Structured representations in visual working memory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEP 1 1 2011

LIBRARIES

by

Timothy F. Brady

B.A., Yale University (2006)

Submitted to the Department of Brain and Cognitive Sciences in partial fulfillment of the requirements for the degree of

ARCHIVES

Doctor of Philosophy in Cognitive Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

© Massachusetts Institute of Technology 2011. All rights reserved.

Author ....
Department of Brain and Cognitive Sciences
June 10, 2011

Certified by ....
Aude Oliva
Associate Professor of Cognitive Science
Thesis Supervisor

Accepted by ....
Earl K. Miller
Picower Professor of Neuroscience
Director, BCS Graduate Program

Structured representations in visual working memory

by

Timothy F. Brady

Submitted to the Department of Brain and Cognitive Sciences
on June 10, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Cognitive Science

Abstract

How much visual information can we hold in mind at once? A large body of research has attempted to quantify the capacity of visual working memory by focusing on how many individual objects or visual features can be actively maintained in memory. This thesis presents a novel theoretical framework for understanding working memory capacity, suggesting that our memory representations are complex and structured even for simple visual displays, and formalizing such structured representations is necessary to understand the architecture and capacity of visual working memory.

Chapter 1 reviews previous empirical research on visual working memory capacity, and argues that an understanding of memory capacity requires moving beyond quantifying how many items people can remember and instead focusing on the content of our memory representations. Chapter 2 argues for structured memory representations by demonstrating that we encode a summary of all of the items on a display in addition to information about particular items, and use both item and summary information to complete working memory tasks. Chapter 3 describes a computational model that formalizes the roles of perceptual organization and the encoding of summary statistics in visual working memory, and provides a way to quantify capacity even in the presence of richer, more structured memory representations. This formal framework predicts how well observers will be able to remember individual working memory displays, rather than focusing on average performance across many displays. Chapter 4 uses information theory to examine visual working memory through the framework of compression, and demonstrates that introducing regularities between items allows us to encode more colors in visual working memory. Thus, working memory capacity needs to be understood by taking into account learned knowledge, rather than simply focusing on the number of items to be remembered. Together, this research suggests that visual working memory capacity is best characterized by structured representations where prior knowledge influences how much can be stored and displays are encoded at multiple levels of abstraction.

Thesis Supervisor: Aude Oliva
Title: Associate Professor of Cognitive Science

Acknowledgments

This thesis would not have been possible without the help of a large number of people. First, I need to thank my advisor, Aude Oliva, who has allowed me free reign to pursue my ideas while always being an advocate for me and providing a broad and creative perspective on research and the scientific process. I also wish to thank Josh Tenenbaum, Ed Vogel and George Alvarez, my thesis committee, for providing useful feedback and mentorship throughout the preparation of this thesis, and the other faculty who have provided intellectual support throughout my time at MIT (in particular, Molly Potter, Jeremy Wolfe, Ruth Rosenholtz and Nancy Kanwisher).

Especially deserving of thanks are Talia Konkle and George Alvarez, by far my most frequent co-authors and contributors to several parts of this thesis. I've been incredibly lucky to work with collaborators who are both a never-ending source of excitement about science, and also excellent co-authors and close friends.

The last five years at MIT have been incredibly fun and rewarding. My labmates – Talia Konkle, Barbara Hidalgo-Sotelo, Soojin Park, and Michelle Greene especially – and my honorary labmates and BCS colleagues – Mike Frank, Ed Vul, Julie Golomb, Todd Thompson and Steve Piantadosi – have been amazing sources of not only engaging and thoughtful conversation, but great fun and constant friendship. Talia and Barbara in particular have made every single day of my graduate school experience more fun and provided just the right amount of encouragement to take breaks *and* to get back to work.

Most importantly, I need to thank my wife, Adena Schachner, for providing constant love and support and improving every aspect of my life. Thanks for being my best friend, companion and being there for me in both life and science.

Finally, I need to thank my parents, Tom and Jeannie Brady, and my sister Jen, who have given me their love and encouragement in everything I've done.

Contents

List of Figures

Chapter 1

Structured representations in visual working memory

1.1 A review of visual working memory capacity[1]

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

[1] Parts of this chapter were published as Brady, T.F, Konkle, T., & Alvarez, G.A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. *Journal of Vision.*

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.

1.1.1 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-drives 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 x 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 x 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 Vogels (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 informa-

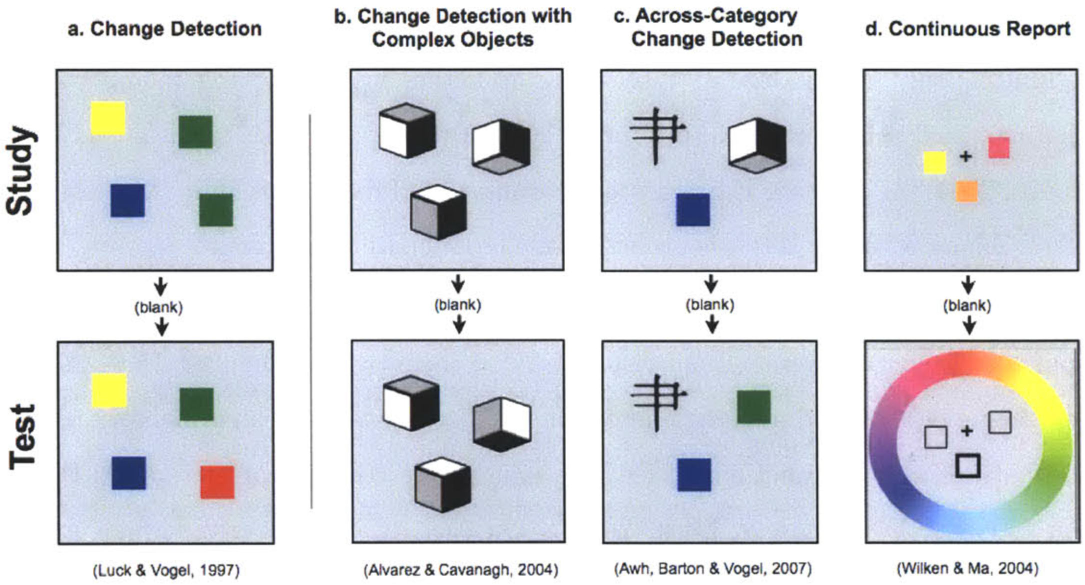

a. Change Detection
b. Change Detection with Complex Objects
c. Across-Category Change Detection
d. Continuous Report
Study
Test
(blank)
(blank)
(blank)
(blank)
(Luck & Vogel, 1997)
(Alvarez & Cavanagh, 2004)
(Awh, Barton & Vogel, 2007)
(Wilken & Ma, 2004)

Figure 1-1: 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.

tion 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 cubes 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 Mas (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).

1.1.2 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 Treismans 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

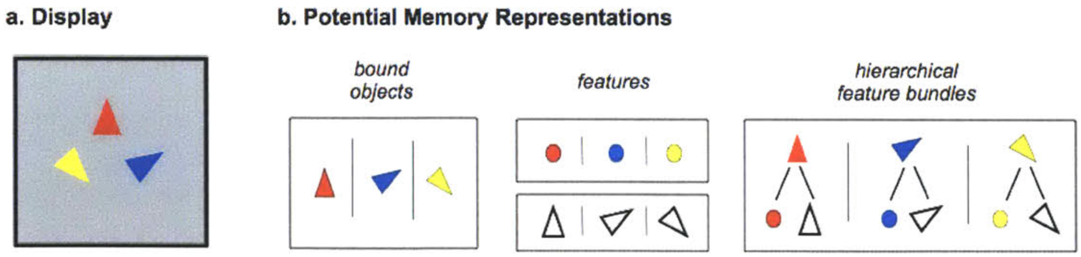

a. Display
b. Potential Memory Representations
bound objects
features
hierarchical feature bundles

Figure 1-2: 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.

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.

1.1.3 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, 2006, 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; Chapter 2). 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; Chapter 2) 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; Chapter 2).

