A review of visual memory capacity: Beyond individual items and towards structured representations

Timothy F. Brady[1], **Talia Konkle**[1], and **George A. Alvarez**[2]

Timothy F. Brady: tfbrady@mit.edu; Talia Konkle: tkonkle@mit.edu; George A. Alvarez: alvarez@wjh.harvard.edu

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2]Vision Sciences Laboratory, Department of Psychology, Harvard University

Abstract

Traditional memory research has focused on identifying separate memory systems and exploring different stages of memory processing. This approach has been valuable for establishing a taxonomy of memory systems and characterizing their function, but has been less informative about the nature of stored memory representations. Recent research on visual memory has shifted towards a representation-based emphasis, focusing on the contents of memory, and attempting to determine the format and structure of remembered information. The main thesis of this review will be that one cannot fully understand memory systems or memory processes without also determining the nature of memory representations. Nowhere is this connection more obvious than in research that attempts to measure the capacity of visual memory. We will review research on the capacity of visual working memory and visual long-term memory, highlighting recent work that emphasizes the contents of memory. This focus impacts not only how we estimate the capacity of the system - going beyond quantifying how many items can be remembered, and moving towards structured representations - but how we model memory systems and memory processes.

Tulving (2000) provided a concise, general definition of memory as the "neurocognitive capacity to encode, store, and retrieve information," and suggested the possibility that there are many separate memory systems that fit this definition. Indeed, one of the primary aims of modern memory research has been to identify these different memory systems (Schacter & Tulving, 1994). This approach has lead to an extensive taxonomy of memory systems which are characterized by differences in timing, storage capacity, conscious access, active maintenance, and mechanisms of operation.

Early on, William James (1890) proposed the distinction between primary memory – the information held in the 'conscious present' – and secondary memory, which consists of information that is acquired, stored outside of conscious awareness, and then later remembered. This distinction maps directly onto the modern distinction between short-term memory (henceforth working memory) and long-term memory (Atkinson & Shiffrin, 1968; Scoville & Milner, 1957; Waugh & Norman, 1965). The most salient difference between these systems is their capacity: the active, working memory system has an extremely limited capacity of only a few items (Cowan, 2001; Cowan, 2005; Miller, 1956), whereas the passive, long-term memory system can store thousands of items (Brady et al. 2008; Standing, 1973; Voss, 2009) with remarkable fidelity (Brady et al., 2008; Konkle et al., 2010a).

The emphasis on memory systems and memory processes has been quite valuable in shaping cognitive and neural models of memory. In general, this approach aims to characterize memory systems in a way that generalizes over representational content (Schacter &

Tulving, 1994). For example, working memory is characterized by a severely limited capacity regardless of whether items are remembered visually or verbally (e.g., Baddeley, 1986), and long-term memory has a very high capacity whether the items remembered are pictures (e.g., Standing, 1973), words (e.g., Shepard, 1967), or associations (e.g., Voss, 2009). However, generalization across content leaves many basic questions unanswered regarding the nature of stored representations: What is the structure and format of those representations?

Research on visual perception takes the opposite approach, attempting to determine what is being represented, and to generalize across processes. For example, early stages of visual representation consist of orientation and spatial frequency features. Vision research has measured the properties of these features, such as their tuning curves and sensitivity (e.g., Blakemore & Campbell, 1969), and shown that these tuning properties are constant across several domains of processing (e.g., from simple detection to visual search).

Thus, the intersection between memory and vision is a particularly interesting domain of research because it concerns both the processes of memory and the nature of the stored representations (Luck & Hollingworth, 2008). Recent research within the vision science community at this intersection between memory and vision has been quite fruitful. For example, working memory research has shown an important item/resolution tradeoff: as the number of items remembered increases, the precision with which each one is remembered decreases, possibly with an upper bound on the number of items that may be stored (Alvarez & Cavanagh, 2004; Zhang & Luck, 2008), or possibly without an upper bound (Bays, Catalao & Husain, 2009; Wilken & Ma, 2004). In long-term memory, it is possible to store thousands of detailed object representations (Brady et al. 2008; Konkle et al., 2010b) but only for meaningful items that connect with stored knowledge (Konkle et al., 2010b; Wiseman and Neisser, 1974).

Here we review recent research in the domains of visual working memory and visual long-term memory, focusing on how models of these memory systems are altered and refined by taking the contents of memory into account.

VISUAL WORKING MEMORY

The working memory system is used to hold information actively in mind, and to manipulate that information to perform a cognitive task (Baddeley, 1986; Baddeley, 2000). While there is a long history of research on verbal working memory and working memory for spatial locations (e.g., Baddeley, 1986), the last 15 years has seen surge in research on visual working memory, specifically for visual feature information (Luck & Vogel, 1997).

The study of visual working memory has largely focused on the *capacity* of the system, both because limited capacity is one of the main hallmarks of working memory, and because individual differences in measures of working memory capacity are correlated with differences in fluid intelligence, reading comprehension, and academic achievement (Alloway & Alloway, 2010; Daneman & Carpenter, 1980; Fukuda, Vogel, Mayr & Awh, 2010; Kane, Bleckly, Conway & Engle, 2001). This relationship suggests that working memory may be a core cognitive ability that underlies, and constrains, our ability to process information across cognitive domains. Thus, understanding the capacity of working memory could provide important insight into cognitive function more generally.

In the broader working memory literature, a significant amount of research has focused on characterizing memory limits based on how quickly information can be refreshed (e.g., Baddeley, 1986) or the rate at which information decays (Broadbent, 1958; Baddeley & Scott, 1971). In contrast, research on the capacity of visual working memory has focused on

the number of items that can be remembered (Luck & Vogel, 1997; Cowan, 2001). However, several recent advances in models of visual working memory have been driven by a focus on the content of working memory representations rather than how many individual items can be stored.

Here we review research that focuses on working memory representations, including their fidelity, structure, and effects of stored knowledge. While not an exhaustive review of the literature, these examples highlight the fact that working memory representations have a great deal of structure beyond the level of individual items. This structure can be characterized as a hierarchy of properties, from individual features, to individual objects, to across-object ensemble features (spatial context and featural context). Together, the work reviewed here illustrates how a representation-based approach has led to important advances, not just in understanding the nature of stored representations themselves, but also in characterizing working memory capacity and shaping models of visual working memory.

The fidelity of visual working memory

Recent progress in modeling visual working memory has resulted from an emphasis on estimating the fidelity of visual working memory representations. In general, the capacity of any memory system should be characterized both in terms of the number of items that can be stored, and in terms of the fidelity with which each individual item can be stored. Consider the case of a USB-drive that can store exactly 1000 images: the number of images alone is not a complete estimate of this USB-drive's storage capacity. It is also important to consider the resolution with which those images can be stored: if each image can be stored with a very low resolution, say 16×16 pixels, then the drive has a lower capacity than if it can store the same number of images with a high resolution, say 1024×768 pixels. In general, the true capacity of a memory system can be estimated by multiplying the maximum number of items that can be stored by the fidelity with which each individual item can be stored (capacity = quantity X fidelity). For a memory system such as your USB-drive, there is only an information limit on memory storage, so the number of files that can be stored is limited only by the size of those files. Whether visual working memory is best characterized as an information limited system (Alvarez & Cavanagh, 2004; Wilken & Ma, 2004), or whether it has a pre-determined and fixed item limit (Luck & Vogel, 1997; Zhang & Luck, 2008) is an active topic of debate in the field.

Luck and Vogel's (1997) landmark study on the capacity of visual working memory spurred the surge in research on visual working memory over the past 15 years. Luck and Vogel (1997) used a change detection task to estimate working memory capacity for features and conjunctions of features (Figure 1a; see also Pashler, 1988; Phillips, 1974; Vogel, Woodman & Luck, 2001). On each trial, observers saw an array of colored squares and were asked to remember them. The squares then disappeared for about one second, and then reappeared with either all of the items exactly the same as before, or with a single square having changed color to a categorically different color (e.g., yellow to red). Observers were asked to say whether the display was exactly the same or whether one of the squares had changed (Figure 1a).

Luck and Vogel (1997) found that observers were able to accurately detect changes most of the time when there were fewer than 3 or 4 items on the display, but that performance declined steadily as the number of items increased beyond 4. Luck and Vogel (1997) and Cowan (2001) have shown that this pattern of performance is well explained by a model in which a fixed number of objects (3–4) were remembered. Thus, these results are consistent with a "slot model" of visual working memory capacity (see also Cowan, 2005; Rouder, Morey, Cowan, Zwilling, Morey & Pratte, 2008) in which working memory can store a fixed number of items.

Importantly, this standard change detection paradigm provides little information about how well each individual object was remembered. The change detection paradigm indicates only that items were remembered with sufficient fidelity to distinguish an object's color from a categorically different color. How much information do observers actually remember about each object?

Several new methods have been used to address this question (see Figure 1b, c, d). First, the change detection task can be modified to vary the amount of information that must be stored by varying the type of changes that can occur. For example, changing from one shade of red to a different, similar shade of red requires a high-resolution representation, whereas a change from red to blue can be detected with a low-resolution representation. Using such changes that require high-resolution representations has proved particularly fruitful for investigating memory capacity for complex objects (Figure 1b, c). Second, estimates of memory precision can be obtained by using a continuous report procedure in which observers are cued to report the features of an item, and then adjust that item to match the remembered properties. Using this method, the fidelity of a simple feature dimension like color can be investigated by having observers report the exact color of a single item (Figure 1d).

Fidelity of storage for complex objects—While early experiments using large changes in a change detection paradigm found evidence for a slot model, in which memory is limited to storing a fixed number of items, subsequent experiments with newer paradigms that focused on the precision of memory representations have suggested an information-limited model. Specifically, Alvarez and Cavanagh (2004) proposed that there is an information limit on working memory, which would predict a trade off between the number of items stored and the fidelity with which each item is stored. For example, suppose working memory could store 8 bits of information. It would be possible to store a lot of information about 1 object (8 bits/object = 8 bits), or a small amount of information about 4 objects (2 bits/object = 8 bits). To test this hypothesis, Alvarez and Cavanagh varied the amount of information required to remember objects, from categorically different colors (low information load), to perceptually similar 3D cubes (high information load). The results showed that the number of objects that could be remembered with sufficient fidelity to detect the changes depended systematically on the information load per item: the more information that had to be remembered from an individual item, the fewer the total number of items that could be stored with sufficient resolution, consistent with the hypothesis that there is a limit to the total amount of information stored.

This result was not due to an inability to discriminate the more complex shapes, such as 3D cubes: observers could easily detect a change between cubes when only a single cube was remembered, but they could not detect the same change when they tried to remember 4 cubes. This result suggests that encoding additional items reduced the resolution with which each individual item could be remembered, consistent with the idea that there is an information limit on memory. Using the same paradigm but varying the difficulty of the memory test, Awh, Barton and Vogel (2007) found a similar result: with only a single cube in memory, observers could easily detect small changes in the cube's structure. However, with several cubes in memory, observers were worse at detecting these small changes but maintained the ability to detect larger changes (e.g., changing the cube to a completely different kind of stimulus, like a Chinese character; Figure 1b, c). This suggests that when many cubes are stored, less information is remembered about each cube, and this low-resolution representation is sufficient to make a coarse discrimination (3D cube vs. Chinese character), but not a fine discrimination (3D cube vs. 3D cube). Taken together, these two studies suggest that working memory does not store a fixed number of items with fixed fidelity: the fidelity of items in working memory depends on a flexible resource that is

shared among items, such that a single item can be represented with high fidelity, or several items with significantly lower fidelity (see Zosh & Feigenson, 2009 for a similar conclusion with infants).

Fidelity of simple feature dimensions—While the work of Alvarez and Cavanagh (2004) suggests a tradeoff between the number of items stored and the resolution of storage, other research has demonstrated this trade off directly by measuring the precision of working memory along continuous feature dimensions (Wilken & Ma, 2004). For example, Wilken and Ma (2004) devised a paradigm in which a set of colors appeared momentarily and then disappeared. After a brief delay, the location of one color was cued, prompting the observer to report the exact color of the cued item by adjusting a continuous color wheel (Figure 1d). Wilken and Ma (2004) found that the accuracy of color reports decreased as the number of items remembered increased, suggesting that memory precision decreased systematically as more items were stored in memory. This result would be predicted by an information-limited system, because high precision responses contain more information than low-precision responses. In other words, as more items are stored and the precision of representations decreases, the amount of information stored per item decreases.

Wilken and Ma's (2004) investigations into the precision of working memory appear to support an information-limited model. However, using the same continuous report paradigm and finding similar data, Zhang and Luck (2008) have argued in favor of a slot-model of working memory, in which memory stores a fixed number of items with fixed fidelity. To support this hypothesis, they used a mathematical model to partial errors in reported colors into two different classes: those resulting from noisy memory representations, and those resulting from random guesses. Given a particular distribution of errors, this modeling approach yields an estimate of the likelihood that items were remembered, and the fidelity with which they were remembered. Zhang and Luck (2008) found that the proportion of random guesses was low from 1 to 3 items, but that beyond 3 items the rate of random guessing increased. This result is naturally accounted for by a slot model in which a maximum of 3 items can be remembered.

However, Zhang and Luck (2008) also found that the fidelity of representations decreased from 1 to 3 items (representations became less and less precise). A slot model cannot easily account for this result without additional assumptions. To account for this pattern, Zhang and Luck proposed that working memory has 3 discrete slots. When only one item is remembered, each memory slot stores a separate copy of that one item, and these copies are then averaged together to yield a higher-resolution representation. Critically, this averaging process improves the fidelity of the item representation because each copy has error that is completely independent of the error in other copies, so when they are averaged these sources of error cancel out. When 3 items are remembered, each item occupies a single slot, and without the benefits of averaging multiple copies, each of the items is remembered with a lower resolution (matching the resolution limit of a single slot).

This version of the slot model was consistent with the data, but only when the number of slots was assumed to be 3. Thus, the decrease in memory precision with increasing number of items stored can be accounted for by re-casting memory slots as 3 quantum units of resources that can be flexibly allocated to at most 3 different items (a set of "discrete fixed-resolution representations"). This account depends critically on the finding that memory fidelity plateaus and remains constant after 3 items, which remains a point of active debate in the literature (e.g., Anderson, Vogel & Awh, 2011; Bays, Catalao & Husain, 2009; Bays & Husain, 2008). In particular, Bays, Catalao and Husain (2009) have proposed that the plateau in memory fidelity beyond 3 items (Zhang & Luck, 2008) is an artifact of an increase in "swap errors" in which the observer accidentally reports the wrong item from the

display. However, the extent to which such swaps can account for this plateau is still under active investigation (Anderson, Vogel & Awh, 2011; Bays, Catalao & Husain, 2009).

Conclusion—To summarize, by focusing on the contents of visual working memory, and on the fidelity of representations in particular, there has been significant progress in models of visual working memory and its capacity. At present, there is widespread agreement in the visual working memory literature that visual working memory has an extremely limited capacity and that it can represent 1 item with greater fidelity than 3–4 items. This finding requires the conclusion that working memory is limited by a resource that is shared among the representations of different items (i.e., is information-limited). Some models claim that resource allocation is discrete and quantized into slots (Anderson, Vogel & Awh, 2011; Awh, Barton & Vogel, 2007; Zhang & Luck, 2008), while others claim that resource allocation is continuous (Bays & Husain, 2008; Huang, 2010; Wilken & Ma, 2004), but there is general agreement that working memory is a flexibly-allocated resource of limited capacity.

Research on the fidelity of working memory places important constraints on both continuous and discrete models. If working memory is slot-limited, then those slots must be recast as a flexible resource, all of which can be allocated to a single item to gain precision in its representation, or which can be divided separately among multiple items yielding relatively low-resolution representations of each item. If memory capacity is information-limited, then it is necessary to explain why under some conditions it appears that there is an upper bound on memory storage of 3–4 objects (e.g. Alvarez & Cavanagh, 2004; Awh, Barton & Vogel, 2007; Luck & Vogel, 1997; Zhang & Luck, 2008) and in other conditions it appears that memory is purely information-limited, capable of storing more-and-more, increasingly noisy representations even beyond 3–4 items (e..g, Bays & Husain, 2008; Bays, Catalao & Husain, 2009; Huang, 2010).

The representation of features vs. objects in visual working memory

Any estimate of memory capacity must be expressed with some unit, and what counts as the appropriate unit depends upon how information is represented. Since George Miller's (1956) seminal paper claiming a limit of 7 +/− 2 chunks as the capacity of working memory, a significant amount of work has attempted to determine the units of storage in working memory. In the domain of verbal memory, for example, debate has flourished about the extent to which working memory capacity is limited by storing a fixed number of chunks vs. time-based decay (Baddeley, 1986; Cowan 2005; Cowan & AuBuchon, 2008). In visual working memory, this debate has focused largely on the issue of whether separate visual features (color, orientation, size) are stored in independent "buffers," each with their own capacity limitations (e.g., Magnussen, Greenlee & Thomas, 1996), or whether visual working memory operates over integrated object representations (Luck & Vogel, 1997; Vogel, Woodman & Luck, 2001; see Figure 2b).

