingly they observed extremely accurate performance for state and exemplar comparisons (87%–88% correct; Brady et al., 2008). These results demonstrate that even when given very brief exposure of images, humans are able to remember thousands of objects (seemingly with no limit) with extremely high accuracy. Furthermore, they suggest that human's VLTM representations contain visual information necessary to assist in making difficult state and exemplar comparisons, beyond simply categorical or semantic knowledge of previous encounters.

Similar results have been found not just for objects but for the visual memory of scenes as well. Konkle et al. (2010a) demonstrated that after studying thousands of images of scenes, participants were able to recognize 96% of the previously seen images in a novel comparison test. Again, they also observed that performance for test comparisons with a similar foil was quite high, with participants 84% correct even when they studied four of the same exemplar (a potential source of interference) during encoding. This suggests that the incredibly large capacity and high-fidelity representations observed in visual long-term memory are not isolated to a specific stimulus class (i.e., objects or scenes), but rather appear to be general properties of the system.

Given the logic of 2AFC discussed previously, one could assume if participants had been given a single image of an object at test and told to discriminate whether it was old or new, performance would be worse (despite the same underlying memory strength). Thus, the results discussed above are best described as a potential upper bound of VLTM performance. Under different testing procedures, performance will likely differ. However, these results still demonstrate that

under potentially ideal testing conditions, visual long-term memory not only has a massive capacity but also contains representations with visual-rich and detailed information.

Delayed estimation (continuous report)

In order to estimate the fidelity of VLTM (i.e., the amount of information in memory) experiments have utilized a delayed estimation paradigm. At encoding participants observe objects embedded in unique colors. Then, at subsequent test, observers see grayscale versions of objects they observed previously and use a color wheel to indicate its original color. By taking the error in degrees between the response and the true value, researchers can create a distribution of long-term color memory responses and measure the standard deviation of the distribution to understand the fidelity of that representation (Brady, Konkle, Gill, Oliva, & Alvarez, 2013).

The precision of information in VLTM representations Brady et al. (2013) used this method to understand the precision of color memory representations across VWM and VLTM. In the study, researchers gave participants two separate tasks. In the VWM condition, participants saw three real-world objects simultaneously for 3 seconds, arranged in a circle around fixation. Participants were instructed to remember the color of all the objects. After a 1-second delay, one of the objects reappeared in grayscale, and participants could alter the color of the image using their mouse and were told to click the mouse when it matched the original color. In the VLTM condition, participants first underwent a study block viewing images sequentially for 1 second each, with a 1-second blank interval between images. Similar to the VWM condition, participants were instructed to remember the color of the object.

After the study block, the color of the items was tested one at a time in a randomly chosen sequence, and participants reported the image color using the same response mechanism used in the short-term memory condition. The precision of participants' memory representations was determined by calculating the distribution of the degree error of each response in color space (with larger mistakes represented by greater degree error). In the VWM condition, at Set Size 3 (and above), participants' precision was 17.8 degrees, which did not significantly differ from the precision observed in the VLTM condition, 19.3 degrees (Brady et al., 2013). Therefore, it appears the precision of color representations across VWM and VLTM have equivalent limits, suggesting they may share or be constrained by similar processes.

Interactions between VLTM and perception Delayed estimation has also been used to investigate the potential role long-term memories may have when biasing new perceptual information. In a series of experiments by Fan, Hutchinson, and Turk-Browne (2016), participants completed a set of initial

exposure trials, where they encountered unique shapes embedded in specific colors. Each shape was shown for half a second, and after a brief delay (1.5 seconds) an achromatic version of the shape reappeared, and participants reported the color of the image using their mouse. During the initial exposure session, participants encountered the same shape three times on separate trials (randomly assorted), always in the same color. As a result, each unique shape was associated with a specific color in long-term memory.

After completing the initial exposure trials, participants were then given final test trials. These final test trials were similar, except now each shape was shown in an unrelated color, and participants had to judge the appearance of the new colors. Researchers found that participants' responses for the final test trials were best characterized as a mixture of the original and current-color representations, suggesting participants had anchored their responses to their representations in long-term memory. Moreover, this anchoring effect increased when perceptual input became more degraded (for example, by shortening the stimulus presentation during the final test trials). These results demonstrate that while perceptual judgments do reflect the current state of the environment, they can be affected by previous experience and long-term memory (Fan et al., 2016).

Core concepts of VWM

As previously discussed, VWM is typically thought of as the interface of multiple processes including perception, short-term memory, and attention (Baddeley & Hitch, 1974; Cowan, 2008). Due to this conception, researchers have generally described the function of VWM as supporting complex cognitive behaviors that require temporarily storing and manipulating information in order to produce actions (Baddeley, 2003; Ma et al., 2014). In particular, a large body of research over the past decade has focused on the capacity limitations of VWM. As a result, many of the models proposed to explain VWM have focused on this limitation.

Fixed slot model

When trying to understand the inherent capacity limitations of VWM, a particularly influential model has been the fixed slot model. It suggests that VWM can only store a discrete number of integrated object representations (see Fig. 6). This model was proposed by a highly influential study conducted by Luck and Vogel (1997), who used a change-detection task to quantify VWM capacity. In the task, participants were instructed to remember an array consisting of items of a single or a conjunction of features (color, orientation, etc.). After a brief delay (900 ms), a test array was presented that was either identical to the previous array or differed in terms of a single feature. Participants were instructed to indicate whether a

change had occurred. Accuracy was assessed as a function of the number of items in the stimulus array in order to determine how many items could be accurately maintained in VWM.