These studies leave open the question of how ensemble representations interact

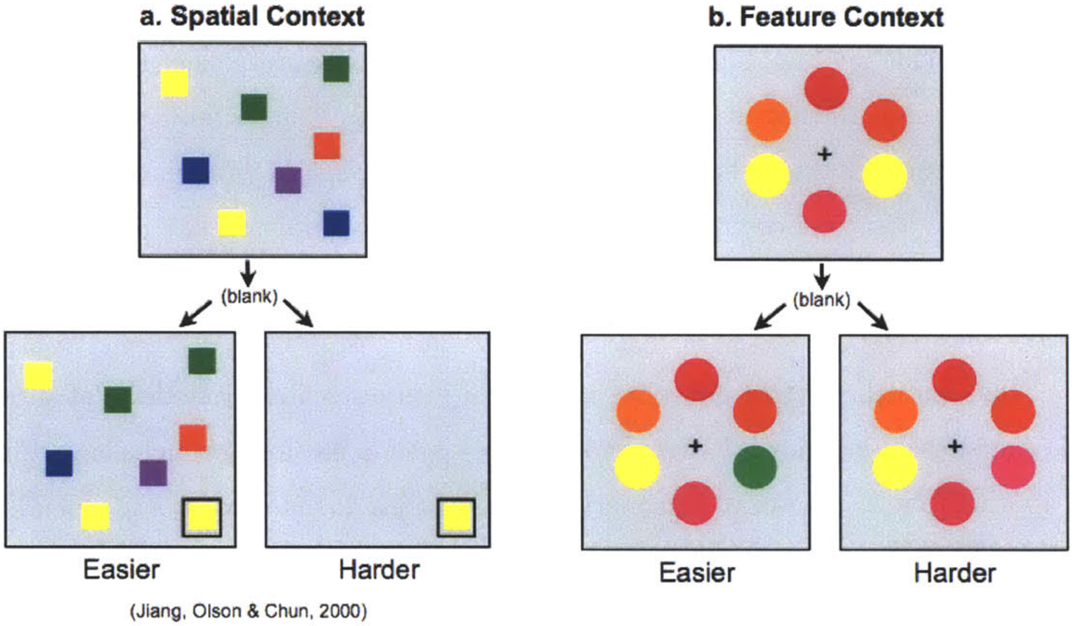


Figure 1-3: Interactions between items in working memory. (a) Effects of spatial context. It is easier to detect a change to an item when the spatial context is the same in the original display and the test display than when the spatial context is altered, even if the item that may have changed is cued (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.

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, & Schner, 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; Chapter 3).

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 (e.g., Chapter 3). 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; Chapter 2). 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 (e.g., Chapter 2; Chapter 3).

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.

1.1.4 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

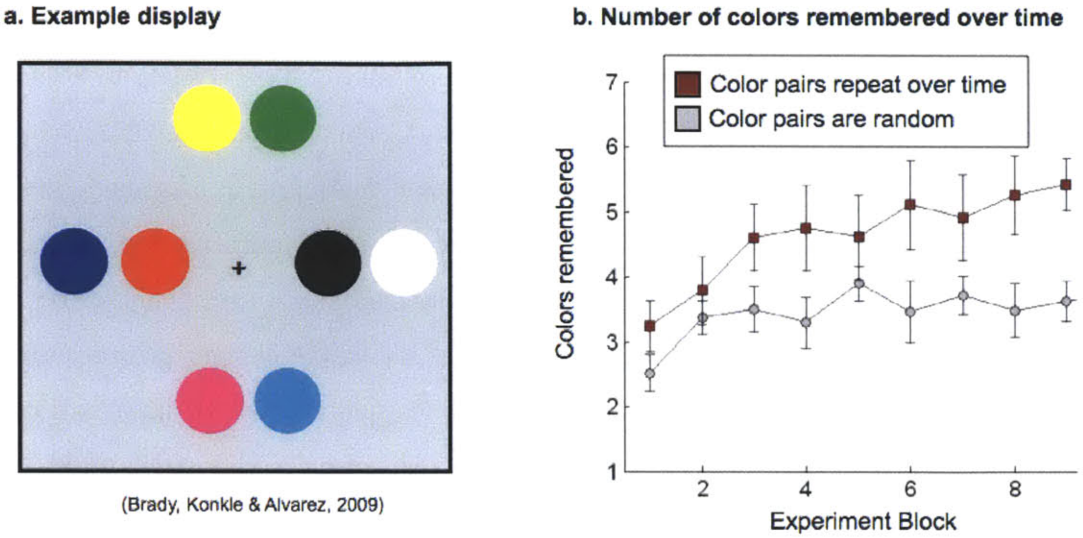
One uncontroversial effect of long-term memory on working memory is that there are biases in working memory resulting from prototypes or previous experience. For example, Huang and Sekuler (2010) have shown that when reporting the spatial frequency of a gabor patch, observers are influenced by stimuli 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 youve 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, childrens 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; Chapter 4) 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



(Brady, Konkle & Alvarez, 2009)

Figure 1-4: Effects of learned knowledge on visual working memory. (a) Example memory display modeled after Brady, Konkle, & Alvarez (2009; Chapter 4). 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; Chapter 4). 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.

representations of familiar color pairs. Furthermore, using an information-theoretic model, Brady, Konkle and Alvarez (2009; Chapter 4) 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; Chapter 4) 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; Chapter 4) 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., Orbn, 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; Chapter 4). 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.

1.1.5 Visual working memory review 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.

1.2 Thesis outline

Visual working memory has often been treated as a system with simple, discrete objects as the unit of storage (e.g., Luck & Vogel, 1997; Cowan, 2005). While many useful models have been built in this framework, and these models have the benefit of being simple and formal (e.g., Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008), such models ultimately depend on the idea that observers' remember discrete items in working memory and that counting how much such items can be remembered is a sufficient measure of capacity.

The claim of this thesis is that such models of visual working memory capacity seriously underestimate the complexity of our memory representations, and thus mischaracterize the nature of our representations even for simple stimuli. In particular, in this thesis I propose that working memory representations are based on existing knowledge and depend critically on both perceptual organization and summary statistics, and thus these factors must be taken into account to accurately characterize our memory representations. I thus aim to investigate working memory capacity from a constructive memory perspective (in the spirit of Bartlett, 1932), rather than quantifying how many independent items can be stored.

The thesis begins with a demonstration that observers represent working memory displays hierarchically – encoding a summary of the display in addition to item-level information (Chapter 2); that they form such hierarchical representations even in standard working memory displays (Chapter 3); and that formal models of working memory can be constructed that allow us to quantify memory capacity in terms of such structured representations (Chapter 3). In addition, I demonstrate that observers take advantage of prior knowledge when representing a display, encoding items more efficiently if they have learned that items are related to each other (Chapter 4), and show that this learning can be formalized using information theory (Chapter 4). Ultimately, the thesis provides empirical evidence that observers use structured knowledge to represent displays in working memory, and provides a set of computational models to formalize these structured memory representations.

Chapter 2

Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items[1]

Influential models of visual working memory treat each item to be stored as an independent unit and assume there are no interactions between items. However, real-world displays have structure, providing higher-order constraints on the items to be remembered. Even displays with simple colored circles contain statistics, like the mean circle size, that can be computed by observers to provide an overall summary of the display. In this chapter we examine the influence of such an ensemble statistic on visual working memory. We find evidence that the remembered size of each individual item is biased toward the mean size of the set of items in the same color, and the mean size of all items in the display. This suggests that visual working memory is constructive, encoding the display at multiple levels of abstraction and integrating across these levels rather than maintaining a veridical representation of each item independently.

[1] Parts of this chapter were published as Brady, T.F, & Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science.*

2.1 Introduction

Observers can quickly and accurately compute ensemble statistics about a display, like the mean size of the items (Ariely, 2001; Chong & Treisman, 2003), the mean facial expression (Haberman & Whitney, 2007), the mean orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), the mean location (Alvarez & Oliva, 2008), and even higher-level spatial layout statistics (Alvarez & Oliva, 2009). However, little work has explored why observers compute these statistics, and in particular, whether the encoding of these higher-order statistics might play a role in how we represent the individual items from such displays in memory.

Nearly all studies of visual working memory use displays consisting of simple stimuli in which the items are chosen randomly. These displays are, as best as possible, prevented from having any overarching structure or gist. Thus, influential models of visual working memory tend to treat each item as an independent unit and assume that items do not influence one another's representation (Alvarez & Cavanagh, 2004; Bays, Catalao & Husain, 2009; Luck & Vogel, 1997; Rouder et al. 2009; Wilken & Ma, 2004; Zhang & Luck, 2008; although see Lin & Luck, 2009 and Johnson, Spencer, Luck and Schner, 2009).

We propose that, contrary to the assumptions of previous models of visual working memory, ensemble statistics allow observers to encode such working memory displays more efficiently: paralleling how people encode real scenes (Lampinen, Copeland, Neuschatz, 2001; Oliva, 2005), observers might encode the 'gist' of simple working memory displays (ensemble statistics like mean size) in addition to information about specific items (their individual information). Such hierarchical encoding would allow observers to represent information about every item in the display simultaneously, significantly improving the fidelity of their memory representations compared to encoding only 3-4 individual items.

To test this hypothesis, we use the ensemble statistic of mean size and the grouping principle of common color, which are known to be automatically and effortlessly computed and could act as a form of higher-order structure in our displays (Chong &

Treisman, 2005). Our results demonstrate a form of hierarchical encoding in visual working memory: the remembered size of individual items is biased towards the mean size of items of the same color and the mean size of all items in the display. This suggests that, contrary to existing models of visual working memory, items are not recalled as independent units, but instead their reported size is constructed by combining information about the specific dot with information about the set of dots at multiple levels of abstraction.