Luck and Vogel (1997) provided the first evidence that visual working memory representations should be thought of as object-based. In their seminal paper (Luck & Vogel, 1997), they found that observers' performance on a change detection task was identical whether they had to remember only one feature per object (orientation or color), two features per object (both color and orientation), or even four features per object (color, size, orientation and shape). If memory was limited in terms of the number of features, then remembering more features per object should have a cost. Because there was no cost for remembering more features, Luck and Vogel concluded that objects are the units of visual working memory. In fact, Luck and Vogel (1997) initially provided data demonstrating that observers could remember 3–4 objects even when those objects each contained 2 colors. In

other words, observers could only remember 3–4 colors when each color was on a separate object, but they could remember 6–8 colors when those colors were joined into bi-color objects. However, subsequent findings have provided a number of reasons to temper this strong, object-based view of working memory capacity. In particular, recent evidence has suggested that, while there is some benefit to object-based storage, objects are not always encoded in their entirety, and multiple features within an object are encoded with a cost.

Objects are not always encoded in their entirety—A significant body of work has demonstrated that observers do not always encode objects in their entirety. When multiple features of an object appear on distinct object parts, observers are significantly impaired at representing the entire object (Davis & Holmes, 2005; Delvenne & Bruyer, 2004; Delvenne & Bruyer, 2006; Xu, 2002a). For instance, if the color feature appears on one part of an object, and the orientation feature on another part of the object, then observers perform worse when required to remember both features than when trying to remember either feature alone (Xu, 2002a). In addition, observers sometimes encode some features of an object but not others, for example remembering their color but not their shape (Bays, Wu & Husain, 2011; Fougnie & Alvarez, submitted), particularly when only a subset of features is task-relevant (e.g., Droll, Hayhoe, Triesch, & Sullivan, 2005; Triesch, Ballard, Hayhoe & Sullivan, 2003; Woodman & Vogel, 2008). Thus, working memory does not always store integrated object representations.

Costs for encoding multiple features within an object—Furthermore, another body of work has demonstrated that encoding more than one feature of the same object does not always come without cost. Luck and Vogel (1997) provided evidence that observers could remember twice as many colors when those colors were joined into bi-color objects. This result suggested that memory was truly limited by the number of objects that could be stored, and not the number of features. However, this result has not been replicated, and indeed there appears to be a significant cost to remembering two colors on a single object (Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002b). In particular, Wheeler and Treisman's work (2002) suggests that memory is limited to storing a fixed number of colors (3–4) independent of how those colors are organized into bi-color objects. This indicates that working memory capacity is not limited only by the number of objects to-be-remembered; instead, some limits are based on the number of values that can be stored for a particular feature dimension (e.g., only 3–4 colors may be stored).

In addition to limits on the number of values that may be stored within a particular feature dimension, data on the fidelity of representations suggests that even separate visual features from the same object are not stored completely independently. In an elegant design combining elements of the original work of Luck and Vogel (1997) with the newer method of continuous report (Wilken & Ma, 2004), Fougnie, Asplund and Marois (2010) examined observers' representations of multi-feature objects (oriented triangles of different colors; see Figure 2a). Their results showed that, while there was no cost for remembering multiple features of the same object in a basic change-detection paradigm (as in Luck and Vogel, 1997), this null result was obtained because the paradigm was not sensitive to changes in the fidelity of the representation. In contrast, the continuous report paradigm showed that, even within a single simple object, remembering more features results in significant costs in the fidelity of each feature representation. This provides strong evidence against any theory of visual working memory capacity in which more information can be encoded about an object without cost (e.g., Luck and Vogel, 1997), but at the same time provides evidence against the idea of entirely separate memory capacities for each feature dimension.

Benefits of object-based storage beyond separate buffers—While observers cannot completely represent 3–4 objects independently of their information load, there is a

benefit to encoding multiple features from the same object compared to the same number of features on different objects (Fougnie, Asplund and Marois, 2010; Olson & Jiang, 2002; Quinlan & Cohen, 2011). For example, Olson and Jiang showed that it is easier to remember the color and orientation of 2 objects (4 features total), than the color of 2 objects and the orientation of 2 separate objects (still 4 features total). In addition, while Fougnie, Asplund and Marois (2010) showed that there is a cost to remembering more features within an object, they found that there is greater cost to remembering features from different objects. Thus, while remembering multiple features within an object led to decreased fidelity for each feature, remembering multiple features on different objects led to both decreased fidelity and a decreased probability of successfully storing any particular feature (Fougnie, Asplund, & Marois, 2010).

Conclusion—So what *is* the basic unit of representation in visual working memory? While there are significant benefits to encoding multiple features of the same object compared to multiple features across different objects (e.g., Fougnie, Asplund and Marois, 2010; Olson & Jiang, 2002), visual working memory representations do not seem to be purely object-based. Memory for multi-part objects demonstrates that the relative location of features within an object limits how well those features can be stored (Xu, 2002a), and even within a single simple object, remembering more features results in significant costs in the fidelity of each feature representation (Fougnie, Asplund & Marois, 2010). These results suggest that what counts as the right "unit" in visual working memory is not a fully integrated object representation, or independent feature representations. In fact, no existing model captures all of the relevant data on the storage of objects and features in working memory.

One possibility is that the initial encoding process is object-based (or location-based), but that the "unit" of visual working memory is *a hierarchically-structured feature-bundle* (Figure 2b): at the top level of an individual "unit" is an integrated object representation, at the bottom level of an individual "unit" are low-level feature representations, with this hierarchy organized in a manner that parallels the hierarchical organization of the visual system. Thus, a hierarchical feature-bundle has the properties of independent feature stores at the lower level, and the properties of integrated objects at a higher level. Because there is some independence between lower-level features, it is possible to modulate the fidelity of features independently, and even to forget features independently. On the other hand, encoding a new hierarchical feature-bundle might come with an "overhead cost" that could explain the object-based benefits on encoding. On this view, remembering any feature from a new object would require instantiating a new hierarchical feature-bundle, which might be more costly than simply encoding new features into an existing bundle.

This proposal for the structure of memory representations is consistent with the full pattern of evidence described above, including the benefit for remembering multiple features from the same objects relative to different objects, and the cost for remembering multiple features from the same object. Moreover, this hierarchical working-memory theory is consistent with evidence showing a specific impairment in object-based working memory when attention is withdrawn from items (e.g., binding failures: Fougnie & Marois, 2009; Wheeler & Treisman, 2002; although this is an area of active debate; see Allen, Baddeley & Hitch, 2006; Baddeley, Allen, & Hitch, 2011; Gajewski & Brockmole, 2006; Johnson, Hollingworth & Luck, 2008; Stevanovski & Jilicoeur, 2011).

Furthermore, there is some direct evidence for separate capacities for feature-based and object-based working memory representations, with studies showing separable priming effects and memory capacities (Hollingworth & Rasmussen, 2010; Wood, 2009; Wood, 2011a). For example, observers may be capable of storing information about visual objects using both a scene-based feature memory (perhaps of a particular view), and also a higher-

level visual memory system that is capable of storing view-invariant, 3D object information (Wood, 2011a; Wood, 2009).

It is important to note that our proposed hierarchical feature-bundle model is not compatible with a straightforward item-based or chunk-based model of working memory capacity. A key part of such proposals (e.g., Cowan, 2001; Cowan et al. 2004) is that memory capacity is limited only by the number of chunks encoded, not taking into account the information within the chunks. Consequently, these models are not compatible with evidence showing that there are limits simultaneously at the level of objects and the level of features (e.g., Fougnie et al. 2010). Even if a fixed number of objects or chunks could be stored, this limit would not capture the structure and content of the representations maintained in memory.

Thus far we have considered only the structure of individual items in working memory. Next we review research demonstrating that working memory representations includes another level of organization that represents properties that are computed across sets of items.

Interactions between items in visual working memory

In the previous two sections, we discussed the representation of *individual items* in visual working memory. However, research focusing on contextual effects in memory demonstrates that items are not stored in memory completely independent of one another. In particular, several studies have shown that items are encoded along with spatial context information (the spatial layout of items in the display), and with featural context information (the ensemble statistics of items the display). These results suggest that visual working memory representations have a great deal of structure beyond the individual item level. Therefore, even a complete model of how individual items are stored in working memory would not be sufficient to characterize the capacity of visual working memory. Instead, the following findings regarding what information is represented, and how representations at the group or ensemble level affect representations at the individual item level, must be taken into account in any complete model of working memory capacity.

Influences of spatial context—Visual working memory paradigms often require observers to remember not only the featural properties of items (size, color, shape, identity), but also where those items appeared in the display. In these cases, memory for the features of individual items may be dependent on spatial working memory as well (for a review of spatial working memory, see Awh & Jonides, 2001). The most prominent example of this spatial-context dependence is the work of Jiang, Olson and Chun (2000), who demonstrated that changing the spatial context of items in a display impairs change detection. For example, when the task was to detect whether a particular item changed color, performance was worse if the other items in the display did not reappear (Figure 3a), or reappeared with their relative spatial locations changed. This interference suggests that the items were not represented independently of their spatial context (see also Vidal et al 2005; Olson & Marshuetz, 2005; and Hollingworth, 2006b, for a description of how such binding might work for real-world objects in scenes). This interaction between spatial working memory and visual working memory may be particularly strong when remembering complex shape, when binding shapes to colors, or when binding colors to locations (Wood, 2011b), but relatively small when remembering colors that do not need to be bound to locations (Wood, 2011b).

Influence of feature context, or "ensemble statistics"—In addition to spatial context effects on item memory, it is likely that there are feature context effects as well. For instance, even in a display of squares with random colors, some displays will tend to have more "warm colors" on average, whereas others will have more "cool colors" on average,

and others still will have no clear across-item structure. This featural context, or "ensemble statistics" (Alvarez, 2011), could influence memory for individual items (e.g., Brady & Alvarez, 2011). For instance, say you remember that the colors were "warm" on average, but the test display contains a green item (Figure 3b). In this case, it is more likely that the green item is a new color, and it would be easier to detect this change than a change of similar magnitude that remained within the range of colors present in the original display.

Given that ensemble information would be useful for remembering individual items, it is important to consider the possibility that these ensemble statistics will influence item memory. Indeed, Brady and Alvarez (2011) have provided evidence suggesting that the representation of ensemble statistics influences the representation of individual items. They found that observers are biased in reporting the size of an individual item by the size of the other items in the same color set, and by the size of all of the items on the particular display. They proposed that this bias reflects the integration of information about the ensemble size of items in the display with information about the size of a particular item. In fact, using an optimal observer model, they showed that observers' reports were in line with what would be expected by combining information from both ensemble memory representations and memory representations of individual items (Brady & Alvarez, 2011).

These studies leave open the question of how ensemble representations interact with representations of individual items in working memory. The representation of ensemble statistics could take up space in memory that would otherwise be used to represent more information about individual items (as argued, for example, by Feigenson 2008 and Halberda, Sires and Feigenson, 2006), or such ensemble representations could be stored entirely independently of representations of individual items and integrated either at the time of encoding, or at the time of retrieval. For example, ensemble representations could make use of separate resource from individual item representations, perhaps analogous to the separable representations of real-world objects and real-world scenes (e.g., Greene & Oliva, 2009). Compatible with this view, ensemble representations themselves appear to be hierarchical (Haberman & Whitney, 2011), since observers compute both low-level summary statistics like mean orientation, and also object-level summary statistics like mean emotion of a face (Haberman & Whitney, 2009).

While these important questions remain for future research, the effects of ensemble statistics on individual item memory suggest several intriguing conclusions. First, it appears that visual working memory representations do not consist of independent, individual items. Instead, working memory representations are more structured, and include information at multiple levels of abstraction, from items, to the ensemble statistics of sub-groups, to ensemble statistics across all items, both in spatial and featural dimensions. Second, these levels of representation are not independent: ensemble statistics appear to be integrated with individual item representations. Thus, this structure must be taken into account in order to model and characterize the capacity of visual working memory. Limits on the number of features alone, the number of objects alone, or the number of ensemble representations alone, are not sufficient to explain the capacity of working memory.

Perceptual grouping and dependence between items—Other research has shown that items tend to be influenced by the other items in visual working memory, although such work has not explicitly attempted to distinguish influences due to the storage of individual items and influences from ensemble statistics. For example, Viswanathan, Perl, Bisscher, Kahana and Sekuler (2010; using Gabor stimuli) and Lin and Luck (2008; using colored squares) showed improved memory performance when items appear more similar to one another (see also Johnson, Spencer, Luck, & Schöner, 2009). In addition, Huang & Sekuler (2010) have demonstrated that when reporting the remembered spatial frequency of a Gabor

patch, observers are biased to report it as more similar to a task-irrelevant stimulus seen on the same trial. It was as if memory for the relevant item was "pulled toward" the features of the irrelevant item.

Cases of explicit perceptual grouping make the non-independence between objects even more clear. For example, Woodman, Vecera and Luck (2003) have shown that perceptual grouping helps determine which objects are likely to be encoded in memory, and Xu and Chun (2007) have shown that such grouping facilitates visual working memory, allowing more shapes to be remembered. In fact, even the original use of the change detection paradigm varied the complexity of relatively structured checkerboard-like stimuli as a proxy for manipulating perceptual grouping in working memory (Phillips, 1974), and subsequent work using similar stimuli has demonstrated that changes which affect the statistical structure of a complex checkerboard-like stimulus are more easily detected (Victor & Conte, 2004). The extent to which such improvements of performance are supported by low-level perceptual grouping – treating multiple individual items as a single unit in memory – versus the extent to which such performance is supported by the representation of ensemble statistics of the display in addition to particular individual items is still an open question. Some work making use of formal models has begun to attempt to distinguish these possibilities, but the interaction between them is likely to be complex (Brady & Tenenbaum, 2010; submitted).

Perceptual Grouping vs. Chunking vs. Hierarchically Structured Memory—

What is the relationship between perceptual grouping, chunking, and the hierarchically structured memory model we have described? Perceptual grouping and chunking are both processes by which multiple elements are combined into a single higher-order description. For example, a series of 10 evenly spaced dots could be grouped into a single line, and the letters F, B, and I can be chunked into the familiar acronym FBI (e.g., Cowan, 2001; Cowan et al. 2004). Critically, strong versions of perceptual grouping and chunking models posit that the resulting groups or chunks are the "units" of representation: if one part of the group or chunk is remembered, all components of the group or chunk can be retrieved. Moreover, strong versions of perceptual grouping and chunking models assume that the only limits on memory capacity come from the number of chunks or groups that can be encoded (Cowan, 2001).

Such models can account for some of the results reviewed here. For example, the influence of perceptual grouping on memory capacity (e.g., Xu & Chun, 2007) can be explained by positing a limit on the number of groups that can be remembered, rather than the number of individual objects. However, such models cannot directly account for the presence of memory limits at multiple levels, like the limits on both the number of objects stored and the number of features stored (Fougnie et al. 2010). Moreover, such models assume independence across chunks or groups and thus cannot account for the role of ensemble features in memory for individual items (Brady & Alvarez, 2011). Any model of memory capacity must account for the fact that groups or chunks themselves have sub-structure, that this sub-structure causes limits on capacity, and that we simultaneously represent both information about individual items and ensemble information across items. A hierarchically structured memory model captures these aspects of the data by proposing that information is maintained simultaneously at multiple, interacting levels of representation, and our final memory capacity is a result of limits at all of these levels.

Conclusion—

Taken together, these results provide significant evidence that individual items are not represented independent of other items on the same display, and that visual working memory stores information beyond the level of individual items. Put another way, every display has multiple levels of structure, from the level of feature representations to

individual items to the level of groups or ensembles, and these levels of structure interact. It is important to note that these levels of structure exist, and vary across trials, even if the display consists of randomly positioned objects that have randomly selected feature values. The visual system efficiently extracts and encodes structure from the spatial and featural information across the visual scene, even when, in the long run over displays, there may not be any consistent regularities. This suggests that any theory of visual working memory that specifies only the representation of individual items or groups cannot be a complete model of visual working memory.

The effects of stored knowledge on visual working memory

Most visual working memory research requires observers to remember meaningless, unrelated items, such as randomly selected colors or shapes. This is done to minimize the role of stored knowledge, and to isolate working memory limitations from long-term memory. However, in the real-world, working memory does not operate over meaningless, unrelated items. Observers have stored knowledge about most items in the real world, and this stored knowledge constrains what features and objects we expect to see, and where we expect to see them. The role of such stored knowledge in modulating visual working memory representations has been controversial. In the broader working memory literature, there is clear evidence of the use of stored knowledge to increase the number of items remembered in working memory (Ericsson, Chase & Faloon, 1980; Cowan, Chen & Rouder, 2004). For example, the original experiments on chunking were clear examples of using stored knowledge to recode stimuli into a new format to increase capacity (Miller, 1956) and such results have since been addressed in most models of working memory (e.g., Baddeley, 2000). However, in visual working memory, there has been less work towards understanding how stored knowledge modulates memory representations and the number of items that can be stored in memory.

Biases from stored knowledge—One uncontroversial effect of long-term memory on working memory is that there are biases in working memory resulting from prototypes or previous experience. For example, Huang and Sekuler (2010) have shown that when reporting the spatial frequency of a gabor patch, observers are influenced by stimuli seen on previous trials, tending to report a frequency that is pulled toward previously seen stimuli (see Spencer & Hund, 2002 for an example from spatial memory). Such biases can be understood as optimal behavior in the presence of noisy memory representations. For example, Huttenlocher et al. (2000) found that observers memory for the size of simple shapes is influenced by previous experience with those shapes; observers' reported sizes are again 'attracted' to the sizes they have previously seen. Huttenlocher et al. (2000) model this as graceful errors resulting from a Bayesian updating process -- if you are not quite sure what you've seen, it makes sense to incorporate what you expected to see into your judgment of what you did see. In fact, such biases are even observed with real-world stimuli, for example, memory for the size of a real-world object is influenced by our prior expectations about its size (Hemmer & Steyvers, 2009; Konkle & Oliva, 2007). Thus, visual working memory representations do seem to incorporate information from both episodic long-term memory and from stored knowledge.