In a series of experiments, Luck and Vogel (1997) gave participants change-detection tasks that varied the number of colored squares presented in an array (one to 12). They observed that performance was at ceiling for arrays of one to three items, and then declined systematically as set size increased from four to 12 items. Overall, the average K value (estimations of VWM capacity) among participants was around three or four items. This finding led to the foundation of the slot model, which posits that individuals can only store between three and four objects in VWM.

In addition to experiments consisting of arrays of single features, Luck and Vogel (1997) also presented participants with arrays consisting of a conjunction of features (i.e., lines differing in orientation and color). Participants completed a change-detection task, but the researchers varied whether participants had to remember a single feature or a conjunction of features. For example, participants would see an array consisting of lines of different orientations and colors. In the color condition, only a color change could occur, and participants were instructed to look for a color change. In the orientation condition, only an orientation change could occur, and participants were instructed to look for an orientation change. And in the conjunction condition, either a color or orientation change could occur, and participants were instructed to remember both features of each item. Thus, in the conjunction condition, participants had to remember eight features but only four integrated objects. If VWM storage capacity is limited by individual features (e.g., color, orientation), then performance should decline at lower set sizes in the conjunction compared to the single feature conditions. However, if VWM storage capacity is limited by integrated objects (e.g., one red, horizontal line), then the same pattern of results should be observed throughout all three conditions.

Consistent with the latter case, they observed that VWM capacities were the same for single feature and conjunction items. Altogether, this provided the basis for the fixed slot model that VWM capacity was constrained by slots of ~3–4 integrated objects. While further research has expanded upon these initial findings, it is also important to note that several experiments have failed to replicate the critical conjunction experiments. These follow-up studies have observed that VWM capacity is, in fact, reduced as feature load increases, irrespective of the number of objects (Fougnie, Asplund, & Marois, 2010; Olson & Jiang, 2002; Wheeler & Treisman, 2002). To explain performance for conjunction conditions, researchers must consider both feature load (i.e., number of features to be remembered) and object load (i.e., number of objects in the display; Hardman & Cowan, 2015). Taken together, the current consensus in the field is that there is a cost to maintaining integrated features in VWM.

A key component of this model is that these VWM slots are considered all or nothing—an observer either remembers every object with the same fidelity (i.e., amount of information) within the capacity limit or fails to remember the object completely. This all-or-nothing component is potentially problematic, as it suggests that an observer has the same amount of information per item whether they viewed single or multiple items. What if an observer sees two encoding displays in a typical change-detection task: one with a single image of an apple and another with four images of very similar looking apples? In each test array, there is either no change, or one of the objects is replaced with another very similar looking image of an apple. According to the fixed slot model, the precision of information available to the observer is the same in both conditions. It does not account for the potential interference four very similar objects may have in terms of their representations in VWM, or how they may affect the decision-making process when an observer decides whether a change has occurred.

Continuous resource model

Another model used to explain apparent limitations in VWM is referred to as the continuous resource model. Unlike the fixed slot model, which defines capacity as being constrained by all-or-nothing slots of integrated objects, the continuous resource model conceptualizes VWM capacity as information based and limited by a finite resource (see Fig. 6). Furthermore, this finite resource can vary unevenly across different items in a display. This unequal division of resources across representations can differ due to a variety of factors, such as top-down goals (e.g., attention) or the total information load of the display (i.e., set size; Bays & Husain, 2008; Wilken & Ma, 2004).

Support for the continuous resource model came from Wilken and Ma (2004), who developed the continuous report method as a way of measuring the fidelity, or amount of information, contained in VWM representations. In the continuous report paradigm, participants are instructed to remember an array consisting of items of a single feature (color, orientation, etc.). After a brief delay (1.5 seconds), a square cue appeared centered at the location of one of the previously presented items. At the same time, a test probe is displayed in the center of the screen, which allows for continuous report of the probed feature. For example, in the color condition, a color wheel containing all possible color values appears in the center of the screen, and participants indicate the color of the probed item by clicking a color on the wheel, using a mouse. Responses are then reported as the degree error from the true color value, and a distribution can be made based on a participant's responses. The standard deviation (SD) of this distribution can then be used to estimate a participant's precision of color information in VWM.