2.2 Experiment 1: Ensemble statistics bias size memory

We examined whether the ensemble statistics of a display would bias memory for individual items in a task where observers attempted to remember the size of multiple colored circles. We hypothesized that on displays with both small red dots and large blue dots, observers would tend to report the size of a particular dot as larger when it was blue than when it was red. This kind of size bias would suggest that observers had taken into account the size of the set of items.

2.2.1 Method

Observers

21 observers were recruited and run using Amazon Mechanical Turk. All were from the U.S., gave informed consent, and were paid 40 cents for approximately 3 minutes of their time.

Procedure

Observers were each presented with the same 30 displays consisting of three red, three blue and three green circles of varying size (see Figure 2-1) and told to remember the size of all of the red and blue circles, but to ignore the green circles. The green distractor items were present in the display because we believed they would encourage

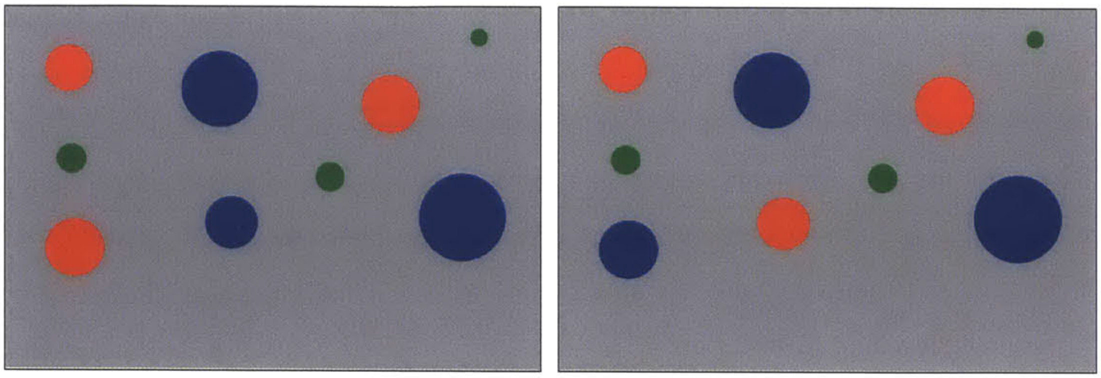


Figure 2-1: An example pair of matched displays from Experiment 1. Observers had to remember the size of the red and blue dots and ignore the green dots. After each trial they were tested on the size of a single dot using a recall procedure. The left display and right display make up a matched pair in which the same items are present, but the tested item (second from the left on the bottom) swaps color with another item (bottom left item). Note that the size of the dots is not to scale in order to more clearly show the display properties.

observers to encode the items by color, rather than selecting all of the items into memory at once (Huang, Treisman, Pashler, 2007; Halberda, Sires, Feigenson, 2006). The order of the 30 displays was randomized across observers. Each display appeared for 1.5 seconds, followed by a 1 second blank, after which a single randomly-sized circle reappeared in black at the location that a red or blue dot had occupied. Observers had to slide the mouse up or down to resize this new black circle to the size of the red or blue dot they had previously seen, and then click to lock in their answer and start the next trial.

Stimuli

The 9 circles appeared in a 600x400 pixel window delineated by a gray box, with each circle at a random location within an invisible 6 x 4 grid, with +/- 10 pixel jitter added to their locations to prevent co-linearities. Observers' monitor size and resolution was not controlled. However, all observers attested to the fact that the entire display was visible on their monitor. Moreover, the critical comparisons are within-subject, and individual differences in absolute size of the display are factored out by focusing on within-subject comparisons between conditions.

Circle sizes were drawn from a separate normal distribution for each color, each with a mean diameter chosen uniformly on each trial from the interval [15px, 95px] and with standard deviation equal to 1/8th their mean. Thus, on a given trial, the three red dots could be sampled from around 35 pixels, the blue dots from 80 pixels and the green dots from 20 pixels. However, which color set was largest and smallest was chosen randomly on each trial; thus, on the next trial it could be the green dots that were largest and the blue dots smallest.

To allow a direct test of the hypothesized bias toward the mean of the same-colored items, the displays were generated in matched pairs. First, 15 displays were generated as described; then another 15 were created by swapping the color of the to-be-tested item with a dot of the other non-distractor color (either red or blue). These 30 displays were then randomly interleaved, with the constraint that paired displays could not appear one after the other. This resulted in 15 pairs of displays, each matched in the size of all of the circles present, with a difference only in the color of the circle that would later be tested. By comparing reported size when the tested item was one color with the reported size when it was another color, we were able to directly test the hypothesis that observers memory for size is biased toward the mean size of all items in the test items color set.

2.2.2 Results

Overall accuracy

We first assessed whether observers were able to accurately perform the size memory task by comparing their performance to an empirical measure of chance (empirical chance: 30.5px, SEM: 0.78px, obtained by randomly pairing a given observers' responses with the correct answer from different trials). Observers performance was significantly better than this measure of chance, with an error of 16.4px on average (SEM: 1.7px; difference from chance, $p< 10^{-9}$).

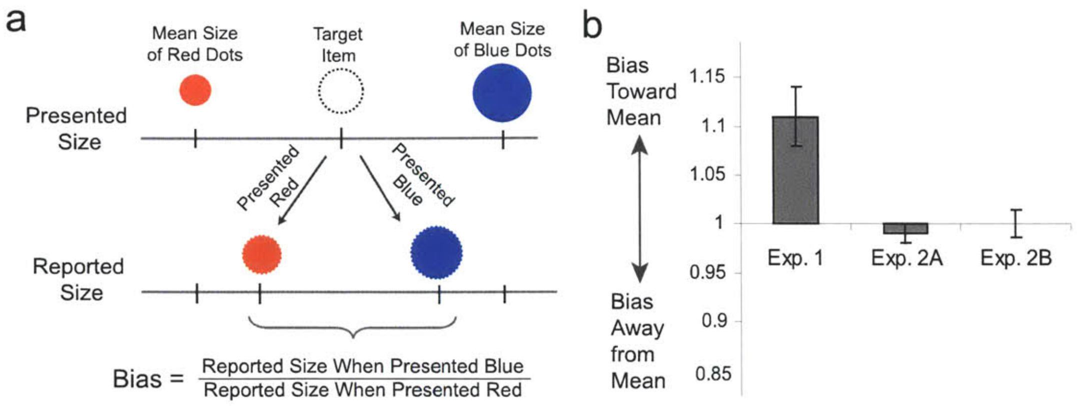


Figure 2-2: (A) Schematic illustration of calculating bias from a pair of matched displays. In this example, the blue dots were larger than the red dots on average. We then measured whether observers reported different sizes for the tested dot when it was red versus when it was blue (the dot was in fact the same size in both presentations). Which color was larger was counterbalanced across trials in the actual experiment, but bias was always calculated by dividing the size reported for the larger color by size reported for the smaller color. (B) The bias from Experiments 1 (color-relevant), 2A (color-irrelevant) and 2B (color-irrelevant). Error bars represent +/- 1 SEM.

Bias from same-colored dots

To address our main hypothesis, we examined whether observers tended to be biased toward the size of the same-colored dots. To do so, we divided the matched pairs based on which of the pair contained smaller same-colored dots on average and which contained larger same-colored dots on average. We then calculated the ratio between the size observers reported in these two cases. If observers were not biased, the ratio between the size observers reported on matched displays with larger versus smaller dots should be 1.0; they should be equally likely to report a larger or smaller size, since the tested item is the same size in each case. However, if observers are biased toward the mean size of the same colored dots, this ratio should be greater than 1.0 (see Figure 2-2).

Observers reported a size on average 1.11 times greater (SEM: +/-0.03) on the half of the displays with larger same-colored dots. This ratio was significantly greater than 1.0 (t(20)=4.17; p=0.0004). In addition, the direction of the effect was highly consistent across observers, with 19 of the 21 observers having a ratio above 1.0.

The maximum possible bias was 1.6, since the same-colored dots were on average 1.6 times larger in these displays than their matched counterparts. Thus the observers reported a size 18% of the way between the correct size and the mean size of the same colored dots. This effect was a result of memory and not a perceptual bias, since in a version of the experiment with a pre-cue indicating which item would be tested, observers (N=22) reported the size accurately (error 6.4px, SEM 0.5px) and with no bias toward the mean size of the same-colored circles (bias: 1.00, S.EM. 0.01).