Stored knowledge effects on memory capacity—While these biases in visual working memory representations are systematic and important, they do not address the question of whether long-term knowledge can be used to store more items in visual working memory. This question has received considerable scrutiny, and in general it has been difficult to find strong evidence of benefits of stored knowledge on working memory capacity. For example, Pashler (1988) found little evidence for familiarity modulating change detection performance. However, other methods have shown promise for the use of

long-term knowledge to modulate visual working memory representations. For example, Olsson and Poom (2005) used stimuli that were difficult to categorize or link to previous long-term representations, and found a significantly reduced memory capacity, and observers seem to perform better at working memory tasks with upright faces (Curby & Gauthier, 2007; Scolari, Vogel & Awh, 2008), familiar objects (see Experiment 2, Alvarez & Cavanagh, 2004), and objects of expertise (Curby et al 2009) than other stimulus classes. In addition, children's capacity for simple colored shapes seems to grow significantly over the course of childhood (Cowan et al., 2005), possibly indicative of their growing visual knowledge base. Further, infants are able to use learned conceptual information to remember more items in a working memory task (Feigenson & Halberda, 2008).

However, several attempts to modulate working memory capacity directly using learning to create new long-term memories showed little effect of learning on working memory. For example, a series of studies has investigated the effects of associative learning on visual working memory capacity (Olson & Jiang, 2004; Olson, Jiang & Moore, 2005), and did not find clear evidence for the use of such learned information to increase working memory storage. For example, one study found evidence that learning did not increase the amount of information remembered, but that it improved memory performance by redirecting attention to the items that were subsequently tested (Olson, Jiang & Moore, 2005). Similarly, studies directly training observers on novel stimuli have found almost no effect of long-term familiarity on change detection performance (e.g., Chen, Eng & Jiang, 2006).

In contrast to this earlier work, Brady, Konkle and Alvarez (2009) have recently shown clear effects of learned knowledge on working memory. In their paradigm, observers were shown standard working memory stimuli in which they had to remember the color of multiple objects (Figure 4a). However, unbeknownst to the observers, some colors often appeared near each other in the display (e.g., red tended to appear next to blue). Observers were able to implicitly learn these regularities, and were also able to use this knowledge to encode the learned items more efficiently in working memory, representing nearly twice as many colors (~5–6) as a group who was shown the same displays without any regularities (Figure 4b). This suggests that statistical learning enabled observers to form compressed, efficient representations of familiar color pairs. Furthermore, using an information-theoretic model, Brady, Konkle and Alvarez (2009) found that observers' memory for colors was compatible with a model in which observers have a fixed capacity in terms of information (bits), providing a possible avenue for formalizing this kind of learning and compression.

It is possible that Brady et al. (2009) found evidence for the use of stored knowledge in working memory coding because their paradigm teaches associations between items, rather than attempting to make the item's themselves more familiar. For instance, seeing the same set of colors for hundreds of trials might not improve the encoding of colors or shapes, because the visual coding model used to encode colors and shapes has been built over a lifetime of visual experience that cannot not be overcome in the time-course of a single experimental session. However, arbitrary pairings of arbitrary features are unlikely to compete with previously existing associations, and might therefore lead to faster updating of the coding model used to encode information into working memory. Another important aspect of the Brady et al. (2009) study is that the items that co-occurred were always perceptually grouped. It is possible that compression only occurs when the correlated items are perceptually grouped (although learning clearly functions without explicit perceptual grouping, e.g., Orbán, Fiser, Aslin & Lengyel, 2008).

Conclusion—Observers have stored knowledge about most items in the real world, and this stored knowledge constrains what features and objects we expect to see, and where we expect to see them. There is significant evidence that the representation of items in working

memory is dependent on this stored knowledge. Thus, items for which we have expertise, like faces, are represented with more fidelity (Curby & Gauthier, 2007; Scolari, Vogel & Awh, 2008), and more individual colors can be represented after statistical regularities between those colors are learned (Brady et al. 2009). In addition, the representation of individual items are biased by past experience (e.g., Huang and Sekuler, 2010; Huttenlocher et al. 2000). Taken together, these results suggest that the representation of even simple items in working memory depends upon our past experience with those items and our stored visual knowledge.

Visual working memory conclusion

A great deal of research on visual working memory has focused on how to characterize the capacity of the system. We have argued that in order to characterize working memory capacity, it is important to take into account both the number of individual items remembered, and the fidelity with which each individual item is remembered. Moreover, it is necessary to specify what the units of working memory storage are, how multiple units in memory interact, and how stored knowledge affects the representation of information in memory. In general we believe theories and models of working memory must be expanded to include memory representations that go beyond the representation of individual items and include hierarchically-structured representations, both at the individual item level (hierarchical feature-bundles), and across individual items. There is considerable evidence that working memory representations are not based on independent items, that working memory also stores ensembles that summarize the spatial and featural information across the display, and further, that there are interactions between working memory and stored knowledge even in simple displays.

Moving beyond individual items towards structured representations certainly complicates any attempt to estimate working memory capacity. The answer to how many items can you hold in visual working memory depends on what kind of items you are trying to remember, how precisely they must be remembered, how they are presented on the display, and your history with those items. Even representations of simple items have structure at multiple levels. Thus, models that wish to accurately account for the full breadth of data and memory phenomena must make use of structured representations, especially as we move beyond colored dot objects probed by their locations towards items with more featural dimensions or towards real-world objects in scenes.

VISUAL LONG-TERM MEMORY

Before discussing the capacity of long-term memory, it is important to make the distinction between visual long-term memory and stored knowledge. By "visual long-term memory," we refer to the ability to explicitly remember an image that was seen previously, but which has not been continuously held actively in mind. Thus, visual long-term memory is the passive storage and subsequent retrieval of visual *episodic* information. By "stored knowledge," we refer to the pre-existing visual representations that underlie our ability to perceive and recognize visual input. For example, when we first see an image, say of a red apple, *stored knowledge* about the visual form and features of apples in general enables us to recognize the object as such. If we are shown another picture of an apple hours later, *visual long-term memory* enables us to decide whether this is the exact same apple we saw previously.

While working memory is characterized by its severely limited capacity, long-term memory is characterized by its very large capacity: people can remember thousands of episodes from their lives, dating back to their childhood. However, in the same way that working memory capacity cannot be characterized simply in terms of the number of items stored, the capacity

of long-term memory cannot be fully characterized by estimating the number of individual episodes that can be stored. Long-term memory representations are highly structured, consisting of multiple levels of representation from individual items to higher-level conceptual representations. Just as we proposed for working memory, these structured representations should be taken into account, both when quantifying and characterizing the capacity of the system, and when modeling memory processes such as retrieval.

Generally, work in the broader field of long-term memory has not emphasized the nature of stored representations, and has focused instead on identifying different memory systems (e.g., declarative vs. nondeclarative, episodic vs. semantic) and understanding the processing stages of those systems, particularly the encoding and retrieval of information (e.g. Squire, 2004). As is the case in the domain of working memory, theories of long-term memory encoding and retrieval are typically developed independent of what particular information is being stored and what particular features are used to represent stored items. A typical approach is to model memory phenomena that result from manipulations of timing (e.g. primacy and recency, rate of presentation), study procedure (e.g. massed or spaced presentation, the number of re-study events), and content-similarity (e.g. the fan-effect, category-size effect, category-length effect). For example, models of memory retrieval and storage that capture many of these phenomena have been proposed by Brown, Neath, and Chater (2007) and Shiffrin and Steyvers (1997).

Critically, in order to account for the range of performance across these manipulations, such models have postulated a role for some form of 'psychological similarity' between items, like how many features they share (e.g., Eysenck, 1979; Nairne, 2006; Rawson & Van Overschelde, 2008; Schmidt, 1985; von Restorff, 1933; see Shiffrin & Steyvers, 1997). For example, the effectiveness of a retrieval cue is based on the extent to which it cues the correct item in memory without cuing competing memories. Put simply, if items share more features, they interfere more in memory, leading to worse memory performance. Thus, in the domain of long-term memory, it is well-known that the nature of the representation, such as the features used in encoding the stimulus, are essential for predicting memory performance.

Clearly the more complete our model of the structure and content of long-term memory representations, the more accurately we will be able to model retrieval processes. Thus, the rich, structured nature of long-term memory representations, and the role of distinctiveness in long-term memory retrieval, pose challenges to quantifying and characterizing the capacity of visual long-term memory.

Here we review recent work that has examined these representation-based issues within the domain of visual long-term memory: What exactly is the content of the representations stored in visual long-term memory? What features of the incoming visual information are critical for facilitating successful memory for those items? By assessing both the quantity and the fidelity of the visual long-term memory representations, we can more accurately quantify the capacity of this visual-episodic-memory system. By measuring the content of visual long-term memory representations, and what forms of psychological similarity cause this information to be forgotten, we can use memory as a probe into the structure of stored knowledge about objects and scenes.

The fidelity of visual long-term memory

Quantifying the number of items observers can remember—In the late 60's and 70's, a series of landmark studies demonstrated that people have an extraordinary capacity to remember pictures (Shepard, 1967; Standing, Conezio, Haber, 1970; Standing, 1973). For example, Shepard (1967) showed observers ~600 pictures for 6 seconds each. Afterwards, he tested memory for these images with a two-alternative forced-choice task where

participants had to indicate which of two images they had seen (Figure 5a). He found that observers could correctly indicate which picture they had seen almost perfectly (98% correct). In perhaps the most remarkable study of this kind, Standing (1973), showed observers 10,000 color photographs scanned from magazines and other sources, and displayed them one at a time for 5 seconds each. The 10,000 pictures were separated into distinct thematic categories (e.g. cars, animals, single-person, two-people, plants, etc.), and within each category only a few visually distinct exemplars were selected. Standing found that even after several days of studying images, participants could indicate which image they had seen with 83% accuracy. These results demonstrate that people can remember a surprisingly large number of pictures, even hours or days after studying each image just once.

By correcting for guessing it is possible to estimate how many images observers must have successfully recognized to achieve a given level of performance. Standing (1973) found that when shown 100 images, observers performance suggested they remembered 90 of these images; when shown 1,000 images, performance suggested memory for 770; and when shown the full set of 10,000, their performance indicated memory for ~6,600 of the 10,000 images they had been shown. Extrapolating the function relating the number of items presented to the number of items recalled suggested no upper bound on the number of pictures that could be remembered (although see Landauer, 1986 for a possible model of fixed memory capacity in these studies; see also Dudai, 1997). These empirical results and models of memory performance have lead many to infer the number of visual items that can be stored in long-term memory is effectively unlimited, with memory performance depending primarily on how distinctive the information is rather than how many items are to be remembered.

Reasons for suspecting low-fidelity representations—These large-scale memory studies always used items that were as semantically and visually distinct as possible. For example, during the study phase there might be a single wedding scene, a single carnival scene, a single restaurant scene, etc; then, at test, an observer might see either the original wedding scene, or a park (e.g., Figure 5a). Under these conditions, to accurately indicate which of two images was studied, an observer would only need to remember the semantic gist of the images. This led many researchers in the field to assume that visual long-term memory stores relatively impoverished representations of each item, perhaps just a gist-like representation capturing the basic category, event, or meaning of the image along with a few specific details (Chun, 2003; Simons & Levin, 1997; Wolfe, 1998).

Influential studies demonstrating 'change blindness' also provided evidence suggesting that people likely only store gist-like representations of images (e.g. Rensink, 2000). Change blindness studies demonstrated that changes to an object part, or even large changes to a surface within a scene, often go undetected if the visual transients are masked (e.g., Rensink et al., 1997; Simons & Levin, 1997). This is even true in cases when memory demands are limited, for example when observers only need to retain information from one scene for a short amount of time before being presented with the altered scene. Together with the large-scale memory studies, change-blindness led to the widely accepted idea that memory representations for real-world stimuli are impoverished and lacked visual detail (see Hollingworth, 2006a for a review).

Evidence of high-fidelity long-term memory representations—A number of recent studies have overturned the assumption that representations of objects and scenes are sparse and lack detail. Experiments using both change detection paradigms (e.g., Mitroff, Simons & Levin, 2004; review by Simons & Rensink, 2005) and long-term memory tasks (e.g.,

Hollingworth, 2004; Brady et al. 2008), have demonstrated that visual memory representations often contain significant detail.

For example, a series of studies by Hollingworth and Henderson (2002) demonstrated that, after briefly attending to objects within a scene, memory for those objects was more visually detailed than just the category of the object, even after viewing 8–10 other objects (Hollingworth & Henderson, 2002; Hollingworth, 2004; Hollingworth, 2005). In fact, people maintained object details sufficient to distinguish between exemplars (this dumbbell vs. that dumbbell) and viewpoints (an object from this view, vs. the same object rotated 90 degrees) with above-chance recognition for around 400 studied objects intervening between initial presentation and memory test (Hollingworth, 2004), or even after a delay of 24 hours (Hollingworth, 2005). These results were the first to demonstrate that observers are capable of storing more than just the semantic category or gist of real-world objects over significant durations with a relatively large quantity of items.

To further assess the fidelity of visual long-term memory representations using a large-scale memory paradigm more closely matched to Standing (1973), Brady, Konkle, Alvarez, and Oliva (2008) had observers view 2500 categorically distinct objects, one at a time, for 3 seconds each, over the course of more than 5 hours. At the end of this study session, observers performed a series of 2-alternative-forced-choice tests that probed the fidelity of the memory representations by varying the relationship between the studied target item and the new foil item (Figure 5c-e). In the *novel* condition, the foil was categorically different from all 2500 studied objects. Success on this type of test required memory only for the semantic category of studied items, as in Standing (1973) and Shepard (1967). In the *exemplar* condition, the foil was a different exemplar from the same basic category as the target (e.g., if the target was a shoe, the foil would be a different kind of shoe). If only the semantic category of the target object was remembered, observers would fail on this type of test. To choose the right exemplar, observers had to remember specific visual details about the target item. In the *state* condition, the foil was the same object as the target, except it was in a different state or pose (e.g., if the target was a shoe with the laces tied, the foil could be the same shoe with the laces untied). To choose the target on state trials, observers would have to remember even more specific visual details about the target item.

As expected based on earlier studies, observers performed at 92% accuracy on the novel test, indicating that they had encoded at least the semantic category of thousands of objects. Surprisingly, observers could successfully perform the exemplar and state tests nearly as well (87% and 88%, respectively). For example, observers could confidently report whether a cup of orange juice they had seen was totally full or only mostly full of juice with almost 90% accuracy. It is important to note that observers did not know which of the 2500 studied items would be tested nor which particular object details had to be remembered for a particular item (e.g. category-level, exemplar-level, or state-level information), indicating that observers were remembering a significant amount of object detail about each item. Thus, visual long-term memory is capable of storing not only thousands of objects, but it can store thousands of *detailed* object representations.

One important difference between the work of Brady et al. (2008) and Standing (1973) is the complexity of the stimuli. In Brady et al. (2008), observers saw individual objects on a white background, whereas in Standing and Shepard's seminal studies, observers saw full scenes (magazine clippings). Thus, it is possible that observers can store detailed *object* representations (as in Brady et al., 2008), but that their memory performance for exemplar-level differences in natural scenes would contain markedly less detail. Recent work has shown this is not the case: Using a paradigm much like that of Brady et al. (2008), Konkle, Brady, Alvarez and Oliva (2010a) demonstrated that thousands of scenes can be

remembered with sufficient fidelity to distinguish between different exemplars of the same scene category (e.g., this garden or that garden, see Figure 5b). Furthermore, performance on these scene stimuli was nearly identical to performance with objects (Konkle, Brady, Alvarez & Oliva, 2010a; Konkle, Brady, Alvarez & Oliva, 2010b).

Conclusion—Quantifying the capacity of a memory system requires determining both the number of items that can be stored and the fidelity with which they are stored. The results reviewed here demonstrate that visual long-term memory is capable of storing not only thousands of objects, but store thousands of detailed object and scene representations (e.g., Brady et al. 2008; Konkle et al. 2010a). Thus, the capacity of visual long-term memory is greater than assumed based on the work of Standing (1973) and Shepard (1967). However, given that we previously believed the capacity of visual long-term memory was "virtually unbounded," what is gained by showing the capacity is even greater than we thought? It is certainly informative to know that our intuitions about the fidelity of our visual long-term memory are incorrect: after studying thousands of unique pictures and tested with very "psychologically similar" foils, we will be much closer to perfect performance than chance performance. It is also valuable to know that, while in everyday life we may often fail to notice the details of objects or scenes (Rensink, O'Regan & Clark, 1997; Simons & Levin, 1997), this does not imply that our visual long-term memory system cannot encode and retrieve a huge amount of information, including specific visual details. Perhaps most importantly, earlier models assumed that visual long-term memory representations lacked detail and were gist-like and semantic in nature. Discovering that visual long-term memory representations can contain significant object-specific detail challenges this assumption, and suggests that visual episodes leave a more complete memory trace that includes more 'visual' or perceptual information.