In a series of experiments, Wilken and Ma (2004) varied the set size of displays as well as the type of feature being probed in memory. Regardless of the type of feature probed, they observed that as set size increased, the precision of VWM representations decreased. However, even at large set sizes, the distribution of responses was still centered around the true value of an item, and the precision (i.e., SD) was large but still well above chance. This led the researchers to conclude that individuals could store a continuous amount of information in VWM, but the precision with which an individual item was represented varied as a function of the total information load of the display (i.e., set size). When set size is greater than four items, observers are able to store more than four items in memory. However, fewer resources are available to allocate to each item. Given the constraints of a typical change-detection task, an item may still be represented in VWM,
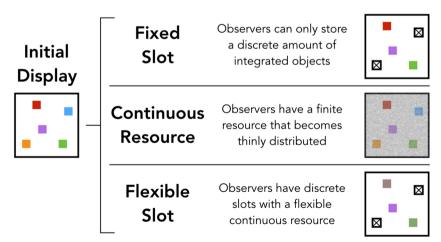

Fig. 6 Illustration of VWM models. According to the fixed slot model, VWM can only store a discrete number of integrated objects. The continuous resource model proposes that VWM has a finite resource that becomes more thinly distributed as the number of items in a display increases. And the flexible slot model is an integration of the latter two, stating VWM has a discrete number of slots, but that a finite resource can be flexibly allocated to each slot

but without the necessary amount of information to make a successful comparison. This suggests that the ~3–4 object limit proposed by the fixed slot model may simply be a behavioral artifact of the tasks being used to assess such limits.

A potential limitation of the research used to support the continuous resource model is that variable precision has not been shown for holistic representations of objects. Studies in support of this model typically probe memory along a single feature dimension (i.e., color), even if observers are viewing images of real-world objects. It remains possible that representations for objects in VWM rely on different kinds of information, some of which may be variable and others that may not. For example, when an observer sees a single image of a real-world object, their representation of that object may contain categorical knowledge (e.g., teddy bear) in addition to other knowledge such as color (e.g., brown). The information of color available to the observer may be variable as a function of set size, but it is likely that categorical knowledge is not—observers either know the categorical identity of an object or they do not.

Work by Schurgin and Flombaum (2015, 2018) provides evidence this may be the case. In their task, participants saw two images of real-world objects in a display, that were then briefly masked, and participants then had to make a 2AFC judgment containing a previously seen object and a completely new object (indicating which of the two was the old object). Critically, they added image noise to stimuli at test by randomly scrambling up to 75% of the pixels in each image. They observed that VWM performance was unaffected by noise at test, with 0% noise and 75% noise demonstrating the same level of performance. The continuous resource model predicts that noise at test should make comparisons in memory harder, as observers would be comparing a noisy internal representation to a noisy external representation (i.e., test stimuli). Thus, when noise is added at test, performance should decrease. In contrast, the fixed slot model would predict no performance decrease. Observers would have no noise in their internal representation, which would allow them to manage noise at test when making a comparison. Altogether, this work suggests that while resolution for memory could vary along a continuous resource for a single feature, such as color, the same may not be true along all dimensions, such as holistic object representations.

Flexible slot model

Contrasting evidence exists that may support either the fixed slot or continuous resource models. The flexible slot model provides a middle ground between the two, proposing that VWM is constrained to a maximum of ~3–4 representations, but that this capacity may also be limited by the amount of information load in the display. In short, VWM is constrained

by slots, but there is flexibility within the system for distributing limited resources across these slots.

One study interpreted by some to support the flexible slot model was conducted by Alvarez and Cavanagh (2004). They utilized a typical change-detection task but varied the information load of displays by changing the type of stimuli presented. On each trial, one to 15 objects were presented for 500 ms, followed by a brief delay (900 ms), and then a test array. On half the trials, one of the objects changed identity, and on the other half, the displays were identical. Participants were instructed to indicate whether one of the objects had changed. Critically, trials could contain stimuli pertaining to a single stimulus class that each differed in their visual complexity: line drawings, shaded cubes, random polygons, Chinese characters, and colored squares.

If VWM is limited by a fixed number of representations (i.e., slots), then performance should be equivalent across all stimulus categories, but if VWM capacity is limited by information load, then these estimates should vary across stimulus categories. After converting responses to K estimates of VWM capacity, Alvarez and Cavanagh (2004) observed there was varying capacity estimates across different stimulus classes, ranging from 1.6 for shaded cubes to 4.4 for colored squares. This provided support that VWM is limited in its number of representations (~4 objects), but also limited by the amount of information (i.e., stimulus complexity) of what is being remembered. However, while some may interpret these results in support of the flexible slot model, a limited number of representations in VWM is not necessarily the same as objects being stored in a slot-like way.

There is disagreement as to whether these differences reflect storage limitations, therefore supporting a flexible slot model, or rather reflect comparison errors made during the decision-making process. It could be that items with higher visual complexity also have higher similarity to one another, and this will lead to greater errors at test even though overall memory capacity for items is the same. Awh et al. (2007) investigated this possibility using Alvarez and Cavangh's (2004) method and stimuli, but with one critical modification: categorical change. When a change occurred at test, they could either be within-category changes (as in Alvarez & Cavanagh; i.e., a Chinese character replaced with a different Chinese character), or they could be between-category changes (i.e., a Chinese character replaced with a line drawing). They observed that for within-category changes, performance varied across stimulus categories, consistent with Alvarez and Cavanagh. Conversely, for between-category changes, they observed no difference in performance relative to stimulus category. This was taken to suggest that variable performance across different kinds of stimuli may be due to differences in similarity, and not information load (consistent with the fixed slot model). However, recent research has found these between-category changes are primarily driven by the use of

global ensemble or texture representations. When objects are clustered in a display by type, individuals can discriminate based on a change in clustering (i.e., using an ensemble or texture representation) rather than one based on individual item memory (Brady & Alvarez, 2015). These results suggest the need for more flexible models of VWM that integrate a role for spatial ensemble representations.