Model: Optimal Integration Across Different Levels of Abstraction

One interpretation of the data is that observers represent the display at multiple levels of abstraction and integrate across these levels when retrieving the size of the tested dot or when initially encoding its size. To more directly test this idea, we formalized how observers might represent a display hierarchically using a probabilistic model (for similar models, see Huttenlocher et al. 2000; Hemmer & Steyvers, 2009). The model had three levels of abstraction, representing particular dots; all dots of a given color; and all dots on the entire display. In the model, observers encode a noisy sample of the size of each individual dot, and the size of each dot is itself considered a noisy sample from the expected size of the dots of that color, which is itself considered a sample of the expected size of the dots on a given display. Then we ask what observers' ought to report as their best guess about the size of the tested dot (assuming normal distributions at each level).

The intuition this model represents is fairly straightforward: if the red dots on a particular display are all quite large, but you encode a fairly small size for one of the red dots, it is more likely to have been a large dot you accidentally encoded as too small than a small dot you accidentally encoded as too large. Thus, in general the model suggests that the optimal way to minimize errors in your responses is to be biased slightly (either when encoding the dots or when retrieving their size) toward the mean of both the set of dots of the same color and the overall mean of the display. Model predictions, along with an alternative representation of the behavioral data from Experiment 1 are represented in Figure 2-3 (for model implementation see

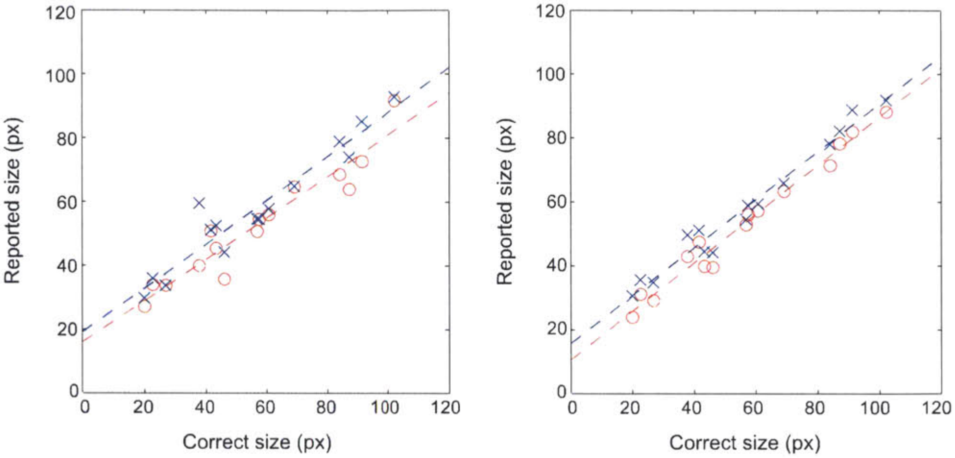


Figure 2-3: (A) Data from Exp. 1, averaged across observers so that each display is a single point. Matched pairs are represented as blue x's for the display in which the same-color dots were larger and red circles for the display in which the same-color dots were smaller. (B) Predictions on the same data from a model that integrates information from multiple levels of abstraction (with SD=25 pixels). Note that in both the observers' data and the model predictions the slope of the line is less than x=y, indicating a bias toward making all dots less extreme in size then they really were, and also note that the blue x's are above the red circles on average, indicating a bias toward the size of the same-color dots.

Chapter Appendix).

The model has a single free parameter, which indicates how noisy the encoding of a given dot is (the standard deviation of the normal distribution from which the encoded size is sampled) and thus how biased toward the means observers' ought be. We set this to 25px in the current experiment by examining the histogram of observers' responses across all of the displays rather than maximizing the fit to the data. While not strictly independent of the data being fit, this method of choosing the parameter is not based on the measures we use to assess the model.

In general the model provides a strong fit to the data on two different metrics: (1) the model predicts the difference between the correct answer and reported answer for each display, ignoring the paired structure of the displays ($r=0.89$, $p<0.0001$); (2) the model predicts the difference in reported size between particular matched displays ($r=0.82$, $p<0.001$). Thus, the model predicts a large amount of the variance even when comparing the matched displays, in which the tested dot is actually the same size for both displays. Any model of working memory which treat items as independent cannot predict a systematic differences on these trials (for example, most slot and resource models, including the mathematical model presented by Zhang & Luck, 2008).

2.2.3 Discussion

We find that observers are biased by the ensemble statistics of the display when representing items in visual working memory. On displays with several different color circles, observers are biased in reporting the size of a given circle by the size of the other circles of that color. This effect is not accounted for by perceptual biases or location noise/swapping, is not a result of observers sometimes guessing based on the mean size of the set of colors (see Chapter Appendix), and is compatible with a simple Bayesian model in which observers integrate information at multiple levels of abstraction to form a final hypothesis about the size of the tested item.

2.3 Experiment 2: Attention to colors is required

In Experiment 1, the color of the items was a task-relevant attribute. In fact, because observers have difficulty attending to more than a single color at a time (Huang, Treisman, Pashler, 2007), observers likely had to separately encode the sizes of the red and blue dots in Experiment 1, perhaps increasing the saliency of the groupings by color. This may be a crucial part of why observers use the mean size of the colors in guiding their memory retrieval. Thus, in Experiment 2A and 2B we removed the green dots from the displays and asked observers to simply remember the sizes of all of the dots. This allowed us to address how automatic the biases we observed in Experiment 1 are; e.g., the extent to which they depend on attentional selection and strategy. In addition, Experiments 2A and 2B provide a control experiment that rules out potential low-level factors that could influence Experiment 1.

2.3.1 Method

25 new observers completed Experiment 2A and a different 20 observers completed Experiment 2B.

The methods and particular 30 displays used were exactly the same as Experiment 1 except that the green dots used as distractor items were not present on the display. The methods of Experiment 2A were otherwise identical to Experiment 1. In Experiment 2B, the dots were shown for only 350ms rather than 1.5 seconds in order to decrease observers performance to the same level as Experiment 1.

2.3.2 Results

Experiment 2A: Overall accuracy

Observers performance in Experiment 2A was very good, with an error of 10.2px on average (SEM: 0.60px). This was significantly less than our empirical measure of chance, ($p < 10^{-19}$; empirical chance: 29px, SEM: 0.28) and significantly less than the error of subjects in Experiment 1 (t(44)=3.73, p<0.001).

Experiment 2A: No bias from same-colored dots

Observers in Experiment 2A displayed no bias as a function of color (M=0.99, SEM: 0.01; not significantly different than 1.0; t(24)=-0.86, p=0.39). This is compatible with the idea that observers do not display a bias when color is not task-relevant.

However, observers in this task had significantly lower error rates than the observers in Experiment 1. Thus, it is possible that observers did not display a bias because they were able to encode all of the dots accurately as individuals. To initially examine this, we selected only the 50% lowest accuracy observers from Exp. 2A and compared them to the 50% highest accuracy observers from Exp. 1. The error rates reverse (Exp.1=10px, Exp 2A=13px), yet the bias remains present only in Exp. 1 (Exp. 1=1.07, Exp. 2A=1.00). This provides preliminary evidence that the difference in accuracy does not drive the difference in bias.

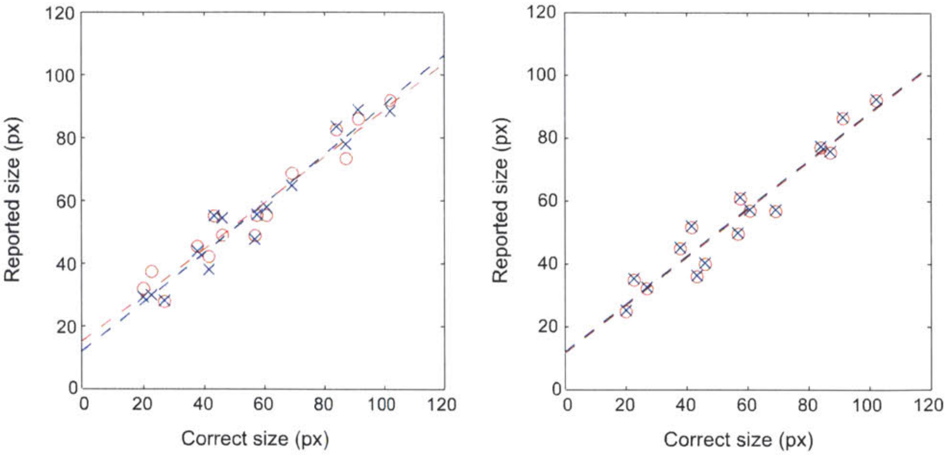


Figure 2-4: (A) Data from Exp. 2B (B) Predictions on the same data from the model (with SD=25 pixels). Note that in both the observers' data and the model predictions the slope of the line is less than x=y, indicating a bias toward making all dots less extreme in size then they really were, and also note that the blue x's are not above the red circles on average, indicating no bias toward the size of the same-color dots.