Effects of Stored Knowledge on Visual Long-Term Memory

Stored knowledge provides a coding model for representing incoming information—Over a lifetime of visual experience, our visual system builds a storehouse of knowledge about the visual world. How does this stored knowledge affect the ability to remember a specific visual episode? In the case of visual working memory, we proposed that stored knowledge provides the coding model used to represent items in working memory. As we learn new information and update our stored knowledge, we update how we encode subsequent information (e.g., remembering more colors after learning regularities in which colors appear together: Brady, Konkle & Alvarez, 2009). We propose that stored knowledge plays the same role in visual long-term memory, providing the coding model used to encode incoming visual information and represent visual episodes.

One way to conceive of this coding model is as a multidimensional feature space, with an axis for each feature dimension. Of course this would be a massive feature space, but for illustration suppose there were some more perceptual feature dimensions (size, shape, and color), and some conceptual dimensions (living/nonliving, fast/slow). For any new incoming visual input, the visual system would extract information along these existing feature dimensions, creating a memory trace that can be thought of as a point in this multi-dimensional feature space. Thus, memory for a particular car is represented as a single point in this space (e.g., a race car might be encoded as small, aerodynamic, red, nonliving, and fast). Memory traces for different cars will fall relatively close to each other in this feature space. Memory traces for similar kinds of objects, like tractors might be nearby cars in this feature space, whereas memory traces for different kinds of objects, like cats, might be far away from cars in this feature space.

Most likely the coding model we derive from our past visual experience is considerably more complex than a single multidimensional space. In higher-level domains like categorization and induction, models of background knowledge based on more structured representations are required to fit human performance (Kemp & Tenenbaum, 2009; Tenenbaum, Griffiths & Kemp, 2006). For example, rather than representing animals in a single multidimensional space, observers seem to have multiple structured representations of animals: some kinds of inference draw on a tree structure expressing animals' biological relatedness, whereas some inferences draw on a food web expressing which animals are likely to eat which other animals (Kemp & Tenenbaum, 2009).

Visual background knowledge is likely to be similarly complex, perhaps based on a hierarchy of features ranging from generic *perceptual* features like color and orientation to mid-level features that have some specificity to particular object classes (e.g. Ullman, Vidal-Naquet, & Sali, 2002) to very high-level *conceptual* features that are entirely object-category specific (e.g. Ullman, 2007) (see Figure 6). In fact, modern models of object recognition propose that stored object knowledge consists of feature hierarchies (e.g. Epshtein & Ullman, 2005; Ommer & Buhman, 2010; Riesenhuber & Poggio, 1999; Torralba, Murphy, & Freeman, 2004; Ullman, 2007). For example, Ullman (2007) proposed that objects are represented by a hierarchy of image fragments: e.g. small image fragments of car parts combine to make larger car fragments, which further combine to make a car. In this sense, stored knowledge about different object concepts can be cached out in a hierarchy of visual features, which may be extracted and stored in visual long-term memory.

Furthermore, according to these computational models, precisely what features are represented in the hierarchy will depend on the task the system must perform (Ullman, 2007): to recognize a face as a face, the model learns one set of features, but to do a finer level of categorization (e.g. that this particular face is George Alvarez) larger features and feature combinations have to be learned by the model (see also Schyns & Rodet, 1997). Thus with increasing knowledge about particular exemplars and subordinate category structure, different kinds of visual features may be created in the visual hierarchy. This could explain why even memory for putatively "visual" information is dependent on conceptual structure, as high-level visual representations themselves are likely shaped by category knowledge.

Regardless of the exact format, this visual background knowledge provides the basis of the coding model for new visual episodes, by defining either the axes of the multidimensional feature space or the particular structured representation (e.g., the features at each level in the hierarchy). What is not known is what the relevant perceptual and conceptual features are for visual long-term memory, and which feature dimensions are most important for retrieving items from this representational space. In the next sections, we review experiments that address the role of stored knowledge in memory encoding, the role of perceptual and conceptual features in visual long-term memory, and the effects of learning on creating new features with which to represent objects.

Role of the "conceptual hook" for supporting visual memory—Several studies have shown that memory for visual images is better when those images are semantically labeled and recognized than when the same images are not labeled and recognized (Wiseman and Neisser, 1974; Koutstaal et al., 2003; see also Bower, Karlin, & Dueck, 1975). For example, Wiseman and Neisser (1974) presented observers with two-tone ambiguous face images (Mooney faces) and asked observers to judge whether or not there was a face present. While all images actually contained faces, subsequent memory was better for images that were recognized as faces relative to images that did not make contact with this organizing concept. Further, there was individual variability in whether or not a

particular image was recognized as a face. Memory for a given image was better in people who saw it as a face compared to those who saw that same image without recognizing it as a face. This provides elegant evidence for the importance of concepts in visual memory while controlling for all low-level visual features. In further support of a "conceptual hook", memory for ambiguous shapes is improved when a disambiguating semantic label is provided during study (Koutstaal et al., 2003, Experiment 1, Figure 7a), and memory for real-world objects is better than memory for perceptually rich but meaningless objects (e.g., Koutstaal et al., 2003, Experiment 2).

These studies show that connecting to existing knowledge - versus not connecting at all - is critical for successful visual long-term memory. Connecting with stored knowledge likely improves memory because it provides a rich and structured coding scheme consisting of both of perceptual and conceptual feature dimensions. But which of these dimensions are important for supporting visual long-term memory? Do all perceptual and conceptual features contribute equally, or are some features more important? To address these questions, it is possible to vary the similarity of items along different perceptual and conceptual dimensions. To the extent that similarity along a particular feature dimension impairs memory, we would conclude that "crowding" along that feature dimension causes interference in memory, and therefore that the feature is important for supporting memory.

Konkle, Brady, Alvarez, & Oliva (2010b) used this approach in a large-scale memory study. Observers viewed 2,800 objects from over 200 distinct categories, where the number of exemplars present from each category varied from 1 to 16. At test, observers indicated which of 2 exemplars they previously studied, requiring detailed memory representations. By varying the number of exemplars per category, they tested the impact of category-information in visual memory. If the category label is a critical feature supporting visual long-term memory, then with more studied exemplars from a category, there should be more interference in memory and worse performance. This effect was observed, but the drop in performance with each doubling of the number of studied exemplars was only 2%. Overall, memory performance was remarkably high (84% with 16 exemplars from a category in mind and thousands of other objects), suggesting that while category information matters for visual memory, it is far from the sole feature supporting detailed visual memory.

Konkle et al. (2010b) next examined the interference effects for each object category. Some categories of objects showed more interference in memory than others. To examine which feature dimensions account for this variation across object categories, a variety of similarity rankings were obtained for each object category. Perceptual similarity was measured separately for color, shape, and for overall visual appearance among exemplars. A conceptual similarity measure captured whether there were few or many different kinds of a particular category (e.g. there are many kinds of cars but few kinds of bean bag chairs) (see Figure 7b). They observed that categories with more conceptually distinctive exemplars showed less interference in memory than categories with conceptually similar exemplars. Surprisingly, they also found that the perceptual measures did not predict memory interference: interference in memory was similar for categories with many perceptually distinctive exemplars and categories with perceptually similar exemplars. These results demonstrated that stored knowledge of both basic-level object categories and subordinate categories is a critical part of the visual long-term memory coding-model, suggesting that distinctiveness along categorical dimensions is necessary for successful memory retrieval. These categorical dimensions appear to provide "conceptual hooks" which enable the recovery of a complete memory trace that includes not only semantic abstract information, but also more perceptual information about object details.

Learning new features and expertise—Stored knowledge about visual concepts provides the coding model for representing incoming information, but it is also constantly changing based on incoming experience and learning. Moreover, the specific features an individual learns will depend on their prior experience. For example, in a category learning experiment in which participants learned to classify different Martian cells, Schyns and Rodet discovered that the order in which participants learned the cells affected which features were learned and used to classify the cells (Rodet and Schyns, 1994; Schyns and Rodet, 1997). In other words, different histories of categorization generate different feature spaces that are used to encode similarities and differences between items. Schyns and Rodet suggested that observers build 'functional features' specifically designed to support performance in a task. If this is the case, then it is possible to gain important insights regarding the format of stored representations from the categorization literature. One conceptualization of these category learning experiments is that they help build an enhanced coding model for learned stimuli. In support of this idea, during typical object recognition experiments, when incoming information makes contact with stored knowledge people name the object at the basic level most quickly (Mervis & Rosch, 1981), but experts will more quickly name objects at subordinate category levels (Joliceour, 1985), suggesting that they extract different features during encoding.

If expertise leads to an enhanced coding model, allowing observers to extract richer and or more distinctive features from an input, then experts should have increased memory capacity for these items (Ericsson & Chase, 1982). Evidence for this has been found when comparing experts and novices' memory for chess configurations (Chase & Simon, 1973; de Groot, 1966) and for baseball-related knowledge (Voss, Vesonder, & Spilich, 1980), in addition to the results with working memory for faces and objects of expertise reviewed earlier (Curby & Gauthier, 2007; Curby et al. 2009). However, there is debate about the nature of the features learned with visual expertise: for example, whether they are combinations of low-level features in a strict hierarchical way, or whether more holistic features can be generated from the earliest stages of representation (see Schyns, Goldstone, & Thibaut, 1998). On either account, the coding-model employed to extract features of the input is changed with learning and expertise.

Even though stored knowledge is different across people and the visual coding model depends on an individuals' history of learning and degree of expertise, it is still possible to generalize across people when studying memory capacity. For example, Konkle et al. (2010b) measured the conceptual distinctiveness of a set of exemplars from different object categories (i.e. how many different kinds of cars and bowties were present in a given set), which is a dimension with clear individual differences in stored knowledge. Nevertheless, ratings from one set of observers were able to predict the memory performance from another set of observers, suggesting a general convergence of basic stored knowledge about real-world objects.

Consequences of coding models: systematic biases of object details in memory—Stored knowledge, specifically about category and subordinate category structure, supports the ability to store and retrieve detailed visual long-term memory representations. However, encoding information with respect to existing knowledge can lead to systematic biases, called constructive memory errors (Bartlett, 1932; Brewer & Treyens, 1981; Roediger & McDermott, 1995). Several classic studies have demonstrated that the details retrieved from visual long-term memory are not necessarily veridical. In one such study, encoding a two circles connected by a line as a "dumbbell" or as "eyeglasses" leads to systematic biases when later drawing the item from memory (Carmichael, Hogan, & Walter, 1932), with a thicker straight connecting line when encoded as a dumbbell, and a thinner more curved connecting line when encoded as eye glasses. Naming an object during

encoding has been argued to shift the representation to be more prototypical (Lupyan, 2008; see also Koutstaal et al, 2003). However, these systematic biases in visual memory can be thought of as graceful errors--any noise in the representation leads the representation to be pulled towards prototypical values, which is what an optimal memory system should do (Huttenlocher, Hedges, & Vevea, 2000).

Thus, bringing to bear conceptual knowledge about objects and scenes during encoding does help support visual detail in memory, but does not enable a more 'photographic-like' memory--rather it likely enables memory for visual details to be connected to, and integrated with, meaningful dimensions of the object or scene. This also implies that measuring systematic biases in visual long-term memory representations can be used as a tool to infer what coding-model observers used to encode the initial episode (e.g., Castel, McCabe, Roediger, & Heitman, 2007). Thus errors in the fidelity of visual long-term memory can be used as a way to discover dimensions of stored knowledge.

Conclusion—What do these results about the role of stored knowledge imply about the capacity of visual long-term memory? All of these studies demonstrate that the capacity of visual long-term memory is critically dependent upon stored knowledge - the coding-model that we use to represent each image. Thus in order to predict memory performance for any given bit of visual information, it is necessary to first characterize what is already 'builtin' to the visual knowledge base for encoding that information. This explains why we are remarkable at remembering natural scenes and real-world objects (e.g., Brady et al. 2008; Konkle et al. 2010a; Hollingworth, 2004; Shepard, 1967; Standing, 1973), for which we have a massive stored knowledge base, and why we cannot not even attempt to remember thousands of random colored dot displays (like those typically used in visual working memory tasks), for which we have no pre-existing knowledge, category structures, or differentiating semantic associations. Research examining visual long-term memory for real-world images has shown that the coding model we use to retrieve representations from memory gives more weight to conceptual features than perceptual features--being perceptually rich and distinctive is not sufficient to support visual long-term memory (e.g., Konkle et al. 2010b; Koutstaal et al. 2003). However, the representation that can be retrieved from visual long-term memory is far more visually detailed than just a category label or gist representation. One possible interpretation of this finding is that visual long-term memory representations are hierarchically structured, with conceptual or category-specific features at the top of the hierarchy, and perceptual or more category-general features at lower levels of the hierarchy. On this view, memory retrieval operates over the top levels of the hierarchy, which includes categorical labels, but successful retrieval activates the full, hierarchical memory trace including lower perceptual features.

Memory for objects within scenes

Just as stored knowledge allows us to bring a rich coding model to represent individual objects, we also have stored knowledge about relationships between objects and other objects, and between objects and the surrounding scene. While a review of research on the format of scene representations is beyond the scope of this review (see Luck & Hollingworth, 2008), here we highlight several key studies that examine the relationship of an object to a scene (e.g. scene schemas which reflect the probability of finding an apple in a kitchen or a bedroom). Combined, these studies illustrate how the rich, structured nature of scene knowledge impacts visual long-term memory representations for objects.

Stored knowledge about scene information has been given various labels, including schemata (Biederman et al., 1982; Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976), scripts (Schank, 1975), frames (Minsky, 1975), and more recently context

frames (Bar and Ullman, 1996; Bar, 2004). In all of these characterizations, stored knowledge about scenes provides predictions about the likely objects to be found in the scenes and the likely positions of those objects in the scene, as well as object relations, relative sizes and positions, and co-occurrence statistics. This knowledge is activated rapidly, even by presentations of single objects that have strong contextual associations (Bar, 2004), and thus can influence the processing and encoding of object information at very early stages of information processing (e.g., Bar et al. 2006).

We note that scenes themselves can also be the "items" of memory (e.g. in Standing 1973, Konkle et al. 2010a), and some have argued that scenes have their own object-less representational basis (e.g., Oliva & Torralba, 2001; Greene & Oliva, 2009; Greene & Oliva, 2010), neural substrates (Epstein & Kanwisher, 1998), and category structures (Tversky & Hemenway, 1983; Xiao, Hayes, Ehinger, Oliva & Torralba, 2010; see also Henderson & Hollingworth, 1999). Indeed, similar degrees of category interference in memory for objects and scenes (Konkle et al, 2010a) suggests that scenes may be thought of as entities at a similar level of abstraction as objects. The nature of the representations of scenes as independent entities and not as background context warrants further study. For the scope of this review, however, we will limit our discussion to scenes as part of ensemble or contextual information. Here we review some of the work showing that this information influences the memory representations of individual objects, by (i) serving as a better retrieval cue for the initial studied episode, (ii) by directing attention to distinctive features of a scene, and (iii) by providing reasonable guesses given uncertainty in memory.

Scene context as a retrieval cue—Several studies have shown that the presence of a background scene helps memory for both the features and spatial position of individual objects in the scene. For example, Hollingworth (2006b) tested whether the presence of background scene information influenced memory for object details when the scene itself was task-irrelevant. Observers studied a scene with many objects in it for 20 seconds. At test, memory for one object was probed, requiring observers to remember the specific exemplar or viewpoint of the cued object. Memory for these object details was better when the objects were tested with their scene backgrounds present. Memory for spatial positions of objects in a scene is also better when the scene is present at retrieval: e.g. memory for an object position (on the screen) is facilitated when the object reappears in the scene at test (Mandler & Johnson, 1976; Hollingworth, 2007) (see Figure 8a). This effect is stronger when the scene information is meaningful and coherent compared to when it is incoherently organized (Mandler & Johnson, 1976; Mandler & Parker, 1976; Mandler & Ritchey, 1977). However, even in the case of meaningful configurations of unrelated objects, probing memory for an object with the same context present shows benefits over conditions with changed context (Hollingworth, 2007).

While these studies were done at short time scales, this idea also holds in the broader long-term memory literature, which shows the importance of context in memory retrieval (the encoding-specificity principle: Tulving & Thomson, 1973). The better the match between the study and retrieval context, the better memory for items, even if the context is irrelevant to the specific items being remembered. This effect was famously demonstrated using memory for word lists studied and tested by scuba divers on the beach or underwater: memory performance is improved when word lists are studied and tested both underwater, for example, compared to studying underwater and testing on the beach; Godden & Baddeley, 1975). These recent visual long-term memory studies add the idea that context can facilitate retrieval of even relatively detailed object information (Hollingworth, 2006b).

The effects of background context or ensemble information likely arise from natural experience, where items are always experienced and learned in a context. These studies

show that benefits for memory persist even when the context is less meaningful (spatially incoherent) or lacks 3D structure and is simply a configuration of items on the screen. This suggests that the representation of items in scenes is never entirely independent of the surrounding scene context, just as in working memory the representation of individuals is not independent of the ensemble statistics of the display (e.g., Brady & Alvarez, 2011; Jiang, Olson, & Chun, 2000).

Semantic consistency, schemas, and encoding distinctive details—How is memory for object information effected when objects are meaningfully related to the surrounding scene? A number of studies have tested memory for objects in a scene as a function of semantic consistency (Brewer & Treyens, 1981; Friedman, 1979; Hollingworth & Henderson, 2000; Hollingworth & Henderson, 2003; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989; Lampinen, Copeland, & Neuschatz, 2001). These studies show that *inconsistent* items are remembered better, e.g. memory for the presence of a coffeemaker is higher when that coffeemaker is shown in a farmyard compared to in a kitchen, and these items are fixated longer during study (Friedman, 1979) (Figure 8b). This benefit for inconsistent items may seem at odds with the claim that memory is supported by stored knowledge - if we have no stored knowledge about coffeemakers *in* farmyards, how can we remember them? However, these results are in fact quite consistent with what we would expect from an efficient encoding system that knows about both coffeemakers and farmyards.