Considerations affecting VWM beyond capacity

An important consideration to our theoretical understanding of working memory is that there are large individual differences in performance, specifically in relation to capacity. Previous research has found capacity estimates of VWM vary substantially across individuals, ranging from 1.5 to 5.0 objects (Vogel et al., 2001; Brady et al., 2011), and that these individual differences in capacity correlate strongly with broad measures of cognitive function, such as academic performance (Alloway & Alloway, 2010) and fluid intelligence (Fukuda, Vogel, Mayr, & Awh, 2010). These individual differences in capacity and relationships with other cognitive functions are likely because VWM is the combination of multiple processes, including short-term memory and executive control mechanisms, many of which vary in performance across individuals (Baddeley, 2003; Conway, Kane, & Engle, 2003).

In the context of the models discussed above, each could be easily amended to account for individual differences. Whether VWM capacity is limited by a fixed or continuous resource, either could conceivably vary across individuals. However, understanding and explaining these individual differences remains key in theoretical discussions of working memory, as they could be the result of potentially different sources. For example, it could be that individual differences in capacity are the result of individual differences across a resource (whether fixed or continuous). Alternatively, these capacity limitations could arise from other components contributing to working memory performance, such as executive function or attention (Baddeley, 2003; Conway et al., 2003). Indeed, research has found potential limitations or biases in VWM can arise from the combination of different sources of information, such as the memory of an item in combination with categorical information (Huttenlocher, Hedges, & Duncan, 1991), nonuniformities in attention (Schurgin & Flombaum, 2014), or ensemble information (Brady & Alvarez, 2011). As a result, understanding the source of individual differences in working memory capacity remains critical to our overall understanding of how capacity limitations arise.

In an effort to explain the capacity limitations of VWM, all the models discussed above also make an implicit assumption that the fidelity of VWM representations is directly affected by the number of objects to be remembered. However, recent research has demonstrated that even with a fixed number of items in a display there is variability across trials in the precision of VWM (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014; Fougnie, Suchow, & Alvarez, 2012; Van den Berg, Shin, Chou, George, & Ma, 2012). This suggests there are not only differences in visual working memory limits across individuals but that the quality of working memory representations varies *within* an individual as well. Moreover, this variability cannot be explained by fluctuations in attention or capacity resources (i.e., slots, continuous resource, etc.; Fougnie et al., 2012).

To better account for these results, researchers have proposed different kinds of variable precision models. These models account for random fluctuations in encoding precision that tend to occur from trial to trial and have been shown to fit data better than traditional fixed slot or continuous resource models (Fougnie et al., 2012; Van den Berg et al., 2012). However, it is important to note that variable precision does not necessarily favor either a slot or continuous resource account of working memory capacity (Van den Berg & Ma, 2017). As a result, the role of variability affecting VWM fidelity should be treated as separate from the source of capacity limitations described by the models above.

Common methods in VWM

Change-detection task

In order to understand the capacity of VWM, many experiments have utilized a change-detection task. In this paradigm, observers first see an array of items (whether a single feature, conjunction of features, or images of real-world objects) and are told to remember what they saw. The array then disappears and after a brief delay reappears with either all the items exactly as before or with a single item changed. Observers are asked to identify whether a change occurred or not (see Fig. 7). The general goal of this paradigm is that by varying the number of items in a display (two, four, eight, etc.), researchers can investigate what kinds of constraints may be limiting VWM capacity (Awh et al., 2007; Luck & Vogel, 1997; Vogel et al., 2001, 2006; Xu, 2002).

A variation of the change-detection task uses a single probe report, showing only one item at test instead of the entire display (with or without a change). A potential advantage of single probe change-detection tasks is that it may be harder for observers to use relational or summary statistical information (i.e., how much the overall display was "blueish" and whether the test display is different in the overall "blueishness") to inform their judgments. A potential disadvantage of this procedure is that if visual memory for certain stimuli (objects, scenes, etc.) relies on relational information (i.e., position, layout, etc.), then probing a single item may remove important information and erroneously reduce performance. It may also

be the case that when participants are given the entire display at test, this encourages them to remember the entire display. It is worth nothing that VWM capacity estimates have been found to be comparable whether they were based on the whole-display or a single-probe procedure (Luck & Vogel, 1997; Jiang, Olson, & Chun, 2000).

Time course of consolidation in VWM Vogel et al. (2006) were interested in understanding the time course of consolidation in VWM using typical change-detection methods. By consolidation, they meant the process whereby VWM becomes durable enough that it is not disrupted by new sensory inputs. The goal of the study was two-fold: to measure the time course of VWM consolidation and whether the rate of consolidation varied as a function of the number of items being consolidated. In the primary experiment, participants viewed arrays of between one and four colored squares. In order to interrupt consolidation, pattern masks at the location of the colored squares were introduced between 117 and 484 ms from the initial onset. After a brief delay (controlled to be the same across different mask interruption timings), a test array was presented that either contained a change or no change. Participants were instructed to judge whether a change occurred. Across Set Sizes 2–4, performance declined when the delay between the initial encoding array and the onset of the mask was shorter, and this effect became larger as set

size increased. Overall, the rate of consolidation was estimated to be ~50 ms per item (although it may be faster, ~10 ms per item; see Sperling, 1963). These results suggest that consolidation is limited in capacity (increasing in time as set size increases) but occurs quite rapidly.