Experiment 2B: Overall accuracy and bias

Experiment 2B experimentally addressed the concern that the lack of bias for observers in Experiment 2A was driven by their high performance level. In Experiment 2B display time was reduced from 1.5 seconds to 350 milliseconds to increase the error rate while maintaining the task-irrelevance of color. Observers in Experiment 2B had an error rate of 15.9px on average (SEM: 2.25px). This was significantly less than our empirical measure of chance, ($p< 10^{-9}$; empirical chance: 31.3px, SEM: 1.13px) but was no longer significantly less than the error of subjects in Experiment 1 (t(39)=-0.17, p=0.86). Thus, Exp. 1 and Exp. 2B were equated on error rate. However, when color was task-irrelevant in Exp. 2B, there was still no bias from the mean of the same-colored items (M=1.00, SEM: 0.01).

Optimal Integration Model

The same model used in Experiment 1 can be applied to the current data, but with only two levels (no grouping by color): information about the particular dot, and information about all the dots in the display. This model once again provides a strong fit to the experimental data (see Figure 2-4). Since the model does not use color information, it predicts exactly the same performance for both of the matched trials. This is in line with observers a bias of 1.00 in the experimental data. Furthermore, the model predicts the overall bias toward the mean size of the display, correlating with the errors people make across all trials with r=0.53 (p=0.002).

2.3.3 Discussion

In Experiment 2 we find that observers do not display a bias toward the mean size of the same-colored dots when color is not task-relevant, even when the experiments are equated on difficulty. However, observers are still biased toward the mean of the overall display. This is compatible with a Bayesian model in which observers treat all items as coming from a single group, rather than breaking into separate groups by color. Furthermore, the results of this experiment help rule out possible confounds of Experiment 1, such as the possibility that location noise causes swapping of items in memory, since the displays used in Experiments 2A and 2B are exactly the same as those used in Experiment 1 except for absence of irrelevant green dots. We have also run Experiments 1 and 2 as separate conditions in a single within-subject experiment, and replicate the finding of a bias only on displays with green dots present (see Chapter Appendix).

2.4 General Discussion

We find that observers are biased by the ensemble statistics of the display when representing items in visual working memory. When asked to report the size of an individual dot, observers tend to report it as larger if the other items in the same color are large and smaller if the other items in the same color are small. These biases are reliable across observers and predicted by a simple Bayesian model that encodes a display at multiple levels of abstraction. Taken together, these findings suggest that items in visual working memory are not represented independently, and, more broadly, that visual working memory is susceptible to the very same hallmarks of constructive memory that are typical of long-term memory (Bartlett, 1932).

2.4.1 Representation of Ensemble Statistics

It is well established that the visual system can efficiently compute ensemble statistics (e.g., Ariely, 2001; Chong & Treisman, 2003; Alvarez & Oliva, 2009) and does so even when not required to do so by the task, causing, for example, a false belief that the mean of the set was present when asked to remember individual items (Haberman & Whitney, 2009; Fockert & Wolfenstein, 2009). However, less work has explored why the visual system represents ensemble statistics. One benefit of ensemble representations is that they can be highly accurate, even when the local measurements constituting them are very noisy (Alvarez & Oliva, 2008, 2009). Another possible benefit of ensemble representations is that they can be used to identify outliers in a display (Rosenholtz & Alvarez, 2007), which can potentially be used to guide attention to items that cannot be incorporated in the summary for the rest of the group (Brady & Tenenbaum, 2010, Chapter 3; Haberman & Whitney, 2009). The current work suggests a new use of ensemble statistics: such statistics can increase the accuracy with which items are stored in visual working memory, reducing uncertainty about the size of individual items by optimally combining item-level information with ensemble statistics at multiple levels of abstraction.

It is interesting that in the current experiments observers only used the mean size of the colors to reconstruct the display when color was task-relevant, despite the fact that using the mean size of the colors would improve memory for the individual items in all conditions. This could suggest that the units over which such ensemble statistics are computed is limited by selective attention (e.g., Chong & Treisman, 2005b). In a different setting, Turk-Browne, Junge and Scholl (2005) suggest that statistical learning, a form of learning about sequential dependencies, may happen automatically but the particular sets over which the statistics are computed may be controlled by selective attention. This is compatible with what we find in the current experiments: when observers did not attend to the colored sets as separate units separate summary statistics may not have been computed for the two colored sets (alternatively, they may have been encoded, but not used in reconstructing the dot sizes). However, when color was attended, the ensemble statistics for each color seem to have been computed in parallel, as found by Chong and Treisman (2005a).

2.4.2 Dependence Between Items in Visual Working Memory

The current results represent a case of non-independence between items in visual working memory: we find that items are represented not just individually but also as a group or ensemble. While not directly addressing such hierarchical effects, non-independence between items in visual working memory has been observed previously. For example, Huang and Sekular (2010) find that observers tend to be biased in reporting the spatial frequency of Gabors, tending to report frequencies as though they have been pulled toward previously presented Gabor patches. In addition, Jiang, Olson and Chun (2000) have shown that the spatial context of the other items improves change detection performance even when only a single item changes (see also Vidal et al. 2005). This suggests that an item is not represented independent of its spatial context in working memory.

Similarly, work by Brady and Tenenbaum (2010; Chapter 3), Sanocki, Sellers, Mittelstadt & Sulman (2010) and Victor and Conte (2004) shows that observers can take advantage of perceptual regularities in working memory displays to remember more individual items from those displays. Brady and Tenenbaum (2010; Chapter 3) investigate checkerboard-like displays and conceptualize their findings as a kind of hierarchical encoding, in which the gist of the display is encoded in addition to specific information about a small number of items that are least consistent with the gist. This is compatible with the model we present for the simpler displays of the current experiment, in which observers seem to encode ensemble information as well as information about specific items.

This dependence between items in memory is not predicted or explained by influential models of visual working memory. Current theories model visual working memory as a flexible resource, in which memory resources are quantized into slots (Zhang & Luck, 2008) or continuously divisible (Alvarez & Cavanagh, 2004; Bays & Hussain, 2008; Wilken & Ma, 2004). According to these models, fewer items can be remembered with higher precision because they receive more memory resources. However, these models assume that items are stored independently, and therefore cannot account for the dependence between items in memory observed in the current study. Expanding these models to account for the current results will require a specification of whether abstract levels of representation compete for the same resources as item-level representations (e.g., Feigenson, 2008), or whether there are essentially separate resources for ensemble representations and item-level representations (e.g., Brady & Tenenbaum, 2010; Chapter 3).

2.4.3 Long-Term Memory Induced Dependency in Visual Working Memory

In addition to dependencies between items and hierarchical encoding of a particular display, there is a significant amount of previous work showing the representation of items in visual working memory depends on long-term memory information (e.g., Brady, Konkle & Alvarez, 2009). For instance, Konkle and Oliva (2007) and Hemmer and Steyvers (2009), have shown biases in the remembered size of an object after a short delay based on knowledge of the size of that object in the real world. Hemmer and Steyvers (2009) provide a model of this as Bayesian inference in a constructive memory framework, similar to the model we propose for the online representation of a display in the current experiments. Convergence on similar models for using ensemble information from the current display and integrating information from long-term memory suggests a promising future direction for understanding the use of higher-order information in memory.

2.4.4 Conclusion

We find that observers are biased by the ensemble statistics of the display when representing items in visual working memory. Rather than storing items independently, observers seem to construct the size of an individual item using information from multiple levels of abstraction. Thus, despite the active maintenance processes involved in visual working memory, it appears to be susceptible to the very same hallmarks of constructive memory that are typical of retrieval from long-term memory and scene recognition (Bartlett, 1932; Lampinen, Copeland, Neuschatz, 2001). Cognitive and neural models of visual working memory need to be expanded to account for such constructive, hierarchical encoding processes.

2.5 Chapter Appendix

2.5.1 Replication and a Within-Subject Experiment

Running the experiment on the internet allowed for variation in the visual angle of the dots and meant that each observer saw only 30 trials . Thus, we ran a control experiment in the lab with 6 observers using the same paradigm. Observers saw 400 trials each (200 matched pairs). These observers in the lab showed the same effects as observers tested on Mechanical Turk. They had a mean error of 20.2 pixels and a bias of 1.04, significantly greater than 1.0 (t(5)=3.47, p=0.02). The maximum possible bias was 1.37, since the same-colored dots were on average 1.37 times larger in the larger of the matched trials than the smaller. Thus the observers run in the lab reported a size 11% of the way between the correct size and the mean of the same colored dots.