Stored knowledge about scenes contains a wealth of information, including what objects are likely to appear in different kinds of scenes and where they should appear within those scenes (Biederman et al., 1982; Hock et al, 1978; Mandler & Johnson, 1976; Bar & Ullman, 1996). For example, stored knowledge about farmyards can assign probabilities on which animals and objects are likely to be in that scene, and a coffeemaker is an extremely low probability object. If the goal is to encode this image so that we can retrieve it later, attention should be directed to the features which are least typical, because those features must be specifically encoded in order to be remembered. In the absence of any episodic memory trace at all, typical objects can easily be inferred. However, the presence of an incongruent object cannot because it is not predictable at all from the scene schema, and it will lead to a much more distinctive trace in memory. Thus in addition to serving as a better contextual retrieval cue, scene information can guide attention during encoding, so that details inconsistent from our stored knowledge can be encoded (Friedman, 1979; Gordon, 2004; Henderson, Weeks, & Hollingworth, 1999; Pezdek et al., 1989; Hollingworth & Henderson, 2000, 2003; see also Vogt & Magnussen, 2007).

Systematic biases in object memory due to scene information—Just as item-specific knowledge can lead to item-specific biases in the retrieval of details (e.g. the eyeglasses versus the dumbbell), so too can stored knowledge about scenes give rise to systematic errors about objects (Brewer & Treyens, 1981; Lampinen et al. 2001; Miller & Gazzaniga 1998; See also Aminoff, Schacter, & Bar, 2008). For example, after studying the objects in a series of scenes (e.g. including a golf course scene), participants are more likely to have false memory for a related object (a golf bag), than for an object that was not related to any of the studied scenes (e.g. a typewriter; Miller & Gazzaniga, 1998). These biases are typically errors of "commission": a refrigerator is remembered in the picture of a kitchen even if was not present because if there is any uncertainty in memory for that detail, there is still a high likelihood that there was a refrigerator visible in the kitchen. In this way, stored knowledge about scenes does not necessarily always support a more accurate or photographic-like memory of the objects in the scene. However, stored scene information can provide 'good guesses' for what was likely to be there, which is often optimal given

uncertainty (though this aspect of memory is problematic for eye-witness testimony, e.g., Loftus, 2004).

Conclusion—There are several ways that scene information plays a role in memory for object information. First, scene information can direct attention to which objects are encoded (e.g., Hollingworth & Henderson, 2000, 2003), just as object information can direct attention to which features within an object are encoded (e.g., Schyns and Rodet, 1997). In this way, scene knowledge, in addition to object knowledge, can be brought to bear during encoding. Efficient strategies suggest that distinctive details should be encoded along meaningful dimensions of variation at the scene-level (e.g. remember improbable objects), just as distinctive details are encoded at the object-level (e.g. remember improbable features). Second, scene information is stored along with item information in memory; it is not the case that scenes simply guide attention to objects and leave no trace in memory. As evidence for this claim, scene information facilitates retrieval of object details, even when the scene is task-irrelevant (e.g., Hollingworth, 2006b). At retrieval, the better the match of the second display to the initial presentation, the better the memory representation is retrieved. Finally, stored knowledge about the scene information can help provide meaningful guesses or graceful errors if we are probed on a detail for which we have a noisier representation. Taken together, these results suggest that what is stored in visual long-term memory includes item-specific and across-item (scene) information, and that all aspects of this representation can impact memory performance for a single object.

Visual long-term memory conclusion

We have reviewed research demonstrating that visual long-term memory can store thousands of items with surprisingly high fidelity (e.g., Brady et al. 2008; Hollingworth, 2004). This ability depends critically on the existence of stored knowledge about the items: the more observers know about the items, the more they can remember about them (e.g., Konkle et al. 2010b; Wiseman & Neisser, 1974). Many computational models suggest that the format of this stored knowledge is hierarchical, with lower levels consisting of basic features that are shared across categories, and higher levers consisting of visual features that are more category-specific. This structured knowledge constitutes the coding model used to extract information from incoming input, resulting in hierarchically structured episodic memory representations. There is some evidence that the structure of this knowledge at both the item-level and the scene-level influences memory for individual items, suggesting that the levels of representation within the hierarchy are mutually informative and constraining (e.g., Hollingworth, 2006b). However, much remains to be discovered regarding the nature and structure of this stored knowledge, and how it influences the content of episodic visual long-term memory representations.

If each visual episode leaves a trace in memory, what does it mean to estimate the capacity of visual long-term memory? The evidence reviewed here suggests that visual long-term memory has content-dependent capacity: the number of items that can be stored and the fidelity of storage depends on what is being remembered. If the content is 10,000 unique pictures of meaningful scenes, performance will be closer to perfect than to chance, even for tests that require detailed discriminations. However, if the content is 10,000 items all from the same semantic category, performance will much lower. If the content is 10,000 random dot displays, observers will be at chance performance. Thus, it seems that the capacity of visual long-term memory is not fixed in terms of either quantity or fidelity. We suggest that understanding the capacity and limitations of visual long-term memory requires characterizing what coding-model can be brought to bear on the content to be remembered, where the quantity and fidelity of retrieved episodes will depend on how many relevant

distinctive features can be extracted during encoding and retrieval of these perceptual episodes.

The work reviewed here on long-term memory dovetails with similar ideas in the section on working memory, where we raised the possibility that individual items in memory are represented as hierarchical feature-bundles, and that items are not represented independently but as part of a scene structure (e.g., making use of ensemble statistics). Broadly this suggests that similar coding mechanisms are brought to bear on any visual stimulus, enabling and creating a hierarchical features representation, whether they operate over simplified displays or natural visual scenes and whether they are actively maintained or passively stored and retrieved (Figure 9).

THE RELATIONSHIP BETWEEN WORKING MEMORY AND LONG-TERM MEMORY

In this review, we have followed the traditional distinction between visual working memory and visual long-term memory, treating them as separate systems with separate capacities. We have distinguished between working memory and long-term memory based primarily on the time-scale of memory storage: at short-time scales we assume that information is held actively in mind and therefore performance is determined primarily by working memory, whereas over longer time scales we assume that information is held passively in mind and therefore performance is determined primarily by long-term memory. However, while the separability of the active storage and passive retrieval systems is relatively uncontroversial, the extent to which the standard working memory and long-term memory paradigms actually isolate these systems remains an important open question. In particular, the extent to which passive retrieval might play an important role even in memory with short delays is a point of significant debate.

This is an important concern because the broader working memory literature --particularly the literature on verbal working memory -- has accumulated significant evidence for shared principles of short-term and long-term memory (McElree, 2006; Nairne, 2002; Jonides et al. 2008), including evidence that items putatively held in active storage are not accessed any faster (McElree, 2006), and that the medial temporal lobe, including the hippocampus, seems to be equally involved in retrieval of items from long-term storage and items that would be expected to be actively maintained (Öztekin, Davachi & McElree, 2010). This suggests we must even explore the possibility that active storage may play only part of the role in the short-term storage of information.

We begin by assessing the role of active and passive storage in working memory paradigms that use stimuli with meaningless, unrelated items, such as randomly selected colored circles, abstract shapes or meaningless characters (henceforth "semantically impoverished stimuli"). We then examine the role of passive storage in paradigms involving real-world semantically rich stimuli, for which observers have richer pre-existing stored knowledge.

The role of active and passive storage in the short-term storage of semantically impoverished stimuli—Evidence from brain imaging suggests that active storage drives a significant part of performance in standard visual working memory paradigms with semantically impoverished stimuli. This active storage seems to be achieved through a combination of continuing activity in frontal and parietal cortex, and either sustained activity or changed patterns of activity in lower-level visual cortex (e.g., Sakai, Rowe, & Passingham, 2002; Todd & Marois, 2004; Xu & Chun, 2006; Harrison & Tong, 2009). For example, studies using paradigms like those used in traditional behavioral studies

of visual working memory (e.g., Todd & Marois, 2004; Xu & Chun, 2006) have found that sustained activity in the intra-parietal sulcus and lateral occipital complex scales with how many items are being actively maintained or how much total information is being retained (Xu & Chun, 2006). Moreover, recent studies have indicated that memory for orientation is reflected in the ongoing activity in area V1 when the orientation is being actively held in mind (Harrison & Tong, 2009; Serences et al. 2009). In addition to fMRI correlates, studies using EEG have shown that there is sustained activity in the contralateral hemisphere when observers hold items in visual working memory, and this contralateral delay activity is greater in magnitude when observers hold more items (Vogel & Machizawa, 2004; Perez & Vogel, 2011). This signal saturates when the number of items reaches the maximum number that can be remembered, with a strong correlation between individual differences in the saturation level and individual differences in the number of items that can be remembered (Vogel & Machizawa, 2004). This suggests that the contralateral delay activity indexes the active storage of items in visual working memory.

Thus, it is clear that short-term storage relies at least in part on active maintenance. However, there is some evidence suggesting that passive retrieval may also play a role in the short-term storage of information. For example, behavioral studies have shown proactive interference from previous trials (Hartshorne, 2008; Makovski & Jiang, 2008) even in working memory paradigms, and systematic biases from previous trials suggest an influence of longer-term storage (Huang & Sekuler, 2010). Moreover, patient studies have demonstrated that the medial temporal lobe and hippocampus, believed to be critical for storage and retrieval in long-term memory (e.g., Davachi, 2006), may also be critical for the short-term storage of simple shapes (Olson et al. 2006). Specifically, patients with medial temporal damage are significantly impaired even in short-term storage of shape and color stimuli (Olson et al. 2006). In addition, parietal regions believed to be specifically involved in retrieval from long-term memory are active and necessary during working memory tasks even for simple, semantically impoverished stimuli (Berryhill & Olson, 2008a; Berryhill & Olson, 2008b). Taken together with the results from verbal working memory (e.g., Öztekin, Davachi & McElree, 2010), this suggests that passive storage may play a role even in standard visual working memory paradigms. However, to our knowledge there is no *direct* evidence to indicate that passive long-term memory plays a major role in the short-term storage of simple visual stimuli.

The role of passive storage in the short-term storage of real-world stimuli—
Observers are often able to hold only 3 or 4 simple colors in visual working memory. But how many semantically rich real-world objects can observers remember in working memory tasks such as change detection? In fact, it is usually found that given sufficient time to encode the objects, observers are able to detect changes to real-world objects easily, successfully remembering all the objects they are shown (Brady et al. 2009; see also Melcher, 2001, 2006). This contrasts markedly with studies of simple colored squares and complex abstract stimuli (e.g., 3D cubes; abstract complex shapes), which find that performance decreases significantly when observers are asked to remember more objects, as though either a resource or item limit had been reached (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Zhang & Luck, 2008). Increased performance with real-world stimuli could result from a number of factors. For example, the large amount of stored knowledge observers' have about such stimuli may provide more extracted features for each item, making items more distinctive from each other. Additionally, real-world stimuli may allow for easier test-foil combinations, because many different features of an object may be changed at once, as compared to only the low-level features of a simple square or even a 3D cube.

However, one intriguing hypothesis is that increased performance with real-world objects is a result of the increased use of the passive storage system for real-world objects compared to semantically impoverished objects, such as simple or complex geometric shapes. In particular, while active storage clearly plays a role in working memory performance for real-world stimuli (e.g., with faces both fMRI activity in the fusiform face area and the contralateral delay activity are increased with greater memory load: Druzgal & D'Esposito, 2001; Ruchkin, Johnson, Grafman, Canoune, & Ritter, 1992), there could be significant effects of passive storage, even at short delays. In other words, 'working memory' paradigms that use semantically rich real-world stimuli make use of not only the active working memory system, but also the passive episodic retrieval ("long-term") memory system, which operates most effectively when objects are semantically distinctive (e.g., Konkle et al. 2010b).

For example, Hollingworth (2004) has demonstrated that both working memory and long-term memory contribute to memory for objects within scenes. Memory is best for the few most recent items (a recency effect), but observers are able to remember significant information about many objects from a scene, even at long delays. This is compatible with performance on the task being driven by both working memory and long-term passive storage: at short time scales, working memory and passive storage both contribute; but with increasing delay, only long-term, passive representations remain, and so performance asymptotes (compatible with standard interpretations of the serial position curve in verbal memory: e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965).

Effects of encoding time and consolidation also suggest that the short-term storage of real-world objects might make use of the passive storage system. For example, in the original study of Luck and Vogel (1997), no difference was found in performance when the stimuli to be encoded were shown for 100ms or for 500 ms, which they took as evidence that the active storage buffer was 'filled up' even with 100ms of exposure (a fact compatible with the results from EEG studies: Perez & Vogel, 2011) (Figure 10b). By contrast, paradigms using real-world objects demonstrate no such 'filling up': instead, performance improves continuously with more time to encode the stimuli (e.g., Brady et al. 2009; Melcher, 2001, 2006) (Figure 10a). In addition, Vogel et al. (2006) find that it takes only 50–100ms to consolidate each item into memory when using simple, semantically impoverished stimuli. However, with real-world stimuli much more time is needed: for example, important work by Molly Potter on conceptual short-term memory and the attentional blink has shown that while real-world stimuli are well remembered later even if visually masked after about 100ms of processing, it requires approximately another 300ms of processing before they are immune to conceptual masking from another meaningful image (Potter, 1976) and up to 500ms to make them fully consolidated (e.g., Chun & Potter, 1995).

There are two implications to the different time courses observed for real-world stimuli. First, real-world stimuli show no evidence of an asymptote that is usually taken as evidence of the 'filling up' the active buffer. This might be because passive storage is playing a significant role in performance with real-world objects. Second, it takes significantly longer to consolidate real-world stimuli. This could be because encoding a trace that is available for passive retrieval takes longer to create.

Conclusion

There is significant reason to believe that passive storage plays a role even in the short-term storage of information. For example, working memory paradigms seem to be dependent on the medial temporal lobe (Olson et al. 2006) and, as the stimuli to be remembered are increasingly semantically rich, there does not seem to be a fixed storage buffer that "fills up"

after some encoding duration (e.g., Brady et al. 2009). Thus, the relative influence of active and passive storage in working memory paradigms remains an important open question.

One hypothesis is that passive storage is *always* used in addition to active maintenance in all paradigms requiring short-term storage. Under this hypothesis passive storage may simply contribute more to performance for real-world objects than for simple stimuli because it is more difficult to make use of the passive storage system for semantically impoverished stimuli. For example, real-world stimuli are much more likely to be conceptually distinctive, allowing them to be encoded in memory more accurately or retrieved with less interference than displays of semantically impoverished stimuli (e.g., Konkle et al. 2010b). Furthermore, real-world stimuli allow experimenters to use different stimuli on each trial, making the passive retrieval task easier by requiring only familiarity, and not recollection (e.g., Yonelinas, 2001).

Insight into these questions will likely require research on memory representations at the intersection of visual working memory and visual long-term memory. Experiments requiring memory for real-world items within naturalistic, structured displays likely engage both active and passive memory systems, and will therefore be a productive avenue for future research. However, the field must first establish behavioral and neural methods for isolating active vs. passive representations in order to correctly attribute aspects of performance to working memory and long-term memory contributions.

SUMMARY & CONCLUSIONS

Research that focuses on memory systems and memory processes often aims to discover principles of memory function that generalize across the type of information that is being remembered. However, understanding the content of memory representations can place important constraints on models of memory. In this review, we have discussed some of the key experiments on visual working memory and visual long-term memory that have focused on the content of memory representations. Specifically, we have focused on studies which have characterized and expanded our knowledge about the fidelity of working and long term memory representations, explored different characterizations of the basic units of memory and the relationship across items in memory, and highlighted the critical effects of stored knowledge on memory.

This representation-based approach has led to several discoveries regarding the content and structure of stored representations, and has led to several claims and constraints on the nature of working and long-term memory models. Taken together, these findings suggest that models of memory must go beyond characterizing how individual items are stored, and move towards capturing the more complex, structured nature of memory representations. We have proposed that information is represented at the individual item level as hierarchical feature-bundles, and across individual items in terms of ensemble or scene context, and that these levels of representation interact (see Figure 9). Moreover, this structure applies to both simple and real-world displays, and to both visual working memory and visual long-term memory.

Importantly, visual memory research can also inform vision research, providing a method to validate or test different models of visual representations: better models of how stimuli are coded should lead to better predictions of memory performance, in both working memory and long-term memory. For example, biases in memory towards prototypical values can be used to investigate which dimensions are used to represent objects (Huang & Sekuler, 2010; Huttenlocher et al. 2000), and the extent to which object features tend to be remembered independently or separately can indicate whether objects are coded as bound units (Bays, Wu & Husain, 2011; Fougnie & Alvarez, submitted). In this way, visual memory paradigms

can be used not only for understanding memory systems and processes, but also for understanding the nature of existing visual representations.