How object based is VWM? Change-detection tasks have also been used to evaluate how object-based encoding in VWM may be affected by remembering objects with two features within the same dimension (i.e., two colors per item) or different dimensions (i.e., color and orientation per item). In one experiment, participants viewed arrays of mushroom-like objects that had two distinct parts: a cap and a stem. In the conjunction condition, five mushroom-like objects were shown, with each cap and stem in an object being a distinct color from one another. In the feature condition, 10 separate parts (five caps and five stems) of varying colors were displayed. Across both conditions, participants were instructed to only monitor cap colors for a possible change, stem colors for a possible change, or both cap and stem colors for a possible change. Participants viewed an array, followed by a brief delay, and then a test array appeared with a possible change present. It was found there was no object-based advantage (i.e., conjunction trials) for remembering one or both colors (Xu, 2002). This suggests that VWM does not always encode visual arrays as integrated objects.
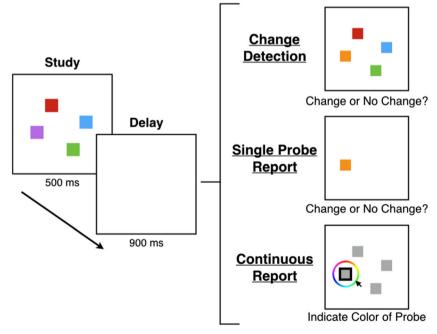

Fig. 7 Illustration of typical VWM methods. In a change-detection task, participants are briefly shown an initial array of items (colors, objects, etc.). After a delay, they are then shown a test display. Half of the time the display is exactly the same, whereas in the other half one of the items in the display has changed. Participants are told to indicate whether a change occurred. In the single probe report task, at test participants are only

shown one item instead of the entire display (with or without a change). In the continuous report task, at test one of the previous items is probed and participants indicate their memory for a single feature in a continuous space (i.e., report the previous color of the item using a color wheel). (Color figure on line)

A follow-up experiment investigated this finding further, exploring whether potential object-based advantages existed for remembering two features along different dimensions. The experiment was similar to the first, but all the mushroom stems were a uniform green color. Instead of varying by color, they varied by orientation (either 45°, 90°, or 135° relative to horizontal). Again, participants viewed conjunction or feature displays, and had to monitor either cap changes (color), stem changes (orientation), or both. Interestingly, participants demonstrated better performance across all conditions in the conjunction relative to the feature display. These results indicate an object-based encoding advantage may exist, but, if so, only for when objects contain two features along different dimensions (Xu, 2002).

Olson and Jiang (2002) used a single probe change-detection task to investigate a similar question of whether VWM was limited by the number of objects or features in a display. Participants first saw a display of three or six color items. After a brief delay, a single item reappeared that was either the same from the previous display or different. Participants indicated whether a change had occurred. All color stimuli were simple objects of the same type (inner square or the outer frame of a square). Crucially, there were two types of stimuli: feature (a single color, either inner square or outer frame) and conjunction (two colors, one in the inner square of the object and another in the outer frame of the object). This created three types of trials. A feature-feature trial consisted only of feature stimuli, and at test a feature item was probed. A conjunction-conjunction condition consisted of both feature and conjunction stimuli, and at test a conjunction item was probed. And a conjunction-feature condition contained both feature and conjunction stimuli, and at test a feature item was probed. It was observed that as set size increased, performance declined across all three trial types. However, performance was higher overall for the feature-feature condition than either of the two conjunction conditions, suggesting there was a cost to maintaining more colors in VWM, even when those colors were bound to a single object. Additional experiments involving color and orientation did observe that integrated objects may improve VWM capacity under certain conditions (Olson & Jiang, 2002).

More recently, Hardman and Cowan (2015) conducted a series of change-detection experiments to investigate whether VWM is limited solely by the number of objects that can be held, or if it is also limited by the complexity of those objects (i.e., number of features to be remembered). Most strikingly, in a direct replication of the four-feature experiment conducted by Luck and Vogel (1997), they found a cost for maintaining more features in VWM in the conjunction condition. Across a variety of experiments (using different encoding times, single probe report, etc.), they consistently observed an effect of feature load, regardless of whether two or four features were contained per object (each along different

dimensions; i.e., color, orientation, gap, and length). These results provide strong evidence that VWM capacity is limited both by the number of objects and number of relevant features of those objects and is inconsistent with the prediction of the fixed slot model that VWM operates in units of integrated objects (Hardman & Cowan, 2015).

Continuous report

In order to measure fidelity of VWM, experiments have used a paradigm often referred to as continuous report. Generally, observers are presented an array of items of a single feature (i.e., color, orientation). After a delay, a single item position is probed (typically highlighted with a square), and participants indicate the specific feature of that item in a continuous space. So if the participant saw four colored squares, after a delay a single item was probed, and participants indicated the color of the object using a continuous color wheel. This allows researchers to create a distribution of how precise the responses for items are (by taking the difference between responses and the true value), and to see how these distributions vary as a function of different features or set sizes. For example, as set size increases a normal distribution with a larger standard deviation (SD) would indicate a less precise VWM representation of that feature. A limitation of change-detection tasks is that they provide little information as to how well each individual object is remembered. However, continuous report allows for a measure that quantifies the amount of information available in memory for an item.