In addition to replicating the experiments in the lab, we also replicated our main results on Mechanical Turk. In particular, to bolster the evidence for our effect we have run both a within-subject experiment (N=17) and replicated both the between-subject experiments (N=16 and N=26, respectively; all conducted on Mechanical Turk). In the within-subject experiment, we combined Exp. 1 with Exp. 2A within observers (thus observers performed 60 trials, 30 with green dots and 30 without green dots present). We found a bias of 1.11 (SEM 0.02) in the trials with green dots and a bias of 1.02 (SEM 0.02) for trials without green dots, a significantly larger bias on green dot trials within-subjects (t(16)=2.90, p=0.01). In addition, the bias was significant in the green-dot displays (t(16)=4.40, p=0.0004) but not the displays without green dots (t(16)=0.82, p=0.42).

In the between-subject replication of Experiment 1 with a different set of displays and different observers, the average bias was 1.09 (N=16), with SEM 0.03. The difference from no bias (1.0) was significant: t(15)=2.29; p=0.037. In the replication of Experiment 2B with a different set of displays and observers, the average bias was 1.00 (N=26; SEM 0.016), not significantly different than 1.00.

2.5.2 Perceptual Effects

Is the bias from same-colored items a result of memory or a perceptual effect caused by crowding or grouping principles in our display? To determine this, we ran a study that was identical to Experiment 1 except that 500ms before the onset of the dots, a single black 'X' appeared at the location of the dot that would later be tested. We instructed observers that this cue indicated which item would be tested (it was 100% valid). If observers have to encode only a single item from the display and know in advance which item will be tested, this should eliminate any bias resulting from memory processes. However, if the locus of our effect is perceptual observers should still be biased toward the size of the same-colored dots. Observers (N=22) reported the size accurately (error 6.4px, SEM 0.5px) and with no bias toward the mean size of the same-colored circles (bias: 1.00, S.E.M. 0.01). This suggests the bias was a result of memory processes, not a perceptual effect from our display.

2.5.3 Potential Reports of the Incorrect Item

Using a similar paradigm but with continuous report of color rather than size, Bays, Catalao and Husain (2009) report that observers sometimes accidentally report the color of the wrong item, perhaps because of noise in their representation of the items' locations. Such location noise would not, in general, affect our conclusion that there is a bias toward the mean of the same colored dots. In particular, if swapping was simply a result of location noise, then since our matched displays contain the exact same size dots in the exact same locations, no difference could arise between them. However, it is possible that observers would be more likely to swap with items in the same color as the target item, and that this could account for the bias we find. If this were the case, we might expect a mixture of correct reports and reports of the incorrect items in our data, resulting in a multimodal distribution. To address this concern, we examined whether the location of the same-colored dots affected the bias we observed, and, additionally, used a mixture model similar to that reported by Bays, Catalao and Husain (2009) to directly examine the possibility of swapping with same-colored items.

To examine the effect of the location of the same-colored dots, we divided the matched pairs by the mean distance of the same-colored dots to the tested dot's location. On those display pairs in which the same-colored dots were much closer in location for one of the matched displays than the other, we might expect a larger bias. Instead, the correlation between the size of the bias and how differently located they were in the two display pairs was not significant, and in fact trended negative (r=-0.27, p=0.33) the opposite of the direction predicted from a swapping account.

As a second measure of the potential of swapping, this time ignoring the location of the items, we used a mixture model to estimate the percentage of swaps directly from the data, effectively examining its bimodality (Bays, Catalao & Husain, 2009). The mixture model attempted to parse the observers' responses into those most likely to have been noisy reports of the correct item, those most likely to have been random guesses, and those most likely to have been swaps . Excluding all responses except those the model considered twice as likely to be noisy reports of the correct item than swaps or guesses still resulted in a substantial bias toward the mean size of the same colored items (M=1.05, SEM:0.016, difference from 1.0: t(20)=3.20, p=0.004). Note that this is an extremely conservative measure, since it effectively counts only responses that are closer to the size of the tested dot than the size of any other dot. Taken together, we believe these analyses help rule out explanations of our data in terms of location noise and reporting the size of the wrong item.

2.5.4 Comparison to an Across-Trial Guessing Model

Rather than performing an integration across different levels of representation on each trial, as proposed in our Bayesian integration model, it is possible that our results could arise from a model in which on some trials observers remember the dot and on other trials the observers' guess based on the dots color. For example, on trials in which the participant retains information about the size of the probed dot, it might be reproduced without bias. On other trials, in which the participant retains no size information about the probed dot, the participant might tend to guess something around the mean of the size of the dots the same color as the probed dot. We will refer to this model as the across-trial guessing model.

While such a model requires observers encode the display at multiple levels of abstraction and integrate across these levels by choosing which kind of information to use in generating a particular response, it is significantly different than the within-trial Bayesian integration model we propose. We believe the evidence from the current experiments heavily supports the within-trial integration model.

First, the across-trial guessing model requires there to be a large number of trials where observers know the color of the tested dot but have no information at all about this dot's size. Both the original work of Luck & Vogel (1997) and important work by Brockmole and colleagues (Logie et al. 2009; Gajewski & Brockmole, 2006) demonstrates that not only is there a benefit to encoding all of the features of a single object, but that observers do so on nearly all trials and represent the objects as bound units. A model which requires observers to frequently know only a single feature of an object is thus theoretically unlikely and in conflict with existing data on binding in visual working memory.

Second, as reported above in the section on modeling location noise and potential item swaps, we can examine trials which are unlikely to have been guess trials by looking at only responses that are closer in size to the size of the correct dot than to the size of any of the other dots (including those of the same color). This still results in a substantial bias toward the mean size of the same-colored dots (see results in location noise section). This is contrary to what you would expect from the across-trial guessing model, which posits a bias arising only from trials where observers do not know the size of the tested dot.

Finally, using model comparison techniques, we can directly compare the distributions predicted by the two models. The within-trial Bayesian integration model assumes the distribution of sizes observers' report for a particular dot has a peak that is shifted toward the mean size of dots of the same color, whereas the across-trial guessing model proposes a mixture between correct responses and responses that are drawn from a distribution around the mean size of the same-colored dots.

The Bayesian model has only a single parameter, the standard deviation of observers' encoding error (this parameter decides both how noisy the distribution is and how much the specific item information is integrated with the ensemble size information). The across-trial guessing model has two parameters, the standard deviation of observers' encoding error and the percentage of trials in which observers report from a distribution around the same-colored mean rather than the correct dot (the guessing rate). In addition, we can choose to make the guessing distribution a normal distribution with the true standard deviation of the dots within the same color, or increase the variance based on the expected sampling error.

For each subject, we performed a leave-one-trial-out cross-validation to find the maximum likelihood parameters for each model. Then we computed the log-likelihood of the observer's response on the left out trial using those parameters. Averaging across all possible left out trials gives us the log-likelihood of each of the two models for each observer. Finally, we can compare these log-likelihoods using AIC (Akaike Information Criterion; Akaike, 1974) . This gives us an AIC score for each model for each observer (lower AIC values indicate a better model fit). We find that across observers, the AIC for the Bayesian model consistently indicates a better fit than the AIC for the across-trial guessing model. This is true both if we assume the guessing distribution is simply a normal with the mean and standard deviation of the true size of the dots of the same color (Bayesian model AIC = 10.8, SEM 0.2, Discrete-guessing model AIC = 13.8, SEM 0.8, t(20)= -3.95, p<0.001) or if we increase this standard deviation by adding in the variance from sampling each dot's size (Discrete-guessing model AIC = 12.7, SEM 0.18, t(20)= -26.8, p$< 10^{-16}$). In fact, using AIC the within-trial Bayesian integration model is preferred in every single observer. Moreover, it is preferred on average even if we do not use AIC to adjust for the greater flexibility of the across-trial guessing model (the log-likelihood of the within-trial integration model is significantly higher than the version of the across-trial guessing model adjusted for measurement error, t(20)=2.23, p=0.038). Thus, in spite of the greater flexibility of the across-trial guessing model, it does not fit the data as well as the within-trial Bayesian integration model.

2.5.5 Optimal Observer Model

To more directly test the idea that observers' represent the display at multiple levels of abstraction and integrate across these levels when retrieving the size of the tested dot, we formalized this theory in a probabilistic model. In the model, observers are assumed to get a single noisy sample from each of the 9 dots on the screen (sampled from a normal distribution centered around the size of the dot and with a standard deviation of 25px). Then, the observer attempts to infer the size of each of the dots on the screen using these samples. A nave, non-hierarchical model simply treats each of the dots independently and thus report the size of each dot as the size that was sampled for that dot. As an alternative, we present a hierarchical Bayesian model that pools information from all of the dots to best estimate the size of any given individual dot. It does so by representing the display at two additional levels of abstraction and partially pooling information at each of these levels: (1) all dots of the same color; (2) all dots on the display. By assuming that dots of the same color and all the dots on a display are sampled from some underlying distribution and therefore provide mutual information about each other, such a model arrives at a more accurate estimate of the size of each dot. Such models are standard in Bayesian statistics (Gelman, Carlin, Stern & Rubin, 2003) and have been previously applied to similar problems in cognitive science (Huttenlocher et al. 2000; Hemmer & Steyvers, 2009).