As research in the domains of visual working and visual long-term memory move forward, it will be essential to keep in mind that paradigms and stimuli do not isolate processes. For example, the process of passive episodic retrieval likely occurs in all memory paradigms, even in paradigms that require short-term storage, and even when using stimuli that long-term memory cannot effectively store. One promising approach to teasing apart the active vs. passive storage of information is to use online neural measures – literally measuring whether representations are active. Additionally, new behavioral methods could be developed to expressly measure and model both the working memory and long-term memory contributions to any task. If the past research on memory systems has helped us isolate and characterize these different processes, future research may now shift the focus towards understanding how these memory systems interact in the same moment in time, operating over similar structured representations.

Acknowledgments

References

Allen RJ, Baddeley AD, Hitch GJ. Is the binding of visual features in working memory resource-demanding? Journal of Experimental Psychology: General. 2006; 135:298–313. [PubMed: 16719655]

Alloway TP, Alloway RG. Investigating the predictive roles of working memory and IQ in academic attainment. Journal of Experimental Child Psychology. 2010; 106(1):20–29. [PubMed: 20018296]

Alvarez GA. Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences. 2011; 15(3):122–131. [PubMed: 21292539]

Alvarez GA, Cavanagh P. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychological Science. 2004; 15(2):106–111. [PubMed: 14738517]

Aminoff E, Schacter DL, Bar M. The cortical underpinnings of context-based memory distortion. Journal of Cognitive Neuroscience. 2008; 20(12):2226–2237. [PubMed: 18457503]

Anderson DE, Vogel EK, Awh E. Precision in Visual Working Memory Reaches a Stable Plateau When Individual Item Limits Are Exceeded. Journal of Neuroscience. 2011; 31(3):1128–1138. [PubMed: 21248137]

Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation: Advances in research and theory. Vol. 2. New York: Academic Press; 1968. p. 742-775.

Awh E, Jonides J. Overlapping mechanisms of attention and working memory. Trends in Cognitive Sciences. 2001; 5(3):119–126. [PubMed: 11239812]

Awh E, Barton B, Vogel EK. Visual working memory represents a fixed number of items, regardless of complexity. Psychological Science. 2007; 18(7):622–628. [PubMed: 17614871]

Baddeley, AD. Working memory. Oxford: Clarendon Press; 1986.

Baddeley AD. The episodic buffer: a new component of working memory? Trends in Cognitive Sciences. 2000; 4(11):417–423. [PubMed: 11058819]

Baddeley AD, Allen RJ, Hitch GJ. Binding in visual working memory: The role of the episodic buffer. Neuropsychologia. 2011

Baddeley AD, Scott D. Short-term forgetting in the absence of proactive inhibition. Q J Exp Psychol. 1971; 23:275–283.

Bar M. Visual objects in context. Nature Reviews Neuroscience. 2004; 5:617–629.

Bar M, Ullman S. Spatial context in recognition. Perception. 1996; 25:343–352. [PubMed: 8804097]

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmidt AM, Dale AM, Hamalainen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E. Top-down facilitation of visual recognition. Proceedings of the National Academy of Science. 2006; 103(2):449–454.

Bartlett, FC. Remembering. University Press; 1932.

Bays PM, Catalao RFG, Husain M. The precision of visual working memory is set by allocation of a shared resource. Journal of Vision. 2009; 9(10):1–11. [PubMed: 19810788]

Bays PM, Husain M. Dynamic Shifts of Limited Working Memory Resources in Human Vision. Science. 2008; 321(5890):851. [PubMed: 18687968]

Bays PM, Wu EY, Husain M. Storage and binding of object features in visual working memory. Neuropsychologia. 2011

Berryhill ME, Olson IR. Is the posterior parietal lobe involved in working memory retrieval? Evidence from patients with bilateral parietal lobe damage. Neuropsychologia. 2008a; 46:1775–1786. [PubMed: 18439630]

Berryhill ME, Olson IR. The right parietal lobe is critical for working memory. Neuropsychologia. 2008b; 46:1767–1774. [PubMed: 18308348]

Biederman I, Mezzanotte RJ, Rabinowitz JC. Scene perception: detecting and judging objects undergoing relational violations. Cognitive Psychology. 1982; 14:143–177. [PubMed: 7083801]

Blakemore C, Campbell FW. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. The Journal of Physiology. 1969; 203(1): 237–60. [PubMed: 5821879]

Bower GH, Karlin MB, Dueck A. Comprehension and memory for pictures. Memory and Cognition. 1975; 3:216–220.

Brady TF, Alvarez GA. Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. Psychological Science. 2011; 22(3):384–392. [PubMed: 21296808]

Brady TF, Konkle T, Alvarez GA. Compression in visual short-term memory: using statistical regularities to form more efficient memory representations. Journal of Experimental Psychology: General. 2009; 138(4):487–502. [PubMed: 19883132]

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

Brady TF, Konkle T, Oliva A, Alvarez GA. Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. Communicative & Integrative Biology. 2009; 2:1, 1–3. [PubMed: 19704852]

Brady, TF.; Tenenbaum, JB. Encoding higher-order structure in visual working memory: A probabilistic model. In: Ohlsson, S.; Catrambone, R., editors. Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science; 2010. p. 411-416.

Brady, TF.; Tenenbaum, JB. A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates. (submitted)

Brewer WF, Treyens JC. Role of schemata in memory for places. Cognitive Psychology. 1981; 13:207–230.

Broadbent, DE. Perception and Communication. Pergamon Press; New York: 1958.

Brown GDA, Neath I, Chater N. A temporal ratio model of memory. Psychological Review. 2007; 114(3):539–576. [PubMed: 17638496]

Carmichael L, Hogan HP, Walter AA. An experimental study of the effect of language on the reproduction of visually perceived forms. Journal of Experimental Psychology. 1932; 15:73–86.

Castel AD, McCabe DP, Roediger HL, Heitman JL. The Dark Side of Expertise. Psychological Science. 2007; 18(1):3–5. [PubMed: 17362368]

Chase WG, Simon HA. Perception in chess. Cognitive Psychology. 1973; 4:55–81.

Chen D, Eng HY, Jiang Y. Visual working memory for trained and novel polygons. Visual Cognition. 2006; 14(1):37–54.

Chun, MM. Scene perception and memory. In: Irwin, D.; Ross, B., editors. Psychology of Learning and Motivation: Advances in Research and Theory: Cognitive Vision. Vol. 42. Academic Press; San Diego, CA: 2003. p. 79-108.

Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance. 1995; 21:109–127. [PubMed: 7707027]

Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences. 2001; 24:87–185. [PubMed: 11515286]

Cowan, N. Working memory capacity. Hove, East Sussex, UK: Psychology Press; 2005.

Cowan N, Chen Z, Rouder JN. Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). Psychological Science. 2004; 15:634–640. [PubMed: 15327636]

Cowan N, Elliott EM, Saults JS, Morey CC, Mattox S, Hismjatullina A, Conway ARA. On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology. 2005; 51:42–100. [PubMed: 16039935]

Cowan N, AuBuchon AM. Short-term memory loss over time without retroactive stimulus interference. Psychonomic Bulletin & Review. 2008; 15(1):230–235. [PubMed: 18605508]

Curby KM, Gauthier I. A visual short-term memory advantage for faces. Psychonomic Bulletin and Review. 2007; 14(4):620–628. [PubMed: 17972723]

Curby KM, Glazek K, Gauthier I. A visual short-term memory advantage for objects of expertise. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35 (1):94–107. [PubMed: 19170473]

Daneman M, Carpenter PA. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior. 1980; 19:450–466.

Davachi L. Item, context and relational episodic encoding in humans. Current Opinion in Neurobiology. 2006; 16:693–700. [PubMed: 17097284]

Davis G, Holmes A. The capacity of visual short-term memory is not a fixed number of objects. Memory & Cognition. 2005; 33(2):185–95.

de Groot, AD. Perception and memory versus thought: Some old ideas and recent findings. In: Kleinmuntz, B., editor. Problem solving. New York: Wiley; 1966. p. 19-50.

Delvenne JF, Bruyer R. Does visual short-term memory store bound features? Visual Cognition. 2004; 11(1):1–27.

Delvenne JF, Bruyer R. A configural effect in visual short-term memory for features from different parts of an object. The Quarterly Journal of Experimental Psychology. 2006; 59(9):1567–1580. [PubMed: 16873109]

Droll JA, Hayhoe MH, Triesch J, Sullivan BT. Task demands control acquisition and storage of visual information. Journal of Experimental Psychology: Human Perception and Performance. 2005; 31(6):1416–1438. [PubMed: 16366799]

Druzgal TJ, D'Esposito M. Activity in fusiform face area modulated as a function of working memory load. Cognitive Brain Research. 2001; 10(3):355–364. [PubMed: 11167061]

Dudai Y. How big is human memory, or on being just useful enough. Learning & Memory. 1997; 3(5): 341. [PubMed: 10456103]

Epshtein B, Ullman S. Feature hierarchies for object classification. IEEE Computer Society. 2005:1550–5499.

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998; 392:598–601. [PubMed: 9560155]

Ericsson KA, Chase WG. Exceptional memory. American Scientist. 1982; 70:607–615. [PubMed: 7181217]

Ericsson KA, Chase WG, Faloon S. Acquisition of a memory skill. Science. 1980; 208:1181–1182. [PubMed: 7375930]

Eysenck, MW. Depth, elaboration, and distinctiveness. In: Cermak, LS.; Craik, FIM., editors. Levels of processing in human memory. Hillsdale, NJ: Erlbaum; 1979. p. 89-118.

Feigenson L. Parallel non-verbal enumeration is constrained by a set-based limit. Cognition. 2008; 107(1):1–18. [PubMed: 17761158]

Feigenson L, Halberda J. Conceptual knowledge increases infants' memory. Proceedings of the National Academy of Sciences. 2008; 105(29):9926–9930.

Fougnie, D.; Alvarez, GA. Breakdown of object-based representations in visual working memory. (submitted)

Fougnie D, Asplund CL, Marois R. What are the units of storage in visual working memory? Journal of Vision. 2010; 10(12):1–11.

Fougnie D, Marois R. Attentive tracking disrupts feature binding in visual working memory. Visual Cognition. 2009; 17:48–66. [PubMed: 19609460]

Friedman A. Framing pictures: the role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General. 1979; 108:316–355. [PubMed: 528908]

Fukuda K, Vogel EK, Mayr U, Awh E. Quantity not quality: The relationship between fluid intelligence and working memory capacity. Psychonomic Bulletin and Review. 2010; 17(5):673–679. [PubMed: 21037165]

Gajewski DA, Brockmole JR. Feature bindings endure without attention: evidence from an explicit recall task. Psychonomic Bulletin & Review. 2006; 13(4):751, 581–587.

Godden DR, Baddeley AD. Context-dependent memory in two natural environments: On land and underwater. British Journal of Psychology. 1975; 66(3):325–331.

Gordon RD. Attentional allocation during the perception of scenes. Journal of Experimental Psychology: Human Perception and Performance. 2004; 30:760–777. [PubMed: 15301623]

Greene MR, Oliva A. Recognition of Natural Scenes from Global Properties: Seeing the Forest Without Representing the Trees. Cognitive Psychology. 2009; 58(2):137–179. [PubMed: 18762289]

Greene MR, Oliva A. High-Level Aftereffects to Global Scene Property. Journal of Experimental Psychology: Human Perception & Performance. 2010; 36(6):1430–1442. [PubMed: 20731502]

Haberman J, Whitney D. Seeing the mean: Ensemble coding for sets of faces. Journal of Experimental Psychology: Human Perception & Performance. 2009; 35(3):718–34. [PubMed: 19485687]

Haberman, J.; Whitney, D. Ensemble perception: Summarizing the scene and broadening the limits of visual processing. Chapter to appear in an edited volume. In: Wolfe, J.; Robertson, L., editors. A Festschrift in honor of Anne Treisman. 2011.

Halberda J, Sires SF, Feigenson L. Multiple spatially-overlapping sets can be enumerated in parallel. Psychological Science. 2006; 17 (7):572–576. [PubMed: 16866741]

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458(7238):632–635. [PubMed: 19225460]

Hartshorne JK. Visual Working Memory Capacity and Proactive Interference. PLoS One. 2008; 3(7)

Hemmer P, Steyvers M. Integrating Episodic Memories and Prior Knowledge at Multiple Levels of Abstraction. Psychonomic Bulletin & Review. 2009; 16:80–87. [PubMed: 19145014]

Henderson JM, Hollingworth A. High-level scene perception. Annual Review of Psychology. 1999; 50:243–271.

Henderson JM, Weeks PA Jr, Hollingworth A. The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:210–228.

Hock HS, Romanski L, Galie A, Williams CS. Real-world schemata and scene recognition in adults and children. Memory & Cognition. 1978; 6:423–431.

Hollingworth A. Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. Journal of Experimental Psychology: Human Perception and Performance. 2004; 30:519–537. [PubMed: 15161384]

Hollingworth A. The relationship between online visual representation of a scene and long-term scene memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:396–411.

Hollingworth A. Visual memory for natural scenes: Evidence from change detection and visual search. Visual Cognition. 2006a; 14:781–807.

Hollingworth A. Scene and position specificity in visual memory for objects. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006b; 32:58–69.

Hollingworth A. Object-position binding in visual memory for natural scenes and object arrays. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:31–47. [PubMed: 17311477]

Hollingworth A, Henderson JM. Semantic informativeness mediates the detection of changes in natural scenes. Visual Cognition. 2000; 7:213–235.

Hollingworth A, Henderson JM. Accurate visual memory for previously attended object in natural scenes. Journal of Experimental Psychology: Human Perception and Performance. 2002; 28:113–136.

Hollingworth A, Henderson JM. Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. Memory & Cognition. 2003; 31:930–940.

Hollingworth A, Rasmussen IP. Binding objects to locations: the relationship between object files and visual working memory. Journal of Experimental Psychology: Human Perception and Performance. 2010; 36(3):543–64. [PubMed: 20515188]

Huang J, Sekuler R. Distortions in recall from visual memory: Two classes of attractors at work. Journal of Vision. 2010; 10(2):24, 1–27.

Huang L. Visual working memory is better characterized as a distributed resource rather than discrete slots. Journal of Vision. 2010; 10(14):8, 1–8. [PubMed: 21135255]

Huttenlocher J, Hedges LV, Vevea JL. Why do categories affect stimulus judgment? Journal of Experimental Psychology: General. 2000; 129(2):220–241. [PubMed: 10868335]

James, W. The Principles of Psychology. New York: Holt; 1890.

Jiang Y, Olson IR, Chun MM. Organization of Visual-Short Term Memory. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2000; 26:683–702.

Johnson JS, Hollingworth A, Luck SJ. The role of attention in the maintenance of feature bindings in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34:41–55. [PubMed: 18248139]

Johnson JS, Spencer JP, Luck SJ, Schöner G. A dynamic neural field model of visual working memory and change detection. Psychological Science. 2009; 20:568–577. [PubMed: 19368698]

Joliceour P. The time to name disoriented natural objects. Memory & Cognition. 1985; 13(4):289–303.

Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. The mind and brain of short-term memory. Annual Review of Psychology. 2008; 59:193–224.

Kane MJ, Bleckley MK, Conway ARA, Engle RW. A controlled-attention view of working-memory capacity. Journal of Experimental Psychology: General. 2001; 130:169–183. [PubMed: 11409097]

Kemp C, Tenenbaum JB. Structured statistical models of inductive reasoning. Psychological Review. 2009; 116(1):20–58. [PubMed: 19159147]

Konkle T, Brady TF, Alvarez GA, Oliva A. Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology: General. 2010b; 139(3):558–78. [PubMed: 20677899]

Konkle T, Brady TF, Alvarez GA, Oliva A. Scene memory is more detailed than you think: the role of categories in visual long-term memory. Psychological Science. 2010a; 21(11):1551–1556. [PubMed: 20921574]

Konkle, T.; Oliva, A. Normative representation of objects: Evidence for an ecological bias in perception and memory. In: McNamara, DS.; Trafton, JG., editors. Proceedings of the 29th Annual Cognitive Science Society. Austin, TX: Cognitive Science Society; 2007. p. 407-413.

Koutstaal W, Reddy C, Jackson EM, Prince S, Cendan DL, Schacter DL. False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2003; 29:499–510.

Lampinen JM, Copeland S, Neuschatz JS. Recollections of things schematic: Room schemas revisited. Journal of Experimental Psychology: Learning, Memory and Cognition. 2001; 27:1211–1222.

Landauer TK. How much do people remember? some estimates of the quantity of learned information in long-term memory. Cognitive Science. 1986; 10(4):477–493.

Lin P-H, Luck SJ. The influence of similarity on visual working memory representations. Visual Cognition. 2008; 17:356–372. [PubMed: 19430536]

Loftus E. Memories of things unseen. Current Directions in Psychological Science. 2004; 13(4):145–147.

Luck, SJ.; Hollingworth, A. Visual Memory. New York: Oxford University Press; 2008.

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997; 390:279–281. [PubMed: 9384378]

Lupyan G. From Chair to 'Chair': A Representational Shift Account of Object Labeling Effects on Memory. Journal of Experimental Psychology: General. 2008; 137(2):348–369. [PubMed: 18473663]

Magnussen S, Greenlee MW, Thomas JP. Parallel processing in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance. 1996; 22(1):202–212. [PubMed: 8742262]

Makovski T, Jiang YV. Proactive interference from items previously stored in visual working memory. Memory and Cognition. 2008; 36(1):43–52.