How precise are VWM representations? Zhang and Luck (2008) were interested in evaluating the role continuous resources and/or slots may play in explaining VWM capacity limitations using a continuous report method. In their experiment, participants were presented with one to six color objects in an array. After a brief delay, a single item was probed (using a highlighted square), and participants reported the color of the probed item using a continuous color wheel surrounding the display. In order to calculate performance, they took the error of each response to create a distribution around the true color value, and through a mixture model analysis obtained an estimate of the quantity (i.e. guess rate) and precision (i.e. SD) of memory. They observed that as set size increased from one to three items, SD increased, but then remained constant as set size increased to six. Further studies have collaborated these results (Brady et al., 2013).

Such results are not compatible with a pure resource model of VWM (which would predict SDs to increase across all set sizes), and instead suggest some variation of a "slot" model. Through a follow-up change-detection experiment, Zhang and Luck (2008) proposed that a slots and averaging model (an amended version of the fixed slot model) best fit the observed data in the continuous report experiment. The logic of this

model is that an observer has a fixed amount of "slots" in VWM, able to store information in an all-or-nothing process with a given precision. The key difference between this model and the fixed slot model is that these slots can be flexibly allocated across different items. So if an observer has a capacity of four slots and a display only contains one item, the observer can allocate all these slots to that single stimulus. By then averaging the slots together, the observer can obtain more precise estimates.

Bays, Catalao, and Husain (2009) also used continuous report to build upon this work, but while also taking into account both errors in color and location. Their experiment was similar to Zhang and Luck's (2008)—participants saw an array of one to six colored items, and after a brief delay a single item from the display was probed. Participants reported the color of the probed item by clicking on a color wheel. When modeling distributions of the responses to the targets, however, they also took into account possible responses to nontargets (i.e., location errors). They observed that responses to nontargets were not uniform (which would have reflected guessing), suggesting participants were making location errors during the task. When this error was taken into account (modeling distributions for both the target and nontarget), they observed that precision decreased as the number of items in the display increased. This pattern of data is best explained by a common continuous resource distributed dynamically across the entire display, and not by a slots and averaging model that constrains the number of items that can be represented.

Can VWM strategically trade off between precision and capacity? To assess the flexibility of VWM, Fougnie, Cormiea, Kanabar, and Alvarez (2016) conducted a VWM experiment where during each trial participants were briefly shown five colors to remember, and after a brief delay they were then asked to report the color of the items using continuous report. Critically, trials were blocked into two conditions. In the standard task, a single random location was cued, and participants were asked to report the color of that item as precisely as possible (with a monetary reward for more precise responses). In the get-them-all task, participants were asked to report the color of all five items and received a monetary reward only if all five responses were on the correct half of the color wheel. Data were analyzed using the same method as Zhang and Luck (2008) to estimate the quantity (i.e., guess rate) and quality (i.e., precision) of memory performance. Interestingly, they observed a trade-off between precision and guess rate across conditions. Specifically, in the standard task, participants displayed significantly more precise representations, but at the cost of reduced capacity (i.e., increased guess rate). In contrast, in the get-them-all task participants displayed significantly more capacity, but at a cost to the precision of those representations (Fougnie et al., 2016).

These results demonstrate that observers can strategically control the content of VWM, suggesting greater insight may be possible by studying flexibility in VWM in addition to capacity limitations.

Synthesizing dichotomies in visual memory

When looking at the memory literature, what one notices very clearly is within a particular paradigm, memory system, or stimulus type we often talk about memory in dichotomies. We think of what memory is through the lens of a specific system, method, or type of stimulus. When one zooms out, it becomes natural to ask to what extent are these dichotomies related to each other?

In fact, many of these dichotomies may be intimately related or may even be the same. For instance, the idea that people simply fail to represent information altogether is a component of some fixed capacity theories of VWM and some theories of recall in VLTM. In a typical change-detection task, if an observer fails to detect a change both the fixed slot and flexible slot model suggest the observer did not represent the changed item in memory. The failure at test is not due to an observer not having enough information to make the correct judgment but rather that they had no information in order to make the correct judgment.

Along similar lines, if during a VLTM test an observer fails to identify a previously seen item, according to the DPSD model this may be the result of a recollection failure. When the observer saw the previously seen image at test, recollection failed to occur, so the observer had no information to inform their decision. Without any information related to the previously seen item, they incorrectly classified the image as new. In both cases, fixed capacity theories of VWM and the DPSD model of VLTM say failures of memory are the result of the observer having no information available at test.