Formally, we assume that observers' treat the dots of a given color as sampled from a normal distribution with unknown mean and unknown variance, and additionally treat these distributions' means as coming from an overall normal distribution that pools information across all of the colors. We put uniform priors over the reasonable range of possible sizes (0-200 pixels) on the parameters of these normal distributions. The exact model is represented in WINBUGS as follows. Note that the normal distribution in WINBUGS is parameterized by a mean and a precision, rather than a mean and standard deviation; nevertheless we put a uniform prior on standard deviation, which is a more standard model (Gelman, Carlin, Stern & Rubin, 2003).

WinBUGS code for the model in Experiment 1:

```
model
% C = number of colors,
% L = number of dots of each color.
% We observe 'sample'.
{
  overallMean ~ dunif(0,100)
  overallMeanStd ~ dunif(0,100)
  overallMeanPrec <- 1/(overallMeanStd*overallMeanStd)

  overallStd ~ dunif(0,100)
  overallStdStd ~ dunif(0,100)
  overallStdPrec <- 1/(overallStdStd*overallStdStd)

  stdev <- 25
  precision <- 1/(stdev*stdev)

  for (i in 1:C)
  {
    groupMean[i] ~ dnorm(overallMean, overallMeanPrec)
    groupStd[i]  ~ dnorm(overallStd, overallStdPrec)
    groupPrec[i] <- 1/(groupStd[i]*groupStd[i])
  }

  for (i in 1:C)
  {
    for (j in 1:L)
    {
      dotMean[i,j] ~ dnorm(groupMean[i], groupPrec[i])
      sample[i,j]  ~ dnorm(dotMean[i,j], precision)
    }
  }
}
```

WinBUGS code for the model in Experiment 2:

```
model
{
  overallMean ~ dunif(0,200)
  overallMeanStd ~ dunif(0,100)
  overallMeanPrec <- 1/(overallMeanStd*overallMeanStd)
  stdev <- 10
  precision <- 1/(stdev*stdev)
  for (i in 1:L)
  {
    dotMean[i] ~ dnorm(overallMean, overallMeanPrec)
    sample[i]   ~ dnorm(dotMean[i], precision)
  }
}
```

Chapter 3

A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates.[1]

When remembering a real-world scene, people encode both detailed information about specific objects and higher-order information like the overall gist of the scene. However, formal models of change detection, like those used to estimate visual working memory capacity, assume observers encode only a simple memory representation which includes no higher-order structure and treats items independently from each other. In this chapter, we present a probabilistic model of change detection that attempts to bridge this gap by formalizing the role of perceptual organization and allowing for richer, more structured memory representations. Using either standard visual working memory displays or displays in which the dots are purposefully ar-

[1] Parts of this chapter were published as Brady, T.F, & Tenenbaum, J.B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. *Proceedings of the Cognitive Science Society.*

ranged in patterns, we find that models which take into account perceptual grouping between items and the encoding of higher-order summary information are necessary to account for human change detection performance. We conclude that even in simple visual working memory displays, items are not represented independently. Thus, models of visual working memory need to be expanded to take into account this non-independence between items before we can make useful predictions about observers' memory capacity, even in simple displays.

3.1 Introduction

Working memory capacity constrains cognitive abilities in a wide variety of domains (Baddeley, 2000), and individual differences in this capacity predict differences in fluid intelligence, reading comprehension and academic achievement (Alloway & Alloway, 2010; Daneman & Carpenter, 1980; Fukuda, Vogel, Mayr & Awh, 2010). The architecture and limits of the working memory system have therefore been extensively studied, and many models have been developed to help explain the limits on our capacity to hold information actively in mind (e.g., Cowan, 2001; Miyake & Shah, 1999). In the domain of visual working memory, these models have grown particularly sophisticated and have been formalized in an attempt to derive measures of the capacity of the working memory system (Alvarez & Cavanagh, 2004; Bays, Catalao & Husain, 2009; Cowan, 2001; Luck & Vogel, 1997; Wilken & Ma, 2004; Zhang & Luck, 2008). However, these models focus on how observers encode independent objects from extremely simple displays of segmented geometric shapes.

By contrast to these simple displays, memory for real-world stimuli depends greatly on the background knowledge and principles of perceptual organization our visual system brings to bear on a particular stimulus. For example, when trying to remember real-world scenes, people encode both the gist and detailed information about some specific objects (Hollingworth & Henderson, 2003; Oliva, 2005). Moreover, they use the gist to guide their choice of which specific objects to remember (Friedman, 1979; Hollingworth & Henderson, 2000), and when later trying to recall the details of the scene, they are influenced by this gist, tending to remember objects that are consistent with the scene but were not in fact present (Brewer & Treyens, 1981; Lampinen, Copeland & Neuschatz, 2001; Miller & Gazzaniga 1998).

In fact, even in simple displays, perceptual organization and background knowledge play a significant role in visual working memory. For example, what counts as a single object may not be straightforward, since even the segmentation of the display depends on our background knowledge about how often the items co-occur. For instance, after learning that pairs of colors often appear together, observers can encode nearly twice as many colors from the same displays (Brady, Konkle, Alvarez, 2009; see Chapter 4). Displays where objects group together into perceptual units also result in better visual working memory performance, as though each unit in the group was encoded more easily (Woodman, Vecera & Luck, 2003; Xu & Chun, 2007; Xu, 2006). Furthermore, observers are better able to recognize changes to displays if those changes alter some statistical summary of the display; for example, if a display is changed from mostly black squares to mostly white squares, observers notice this change more easily than a matched change that does not alter the global statistics (Victor & Conte, 2004; see also Alvarez & Oliva, 2009).

There is thus significant behavioral evidence that even in simple visual working memory displays, items are not treated independently (for a review, see Brady, Konkle & Alvarez, 2011; Chapter 1). However, existing formal models of the architecture and capacity of visual working memory do not take into account the presence of such higher-order structure and prior knowledge. Instead, they most often depend on calculating how many individual items observers remember if the items were treated independently.

3.1.1 Existing models of capacity estimates

The most common paradigm for examining visual working memory capacity is a change detection task (e.g., Luck & Vogel, 1997; Pashler, 1988). In a typical change detection task, observers are presented with a study display consisting of some number N of colored squares (see Figure 3-1). The display then disappears, and a short time

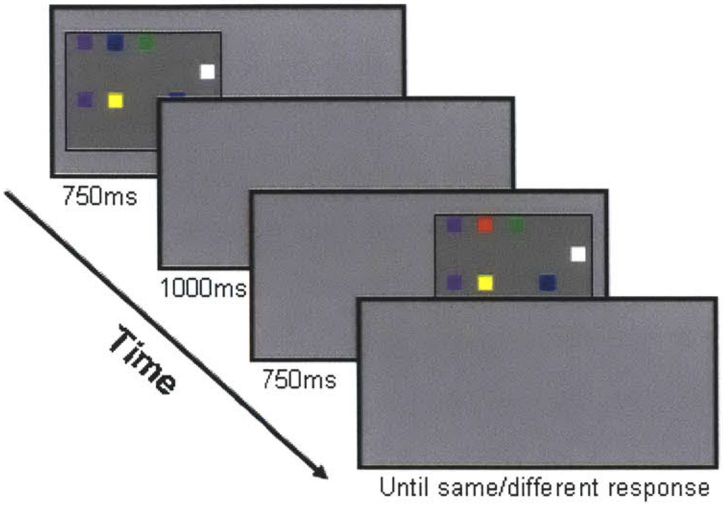


Figure 3-1: Methods of a change detection task (as used in Experiments 1 and 2). Observers are first briefly presented with a display (the study display), and then after a blank are presented with another display where either the items are exactly the same or one item has changed color (the test display). They must say whether the two displays were the same or different.

later another display reappears that is either identical to the study display or in which a single square has changed color. Observers must decide whether this test display is identical to the study display or there has been a change. Observers are told that at most a single item will change color.

The standard way of reporting performance in such a visual working memory task is to report the "number of colors remembered", often marked by the letter 'K'. These values are calculated using a particular model of change detection (a 'slot model'), which supposes that the decline in observers' performance when more squares must be remembered is caused solely by a hard limit in the number of items that can be remembered (Cowan, 2001; Pashler, 1988). Such estimates thus assume complete independence between the items.