Mandler JM, Johnson NS. Some of the thousand words a picture is worth. Journal of Experimental Psychology: Human Learning and Memory. 1976; 2:529–540. [PubMed: 1003126]

Mandler JM, Parker RE. Memory for descriptive and spatial information in complex pictures. Journal of Experimental Psychology: Human Learning and Memory. 1976; 2:38–48. [PubMed: 1249532]

Mandler JM, Ritchey GH. Long-term memory for pictures. Journal of Experimental Psychology: Human Learning and Memory. 1977:386–396.

McElree, B. Accessing recent events. In: Ross, BH., editor. The psychology of learning and motivation. Vol. 46. San Diego: Academic Press; 2006.

Melcher D. Persistence of visual memory for scenes. Nature. 2001; 412:401. [PubMed: 11473303]

Melcher D. Accumulation and persistence of memory for natural scenes. Journal of Vision. 2006; 6(1): 8–17. [PubMed: 16489855]

Mervis CB, Rosch E. Categorization of natural objects. Annual Review of Psychology. 1981; 32(1): 89–115.

Miller GA. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review. 1956; 63:81–97. [PubMed: 13310704]

Miller MB, Gazzaniga MS. Creating false memories for visual scenes. Neuropsychologia. 1998; 36:513–520. [PubMed: 9705061]

Minsky, M. A framework for representing knowledge. In: Winston, P., editor. The psychology of computer vision. McGraw-Hill; 1975. p. 211-277.

Mitroff SR, Simons DJ, Levin DT. Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. Perception & Psychophysics. 2004; 66:1268–1281. [PubMed: 15813193]

Nairne JS. Remembering over the short-term: The case against the standard model. Annual Review of Psychology. 2002; 52:53–81.

Nairne, JS. Modeling distinctiveness: Implications for general memory theory. In: Hunt, RR.; Worthen, J., editors. Distinctiveness and memory. New York: Oxford University Press; 2006.

Oliva, A. Gist of the scene. In: Itti, L.; Rees, G.; Tsotsos, JK., editors. In the Encyclopedia of Neurobiology of Attention. Elsevier; San Diego, CA: 2005. p. 251-256.

Oliva A, Torralba A. Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. International Journal in Computer Vision. 2001; 42:145–175.

Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381:607–609. [PubMed: 8637596]

Olson IR, Marshuetz C. Remembering "What" Brings Along "Where" in visual working memory. Perception & Psychophysics. 2005; 67(2):185–194. [PubMed: 15971683]

Olson IR, Jiang Y. Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. Perception & Psychophysics. 2002; 64:1055–1067. [PubMed: 12489661]

Olson IR, Jiang Y. Visual short-term memory is not improved by training. Memory and Cognition. 2004; 32:1326–1332.

Olson IR, Jiang Y, Moore KS. Associative learning improves visual working memory performance. Journal of Experimental Psychology: Human Perception and Performance. 2005; 31:889–900. [PubMed: 16262486]

Olson IR, Moore KS, Stark M, Chatterjee A. Visual working memory is impaired when the medial temporal lobe is damaged. Journal of Cognitive Neuroscience. 2006; 18(7):1–11. [PubMed: 16417678]

Olsson H, Poom L. Visual memory needs categories. Proceedings of the National Academy of Sciences, USA. 2005; 102:8776–8780.

Ommer B, Buhmann JM. Learning the compositional nature of visual object categories for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010; 32(3):501–516. [PubMed: 20075474]

Orbán G, Fiser J, Aslin RN, Lengyel M. Bayesian learning of visual chunks by human observers. Proceedings of the National Academy of Sciences USA. 2008; 105(7):2745–50.

Öztekin I, Davachi L, McElree B. Are representations in working memory distinct from those in long-term memory? Neural evidence in support of a single store. Psychological Science. 2010; 21:1123–1133. [PubMed: 20624934]

Pashler H. Familiarity and the detection of change in visual displays. Perception & Psychophysics. 1988; 44:369–378. [PubMed: 3226885]

Perez, VB.; Vogel, EK. What ERPs can tell us about Visual Working Memory. In: Luck, SJ.; Kappenman, E., editors. Oxford Handbook of Event-related Potential Components. 2011.

Pezdek K, Whetstone T, Reynolds K, Askari N, Dougherty T. Memory for real-world scenes: The role of consistency with schema expectation. Journal of Experimental Psychology: Learning, Memory & Cognition. 1989; 15:587–595.

Phillips WA. On the distinction between sensory storage and short-term visual memory. Perception & Psychophysics. 1974; 16(2):283–290.

Potter MC. Short-term conceptual memory for pictures. Journal of Experimental Psychology: Human Learning and Memory. 1976; 2:509–522. [PubMed: 1003124]

Quinlan PT, Cohen DJ. Object-based representations govern both the storage of information in visual short-term memory and the retrieval of information from it. Psychonomic Bulletin & Review. 2011

Rawson KA, Van Overschelde JP. How does knowledge promote memory? The distinctiveness theory of skilled memory. Journal of Memory and Language. 2008; 58:646–668.

Rensink RA. The dynamic representation of scenes. Visual Cognition. 2000; 7(1–3):17–42.

Rensink RA, O'Regan JK, Clark JJ. To see or not to see: The need for attention to perceive changes in scenes. Psychological Science. 1997; 8:368–373.

Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience. 1999; 2:1019–1025.

Rodet, L.; Schyns, PG. Learning features of representation in conceptual context. Proceedings of the XVI Meeting of the Cognitive Science Society; Erlbaum. 1994. p. 766-71.

Roediger HL, McDermott KB. Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory and Cognition. 1995; 21(4):803–814.

Rouder JN, Morey RD, Cowan N, Zwilling CE, Morey CC, Pratte MS. An assessment of fixed-capacity models of visual working memory. Proceedings of the National Academy of Sciences. 2008; 105(16):5975–5979.

Ruchkin DS, Johnson R, Grafman J, Canoune H, Ritter W. Distinctions and similarities among working memory processes: An event-related potential study. Cognitive Brain Research. 1992; 1:53–66. [PubMed: 15497435]

Sakai K, Rowe JB, Passingham RE. Active maintenance in prefrontal area 46 creates distractor-resistant memory. Nature Neuroscience. 2002; 5(5):479–8.

Schacter, DL.; Tulving, E. What are the memory systems of 1994?. In: Schacter, DL.; Tulving, E., editors. Memory Systems. Cambridge, MA: MIT Press; 1994.

Schank, RC. Conceptual information processing. New York: Elsevier; 1975.

Schmidt SR. Encoding and retrieval processes in the memory for conceptually distinctive events. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1985; 11:565–578.

Schyns PG, Rodet L. Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1997; 23:681–696.

Schyns PG, Goldstone RL, Thibaut JP. The development of features in object concepts. Behavioral and Brain Sciences. 1998; 21:1–17. [PubMed: 10097010]

Scolari M, Vogel EK, Awh E. Perceptual expertise enhances the resolution but not the number of representations in working memory. Psychonomic Bulletin & Review. 2008; 15(1):215–222. [PubMed: 18605506]

Scoville WB, Milner B. Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology and Neurosurgery and Psychiatry. 1957; 20:11–21.

Serences J, Ester E, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. Psychological Science. 2009; 20(2):207–214. [PubMed: 19170936]

Shepard RN. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior. 1967; 6:156–163.

Shiffrin RM, Steyvers M. A model for recognition memory: REM—retrieving effectively from memory. Psychonomic Bulletin & Review. 1997; 4(2):145–166. [PubMed: 21331823]

Simons DJ, Levin DT. Change blindness. Trends in Cognitive Sciences. 1997; 1(7):261–267. [PubMed: 21223921]

Simons DJ, Rensink RA. Change blindness: Past, present, and future. Trends in Cognitive Sciences. 2005; 9(1):16–20. [PubMed: 15639436]

Spencer JP, Hund AM. Prototypes and particulars: Geometric and experience-dependent spatial categories. Journal of Experimental Psychology: General. 2002; 131(1):16–36. [PubMed: 11900101]

Squire LR. Memory systems of the brain: A brief history and current perspective. Neurobiology of learning and memory. 2004; 82(3):171–177. [PubMed: 15464402]

Standing L. Learning 10,000 pictures. Quarterly Journal of Experimental Psychology. 1973; 25:207–222. [PubMed: 4515818]

Standing L, Conezio J, Haber RN. Perception and memory for pictures: Single trial learning of 2560 visual stimuli. Psychonomic Science. 1970; 19:169–179.

Stevanovski B, Jolicur P. Consolidation of multifeature items in visual working memory: Central capacity requirements for visual consolidation. Attention, Perception, & Psychophysics. 2011

Tenenbaum JB, Griffiths TL, Kemp C. Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences. 2006; 10(7):309–318. [PubMed: 16797219]

Todd JJ, Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature. 2004; 428:751–754. [PubMed: 15085133]

Torralba A, Murphy KP, Freeman WT. Sharing features: efficient boosting procedures for multiclass object detection. IEEE Computer Society. 2004:1063–6919.

Triesch JJ, Ballard D, Hayhoe M, Sullivan B. What you see is what you need. Journal of Vision. 2003; 3:86–94. [PubMed: 12678628]

Tulving, E. Concepts of memory. In: Tulving, E.; Craik, FIM., editors. The Oxford Handbook of Memory. New York, NY: Oxford University Press, Inc; 2000. p. 33-43.

Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. Psychological Review. 1973; 80(5):362–373.

Tversky B, Hemenway K. Categories of environmental scenes. Cognitive Psychology. 1983; 15:121–149.

Ullman S. Object recognition and segmentation by a fragment-based hierarchy. Trends in Cognitive Sciences. 2007; 11(2):58–64. [PubMed: 17188555]

Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. Nature Neuroscience. 2002; 5(7):1–6.

Victor JD, Conte MM. Visual working memory for image statistics. Vision Research. 2004; 44(6):541–556. [PubMed: 14693182]

Vidal JR, Gauchou HL, Tallon-Baudry C, O'Regan JK. Relational information in visual short-term memory: The structural gist. Journal of Vision. 2005; 5(3):244–256. [PubMed: 15929649]

Viswanathan S, Perl DR, Visscher KM, Kahana MJ, Sekuler R. Homogeneity computation: How inter-item similarity in visual short term memory alters recognition. Psychonomic Bulletin & Review. 2010; 17:59–65. [PubMed: 20081162]

Vogel EK, Machizawa MG. Neural activity predicts individual differences in visual working memory capacity. Nature. 2004; 428:784–751.

Vogel EK, Woodman GF, Luck SJ. The Time course of consolidation in visual working memory. Journal of Experimental Psychology: Human Perception and Performance. 2006; 32(6):1436–1451. [PubMed: 17154783]

Vogel EK, Woodman GF, Luck SJ. Storage of features, conjunctions, and objects in visual working memory. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27:92–114. [PubMed: 11248943]

Vogt S, Magnussen S. Long-term memory for 400 pictures on a common theme. Experimental Psychology. 2007; 54(4):298–303. [PubMed: 17953150]

von Restorff H. Uber die wirkung von bereichsbildungen im spurenfeld [The effects of field formation in the trace field]. Psychologische Forschung. 1933; 18:299–342.

Voss JF, Vesonder GT, Spilich GJ. Text generation and recall by high-knowledge and low-knowledge individuals. Journal of Verbal Learning and Verbal Behavior. 1980; 19:651–667.

Voss JL. Long-term associative memory capacity in man. Psychonomic Bulletin & Review. 2009; 16(6):1076–81. [PubMed: 19966258]

Waugh NC, Norman DA. Primary memory. Psychological Review. 1965; 72:89–104. [PubMed: 14282677]

Wheeler ME, Treisman AM. Binding in short-term visual memory. Journal of Experimental Psychology: General. 2002; 131:48–64. [PubMed: 11900102]

Wilken P, Ma WJ. A detection theory account of change detection. Journal of Vision. 2004; 4:1120–1135. [PubMed: 15669916]

Wiseman S, Neisser U. Perceptual Organization as a Determinant of Visual Recognition Memory. The American Journal of Psychology. 1974; 87 (4):675–681. [PubMed: 4463729]

Wolfe JM. Visual Memory: What do you know about what you saw? Current Biology. 1998; 8:R303–R304. [PubMed: 9560330]

Wood JN. Distinct visual working memory system for view-dependent and view-invariant representation. PLoS One. 2009; 11, 4(8):e6601. [PubMed: 19668380]

Wood JN. A core knowledge architecture of visual working memory. Journal of Experimental Psychology: Human Perception and Performance. 2011a; 37(2):357–81. [PubMed: 21463083]

Wood JN. When do spatial and visual working memory interact? Attention, Perception and Psychophysics. 2011b; 73(2):420–39.

Woodman GF, Vogel EK. Top-down control of visual working memory consolidation. Psychonomic Bulletin & Review. 2008; 15:223–229. [PubMed: 18605507]

Woodman GF, Vecera SP, Luck SJ. Perceptual organization influences visual working memory. Psychonomic Bulletin & Review. 2003; 10:80–87. [PubMed: 12747493]

Xiao, J.; Hayes, J.; Ehinger, K.; Oliva, A.; Torralba, A. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition; IEEE Computer Society. 2010. p. 3485-3492.

Xu Y, Chun MM. Visual grouping in human parietal cortex. Proceedings of the National Academy of Sciences, USA. 2007; 104:18766–18771.

Xu Y. Encoding color and shape from different parts of an object in visual short-term memory. Perception & Psychophysics. 2002; 64:1260–1280. [PubMed: 12519024]

Xu Y. Limitations in object-based feature encoding in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance. 2002b; 28:458–468. [PubMed: 11999866]

Xu Y, Chun MM. Dissociable neural mechanisms supporting visual short-term memory for objects. Nature. 2006; 440:91–95. [PubMed: 16382240]

Yonelinas AP. Components of episodic memory: The contribution of recollection and familiarity. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2001; 356(1413):1363–1374.

Zhang W, Luck SJ. Discrete fixed-resolution representations in visual working memory. Nature. 2008; 452:233–235. [PubMed: 18385672]

Zosh, JM.; Feigenson, L. Beyond 'what' and 'how many': Capacity, complexity, and resolution of infants' object representations. In: Hood, B.; Santos, L., editors. The Origins of Object Knowledge. New York: Oxford University Press; 2009. p. 25-51.
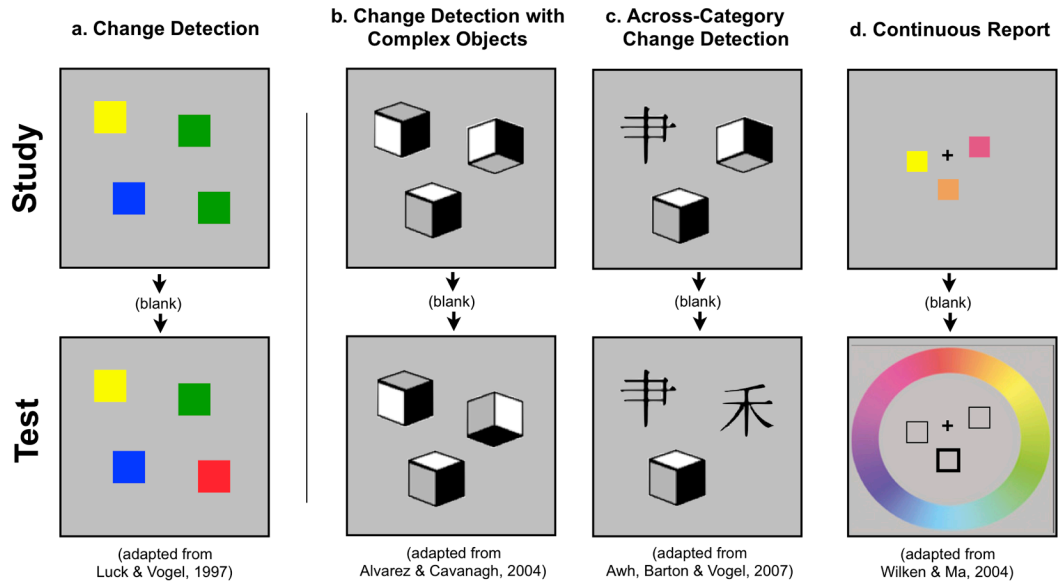
Figure 1.
Measures of visual working memory fidelity. (a) A change detection task. Observers see the 'Study' display, then after a blank must indicate whether the 'Test' display is identical to the Study display or whether a single item has changed color. (b) Change detection with complex objects. In this display, the cube changes to another cube (within-category change), requiring high-resolution representations to detect. (c) Change detection with complex objects. In this display, the cube changes to a Chinese character (across-category change), requiring only low-resolution representations to detect. (d) A continuous color report task. Observers see the Study display, and then at test are asked to report the exact color of a single item. This gives a continuous measure of the fidelity of memory.

a. Display   **b. Potential Memory Representations**
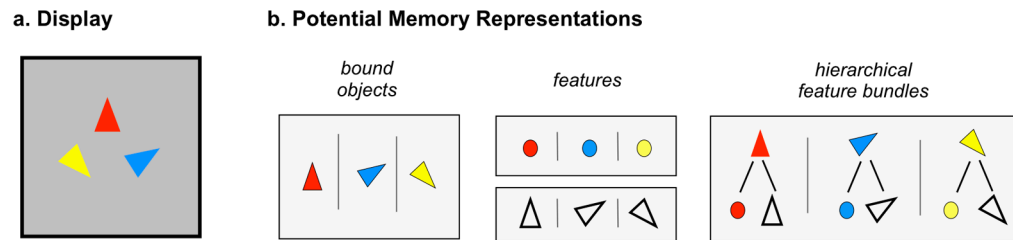
Figure 2.
Possible memory representations for a visual working memory display. (a) A display of oriented and colored items to remember. (b) Potential memory representations for the display in (a). The units of memory do not appear to be integrated bound objects, or completely independent feature representations. Instead, they might be characterized as hierarchical feature-bundles, which have both object-level and feature-level properties.