A natural question arises as to what extent are the fixed/flexible slot model and recollection describing the same kinds of phenomena but using distinct methods and theories? By utilizing methods across these fields, it may be possible to inform both perspectives. For instance, many VWM researchers studying the fixed slot model (and in VWM research generally) use some variation of a change-detection task. What if classical constructs studied in the broader episodic memory literature, such as familiarity versus recollection, were measured in the context of VWM? For example, participants could be given a working memory display to remember and at subsequent test asked to classify a test object as either old or new, as well as make a source localization judgment. Would it be possible for an observer to incorrectly classify an old object, but make a correct source judgment? What would the implications of such results be for our understanding of VWM capacity? Typically, these results would be interpreted in the context of swap errors, perhaps through misbinding features of items

or incorrectly reporting another memory item present in the display (Fougnie & Alvarez, 2011; Ma et al., 2014). However, what if such results were interpreted under the framework guiding long-term memory? If observers are able to make a correct source but incorrect identity judgment, this might suggest they have different *kinds* of information available at test, consistent with the long-term memory literature. Such data would be inconsistent with a fixed slot model, as an observer would have all or none of the information about an item available at test. It would also have potentially far-reaching implications for continuous resource models of VWM (continuous resource or flexible slot), as it would suggest that not only the amount of information about an item can vary, but the *kinds* of information available about an item can vary as well.

There is also the idea that failures during memory tests are not the result of failing to represent information, but rather that the information available to observers to make decisions declines or becomes less precise under certain conditions. This is a notion shared by both the continuous resource model of VWM and theories of familiarity in VLTM. In a typical change-detection task, it may not be the case that observers fail to represent an item if they do not detect a change occurred. The continuous resource model explains such performance failures as the result of a finite continuous resource that becomes more thinly distributed as the number of items in VWM increases. An observer may incorrectly report no change occurred, but they may still have some information of representation of that item. It was simply that they did not have enough information about that item in order to make the required judgment (i.e., change or no change).

In VLTM, both the DSPD and CDP models of familiarity suggest that sometimes an observer may fail to correctly recognize a previously seen item but still have some information in memory about that item. An observer may have a familiarity signal related to a previously seen item. Familiarity operates via a signal-detection-based process, so even if the signal is in memory, an observer may still fail to recognize a previously seen item depending on a variety of factors. The memory signal used in the comparison process could be weak, or the observer may have a particularly conservative decision criterion (avoiding all potential false alarms at the cost of recognizing some previously seen items). Regardless, this means an observer may fail to recognize an object, but still have some information about that object in memory.

Again, it becomes natural to ask to what extent are such theories describing the same kinds of phenomena? Are certain kinds of memory performance classified as "familiarity" in VLTM simply measuring memories with more thinly distributed resources? What happens in VLTM tasks as set size during encoding increases? Do observers transition from "recollection" based responses to "familiarity" responses? At high set sizes, can observers still demonstrate responses

typically associated with recollection for certain items (consistent with the fixed slot model of VWM), or does performance reliably conform to familiarity-based responses (consistent with the continuous resource model)? Interestingly, variable performance at high set sizes for recollection-based and familiarity-based responses in VLTM would be analogous to the flexible slots model, and entirely consistent with DPSD and CDP models of familiarity and recollection.

Pattern separation and pattern completion bear obvious resemblance to tolerance and discrimination. Tolerance refers to our ability to recognize something despite considerable changes in inputs across encounters. Pattern completion, which is the process by which varying inputs are recognized as belonging to the same source, seems to be one such process through which tolerance may be achieved. Discrimination refers to our ability to distinguish similar but distinct inputs from one another, and pattern separation is defined as the process of recognizing overlapping inputs as distinct. Given these definitions, it becomes abundantly clear that researchers studying VLTM in terms of pattern separation and pattern completion, and vision researchers studying object recognition in terms of tolerance and discrimination are investigating the same processes. But these two fields of research have operated largely independently of one another, devising their own methods and terminology to investigate the same phenomena. How, then, might methods from vision scientists studying object recognition inform episodic long-term memory researchers studying pattern separation/completion? Do rules of perception that guide object identity also affect visual episodic long-term memory? Recent research suggests this is the case, as expectations about spatiotemporal continuity (broadly known to support object perception; see Scholl, 2001, for a review) appear to be critical for the construction of stronger, more tolerant long-term memories (Schurgin & Flombaum, 2017). Broadly, beginning to answer these questions may provide further insight to our understanding of memory, in addition to bridging sometimes disparate but related fields of study across vision science, psychology, and neuroscience.

Conclusion

Despite generally being studied separately, VWM, VLTM, and even object recognition research are clearly intimately related to one another. Indeed, many dichotomies established across these fields may be describing the same kinds of processes or behaviors. Here, I have provided a review of the major theories and methods guiding VWM and VLTM research, with the goal to provide background of one field to researchers primarily operating in the other. While tremendous advances in our understanding of visual memory have been made operating separately of one another, each of these fields may greatly benefit through increased interaction and

application with one another. By more closely relating how aspects already identified in visual working memory and long-term memory may interact and support one another, we can better address the major unresolved questions of visual memory. Namely, what are the representational structure and contents of our memories, and how do we acquire them?

References

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29.

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding, 117*(8), 827–891.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science, 15*(2), 106–111.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*(7), 622–628.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89.

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

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), 7.1–7.11. https://doi.org/10.1167/9.10.7.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851–854. https://doi.org/10.1126/science.1158023.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*(1), 38–64.

Bowles, B., Crupi, C., Pigott, S., Parrent, A., Wiebe, S., Janzen, L., & Köhler, S. (2010). Double dissociation of selective recollection and familiarity impairments following two different surgical treatments for temporal-lobe epilepsy. *Neuropsychologia, 48*(9), 2640–2647.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22*(3), 384–392.