a. Spatial Context

(blank)

Easier   Harder

(Jiang, Olson & Chun, 2000)

b. Feature Context

(blank)

Easier   Harder
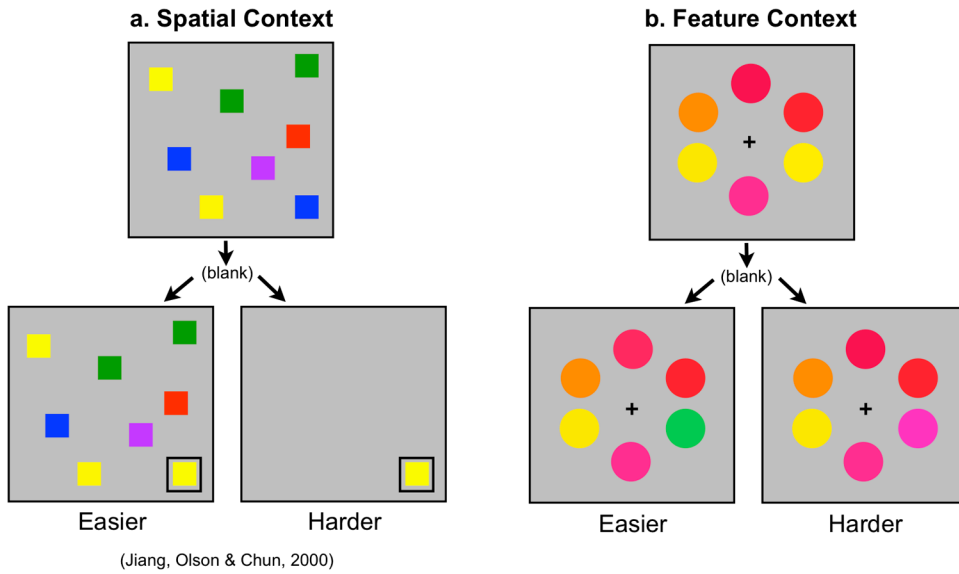
Figure 3.
Interactions between items in working memory. (a) Effects of spatial context. It is easier to detect a change to an item when the spatial context is the same in the original display and the test display than when the spatial context is altered, even if the item that may have changed is cued (with a black box). Displays adapted from the stimuli of Jiang, Olson & Chun (2000). (b) Effects of feature context on working memory. It is easer to detect a change to an item when the new color is outside the range of colors present in the original display, even for a change of equal magnitude.

a. Example display
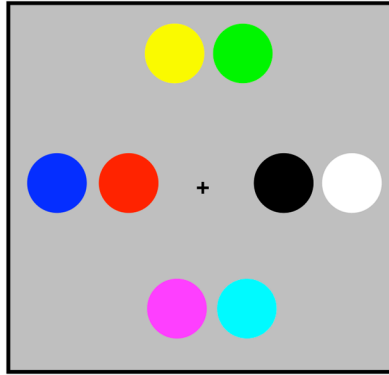
b. Number of colors remembered over time

(Brady, Konkle & Alvarez, 2009)
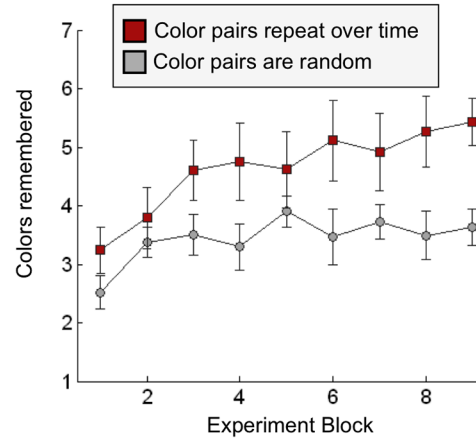
Figure 4.
Effects of learned knowledge on visual working memory. (a) Example memory display modeled after Brady, Konkle, & Alvarez (2009). The task was to remember all 8 colors. Memory was probed with a cued-recall test: a single location was cued, and the observer indicated which color appeared at the cued location. (b) Number of colors remembered over time in Brady, Konkle & Alvarez (2009). One group of observers saw certain color pairs more often than others (e.g., yellow and green might occur next to each other 80% of the time), whereas the other group saw completely random color pairs. For the group that saw repeated color pairs, the number of color remembered increased across blocks, nearly doubling the number remembered by the random group by the end of the session.

a. Novel Scene Categories

b. Scene Exemplars

(Standing, 1973)

(Konkle et al 2010a)

c. Novel Object Categories

d. Object Exemplars
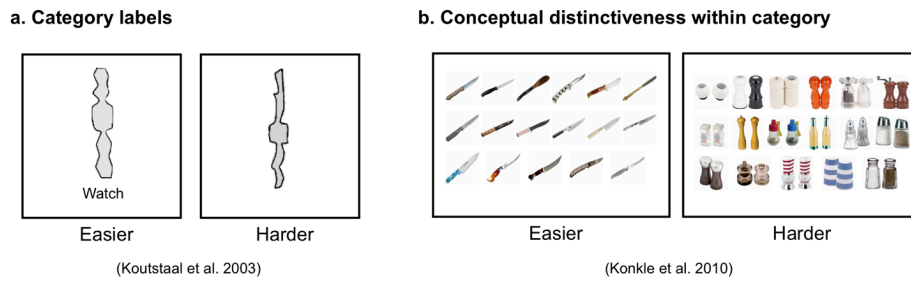
e. Object States

(Brady et. al, 2008)

Figure 5.
Explorations of fidelity in visual long-term memory. (a) Examples of scenes from different, novel categories (modeled after Standing, 1973). (b) Exemplars of scenes from the same category, greenhouse garden (as in Konkle et al., 2010a). (c) Objects from different, novel categories, as in Brady et al. (2008). (d) Examples of objects exemplars from the same category (globes and soap). (e) Examples of objects with a different state (full vs. empty mug) or different pose (mailbox with flag up vs. down).

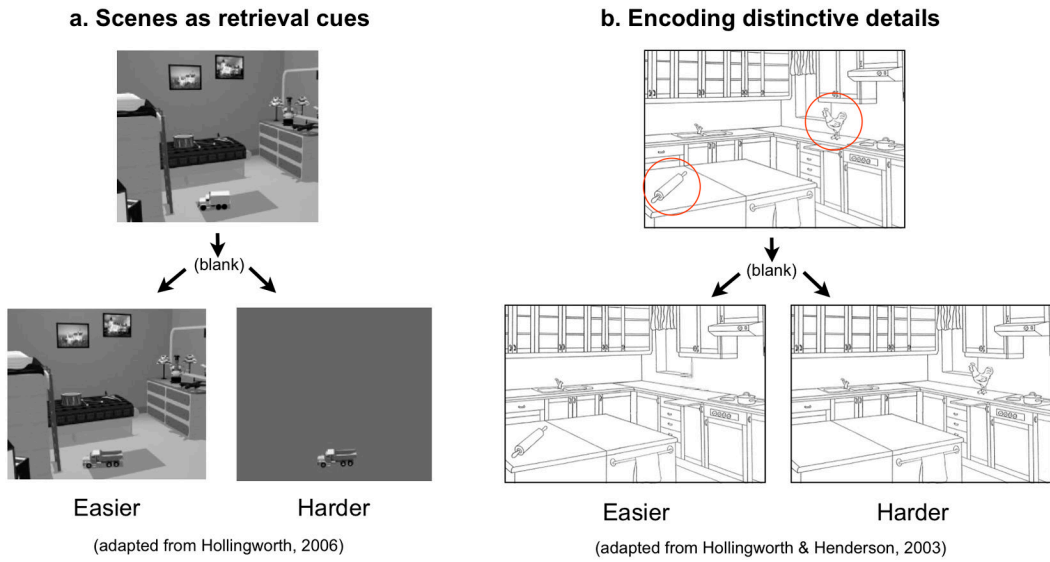# Hierarchy of object-generic and object-specific parts
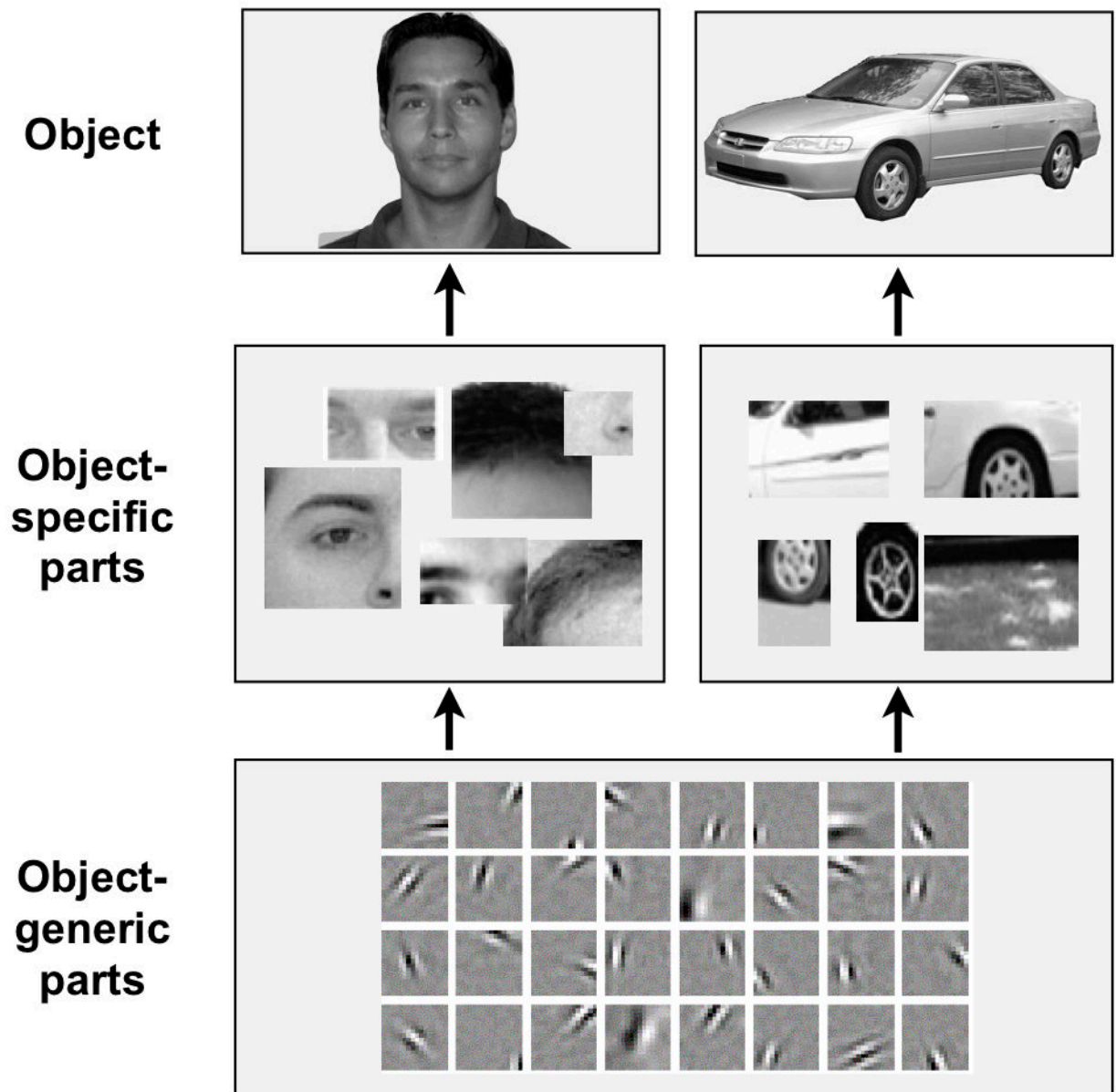
Figure 6.
Hierarchy of visual knowledge, from object-generic parts, to object-specific parts, to whole objects. Gabor patch stimuli adapted from Olshausen & Field (1996). Meaningful object fragments adapted from Ullman (2007).

a. Category labels

b. Conceptual distinctiveness within category

Easier	Harder
(Koutstaal et al. 2003)	(Konkle et al. 2010)

Figure 7.
Explorations of the role of conceptual information in visual memory. (a) Category labels that connect shapes with stored knowledge make it easier to remember shapes. Stimulus example adapted from Koutstaal et al. 2003. (b) Memory for multiple exemplars of the same category is better when the items are conceptually distinctive than when they are conceptually similar, independent of perceptual distinctiveness within the category. Stimuli from Konkle et al. 2010b.

a. Scenes as retrieval cues   **b. Encoding distinctive details**

(adapted from Hollingworth, 2006)   (adapted from Hollingworth & Henderson, 2003)
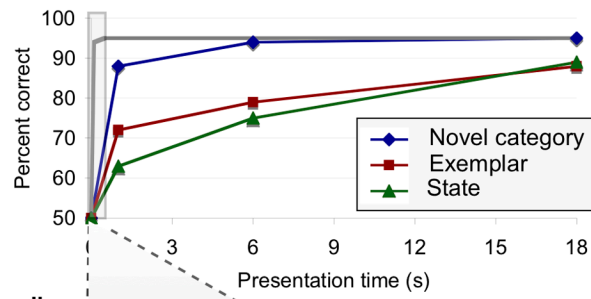
Figure 8.
Scenes influence the encoding and retrieval of objects. (a) Scenes as retrieval cues. It is easier to remember objects that are presented in the same scene at encoding and test, suggesting that scene context serves as a useful retrieval cue. Task and stimuli are adapted from Hollingworth 2006b. (b) Encoding distinctive details. Items that are inconsistent with the scene context are more likely to be remembered than items that are consistent with the scene context. This finding suggests that scene context guides encoding towards distinctive details within the scene. Task and stimuli are adapted from Hollingworth & Henderson, 2003.
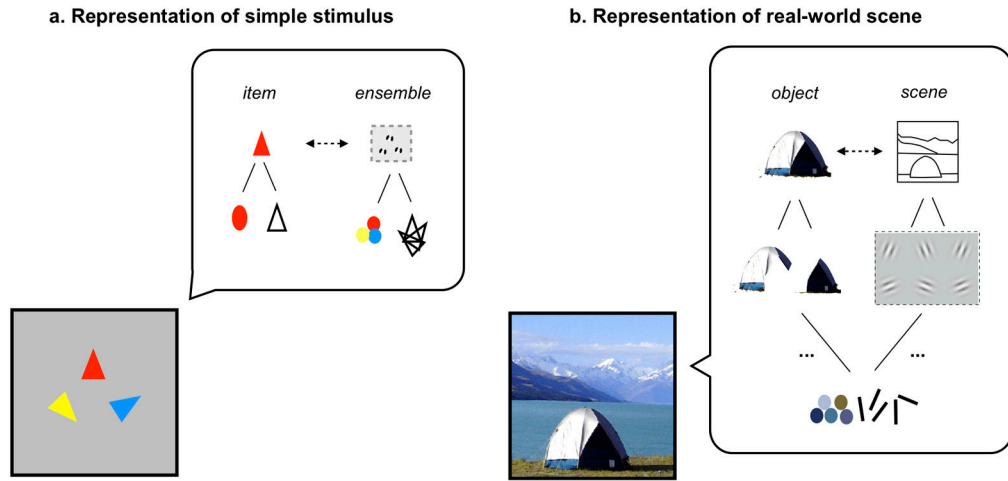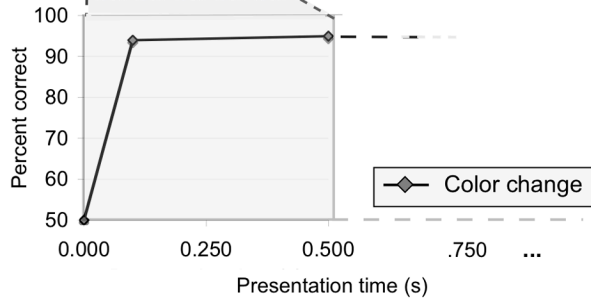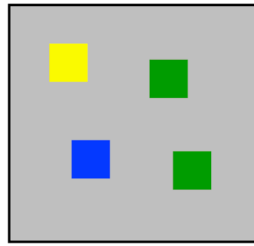
Figure 9.
Proposed structure of memory representations in both simple and real-world displays. (a) In simple displays of meaningless shapes, information is represented both at the item level (perhaps as a hierarchical feature-bundle), and across individual items at the ensemble level. (b) Real world displays have information represented at the object level (as a hierarchical feature bundle), and at the scene level (including scene statistics computed over basic features). In both simple and real-world displays, information is represented at the individual item level, and across individual items, possibly in parallel but interacting processing streams.

a. Encoding rate for real-world stimuli

b. Encoding rate for simple stimuli

Figure 10.
Differential encoding rates for real-world and simple stimuli. (a) Encoding rate for real-world stimuli. For real-world stimuli, sufficient detail to discriminate categorically different items is encoded in less than 1 second, but information continues to accrue over the course of seconds. With enough time, sufficient detail to discriminate items at the exemplar-level or state-level can be encoded. Data and stimuli adapted from Brady et al., 2009. (b) Encoding rate for simple stimuli. For basic features and shapes, information is rapidly encoded into memory, typically reaching an asymptote at or before 100 ms. Data adapted from Vogel, Woodman & Luck, 2001.