Brady, T. F., & Alvarez, G. A. (2015). No evidence for a fixed object limit in working memory: Ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41*(3), 921–929.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5). https://doi.org/10.1167/11.5.4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325–14329.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*(6), 981–990.

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology, 2*(1), 1–3.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex, 12*(10), 1048–1056.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547–552.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research, 169*, 323–338.

Cox, D. D., & DiCarlo, J. J. (2008). Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 28*(40), 10045–10055.

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). "Breaking" position-invariant object recognition. *Nature Neuroscience, 8*(9), 1145–1147.

Dede, A. J., Squire, L. R., & Wixted, J. T. (2014). A novel approach to an old problem: Analysis of systematic errors in two models of recognition memory. *Neuropsychologia, 52*, 51–56.

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*(9), 379–386.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*(8), 333–341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological review, 111*(2), 524.

Fan, J. E., Hutchinson, J. B., & Turk-Browne, N. B. (2016). When past is present: Substitutions of long-term memory for sensory evidence in perceptual judgments. *Journal of Vision, 16*(8), 1–12.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature, 431*(7005), 188–191.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11*(12), 3–3. https://doi.org/10.1167/11.12.33.

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

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*(8), 1231. https://doi.org/10.1037/xhp0000211.

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

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17*(5), 673–679.

Gabrieli, J. D., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science, 6*(2), 76–82.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics (Vol. 1). New York: Wiley.

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, 66(1), 68–78.

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 325–347. https://doi.org/10.1037/xlm0000031.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632–635. https://doi.org/10.1038/nature07832.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'Reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, 12(3), 341–351.

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*, 24(5), 790–796.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, 137(1), 163–181. https://doi.org/10.1037/0096-3445.137.1.163.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352–376.

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 325–339. https://doi.org/10.1037/a0025483.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, memory, and cognition, 26*(3), 683.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, 54(1), 99–112.

Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus*, 23(4), 287–294.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558–578. https://doi.org/10.1037/a0019165.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*(11), 1551–1556.

Loiotile, R. E., & Courtney, S. M. (2015). A signal detection theory analysis of behavioral pattern separation paradigms. *Learning & Memory*, 22(8), 364–369.

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

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356.

Macmillan, N. A., & Creelman, C. D. (2004). Detection theory: A user's guide. New York: Psychology Press.

Miyamoto, K., Adachi, Y., Osada, T., Watanabe, T., Kimura, H. M., Setsuie, R., & Miyashita, Y. (2014). Dissociable memory traces within the macaque medial temporal lobe predict subsequent recognition performance. *The Journal of Neuroscience*, 34(5), 1988–1997.

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, 81(2), 416–427.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics*, 64(7), 1055–1067.

Pessoa, L., Gutierrez, E., Bandettini, P. A., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35*(5), 975–987.

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLOS Computational Biology, 4*(1), e27. https://doi.org/10.1371/journal.pcbi.0040027.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509–522.

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

Reagh, Z. M., & Yassa, M. A. (2014). Object and spatial mnemonic interference differentially engage lateral and medial entorhinal cortex in humans. *Proceedings of the National Academy of Sciences*, 111(40), E4264-E4273.

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current Opinion in Neurobiology*, 20(3), 382–388.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B, 372*(1711). https://doi.org/10.1098/rstb.2016.0049.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*(1/2), 1–46.

Scholl, B. J., & Flombaum, J. I. (2010). Object persistence. In B. Goldstein (Ed.), Encyclopedia of perception (Vol. 2, pp. 653–657). Thousand Oaks: SAGE.

Schurgin, M. W., & Flombaum, J. I. (2014). How undistorted spatial memories can produce distorted responses. *Attention, Perception, & Psychophysics, 76*(5), 1371–1380.

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, 23(7), 859–862.

Schurgin, M. W., & Flombaum, J. I. (2017). Exploiting core knowledge for visual object recognition. *Journal of Experimental Psychology: General, 146*(3), 362–375.

Schurgin, M. W., & Flombaum, J. I. (2018). Visual working memory is more tolerant than visual long-term memory. *Journal of Experimental Psychology: Human Perception and Performance*.

Schurgin, M. W., Reagh, Z. M., Yassa, M. A., & Flombaum, J. I. (2013). Spatiotemporal continuity alters long-term memory representation of objects. *Visual Cognition*, 21(6), 715–718.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Sperling, G. (1963). A model for visual memory tasks. *Human factors, 5*(1), 19–31.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171–177.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology, 25*(2), 207–222.

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, 51(12), 2442–2449.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751–754.

Van den Berg, R., & Ma, W. J. (2017). A rational theory of the limitations of working memory and attention. *bioRxiv.* https://doi.org/10.1101/151365.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences, 109*(22), 8780–8785.

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*(1), 92–114.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance, 32*(6), 1436–51.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences, 3*(1), 22–31.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98*(8), 4800–4804.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48-64.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 1120–1125.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616–641.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 458–468.

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus, 21*(9), 968–979.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*(10), 515–525.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language, 46*(3), 441–517.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20*(11), 1178–1194.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.