For example, imagine that an observer is shown a display of N colored squares and afterwards shown a single square, and asked if it is the same or different as the item that appeared at the same spatial location in the original display (Cowan, 2001). According to the slot model of change detection, if the observer encoded the item in memory, then they will get the question correct; and this will happen on $\frac{K}{N}$ trials. For example, if the observer can encode 3 items and there are 6 on the display, on 50% of the trials they will have encoded the item that is tested and will get those 50% of trials correct. Such models suppose no noise in the memory representation: if the item is encoded it is remembered perfectly. On the other hand, if the observer does not encode the item in memory, then the model supposes they guess randomly (correctly choose same or different 50% of the time). Thus, the total chance of getting a trial correct is:

$$PC(\%) = \frac{K}{N} * 100\% + \frac{N-K}{N} * 50\%$$

By solving for 'K', we can take the percent correct at change detection for a given observer and determine how many items they remembered out of the N present on each trial (Cowan, 2001). Such modeling predicts reasonable values for a variety of simple displays (e.g., Vogel, Woodman & Luck, 2001; Cowan, 2001, 2005), suggesting observers have a roughly fixed capacity of 3-4 items, independent of a number of factors that affect percent correct (like set size, N).

However, nearly all visual working memory papers report such values, often without considering whether the model that underlies them is an accurate description of observers working memory representation of those stimuli. Thus, even in displays where observers perform grouping or encode summary statistics in addition to specific items, many researchers continue to report how many items observers can remember (K values) using the standard formula in which each item is treated as an independent unit (e.g., Brady, Konkle, Alvarez, 2009; Xu & Chun, 2007). This results in K values that vary by condition, which would indicate a working memory capacity that is not fixed In these cases, the model being used to compute capacity is almost certainly incorrect observers may not be encoding items independently.

Other models have also been used to quantify working memory capacity (e.g., Bays, Catalao & Husain, 2009; Wilken & Ma, 2004; Zhang & Luck, 2008). However, these models also operate without taking into account the presence of higher-order structure and prior knowledge, as they model displays that are sampled uniformly, limiting any overarching structure or gist. It is thus difficult to make claims about observers capacities using such models. Due to the nature of the models (e.g. multinomial processing tree models like Cowan's K), it is also extremely difficult to expand existing models to account for summary representations, or representations of items which are not independent of one another.

3.1.2 Change detection as Bayesian inference

In this paper we reformulate change detection as probabilistic inference in a generative model. We first formalize how observers encode an initial study display, and then we model the change detection task as an inference from the information about the test display and the information in memory to a decision about whether a change occurred. Modeling change detection in this Bayesian framework allows us to use more complex and structured knowledge in our memory encoding model (e.g., Tenenbaum, Griffiths & Kemp, 2006), allowing us to make predictions about memory capacity under circumstances where items are non-independent or summary statistics are encoded in addition to specific items.

We begin by modeling a simple change detection task, as shown in Figure 3-1 and described above. To create a model of this change detection task, we first specify how observers' encode the study display. For the simplest case, in order to most closely match the models used to quantify K-values in standard displays (Cowan, 2001), we assume that memory takes the form of a discrete number of slots, K, each of which stores which color was present on the display in a particular location (using seven categorically distinct colors: black, white, red, green, blue, yellow and purple). Also in line with standard slot models, we initially assume that observers choose which K of the N dots to encode at random. To model the change detection task, we then formalize how observers make a decision about whether there was a change when the test display is presented.

When observers must decide if there has been a change, they have access to all of the items in the test display and to the items they encoded in memory from the study display. Using the information that at most a single item can change color between the two displays, we model the observer as performing an optimal inference to arrive at a judgment for whether the display has changed. In other words, the observer places probabilities on how likely each possible display is to have been the study display, and then effectively "rules out" all possible displays inconsistent with the items in memory and all displays that have more than a single change from the test display. They can then arrive at a probability that indicates how likely it is that the study display was exactly the same as the test display. Interestingly, this Bayesian model of change detection has Cowan's K as a special case (for details, see Chapter Appendix, 3.8.1).

Importantly, however, by framing the model in terms of probabilistic inference we make explicit the assumptions about the architecture of working memory the model entails. First, we assume that observers remember information about a specific subset K of the N items. Second, we assume that memory for these items is without noise. Both of these assumptions are simply properties of the probability distributions we choose and can be relaxed or generalized without changing the model architecture. Thus, the Bayesian framework we adapt allows a much greater range of memory architectures to be tested and made explicit.

3.1.3 The current experiments

In the current paper we use this reformulated model of change detection to examine the use of higher-order information in visual working memory. While such higher-order information can take many forms, we begin with two possible representations: (1) a model that encodes both specific items and also a summary representation (how likely neighboring items are to be the same color); and (2) a model in which observers first use basic principles of perceptual organization to chunk the display before encoding a fixed number of items.

To examine whether such representations can account for human memory performance, we not only look at the overall level of performance achieved by using a particular memory representation in the model, but also examine how human performance varies from display to display. In Experiments 1A and 1B, we test our proposed memory representations on displays where the items are purposefully arranged in patterns. In Experiment 2, we generalize these results to displays of randomly chosen colored squares (as in Luck & Vogel, 1997). We show for the first time that observers are highly consistent in which changes they find easy or difficult to detect, even in standard colored square displays. In addition, we show that models which have richer representations than simple slot models provide good fits to the difficulty of individual displays, because these more structured models representations capture which particular changes people are likely to detect. By contrast, the simpler models of change detection typically used in calculations of visual working memory capacity (e.g., the model underlying K-values) do not predict any reliable differences in difficulty between displays. We conclude that even in simple visual working memory displays items are not represented independently, and that models of working memory with richer representations are needed to understand observers' working memory capacity.

3.2 Experiments 1A and 1B: Patterned dot displays

Rather than being forced to treat each item as independent, our Bayesian model of change detection can be modified to take into account the influences of perceptual organization, summary statistics and long-term knowledge. We thus had observers perform a memory task with displays where the items were arranged in spatial patterns. Observers are known to perform better on such displays than on displays without patterns (e.g., Garner, 1974; see also Hollingworth, Hyun & Zhang, 2005; Phillips, 1974; Sebrechts & Garner, 1981). Because observers' memory representations in these displays are likely to be more complex than simple independent representations of items, such displays provide a test case for modeling higher-order structure in visual working memory. To examine the generality of observers' memory representations, we used two similar sets of stimuli (Exp 1A: red and blue circles; Exp 1B: black and whites squares), which vary basic visual properties of the stimuli but keep the same high-level grouping and object structure.

3.2.1 Method

Observers

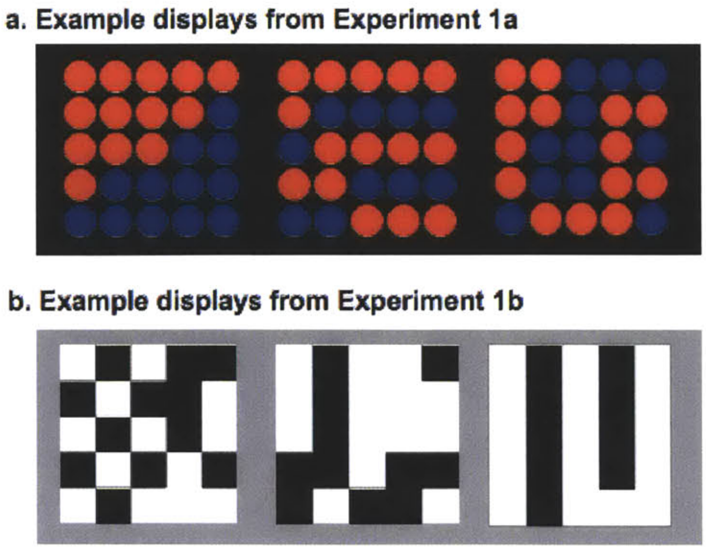
130 observers were recruited and run using Amazon Mechanical Turk (see Brady & Alvarez, 2011 for a validation of using Mechanical Turk for visual working memory studies). All were from the U.S., gave informed consent, and were paid 30 cents for approximately 4 minutes of their time. 65 of the observers participated in Experiment 1A and 65 different observers in Experiment 1B.

Procedure

To examine human memory performance for patterned displays, we had observers perform a change detection task. We showed each of our observers the exact same set of 24 displays (see Figure 3-1). Each display was presented to each observer in both a "same" and "different" trial, so observers completed 48 trials each. On each trial, the study display was presented for 750ms, followed by a 1000ms blank period; then either an identical or a changed version of this original display was presented for 750ms in a different screen location (the test display). Observers' task was simply to indicate, using a set of buttons labeled 'Same' and 'Different', whether the two displays were identical or whether there had been a change. The order of the 48 trials was randomly shuffled for each subject. Observers started each trial manually by clicking on a button labeled 'Start this trial', after which the trial began with a 500ms delay.

Stimuli

Unlike traditional displays used to assess visual working memory capacity, we used displays where the items to be remembered were not simply colored squares in random



a. Example displays from Experiment 1a

b. Example displays from Experiment 1b

Figure 3-2: a) Example study displays from Experiment 1A b) Example study displays from Experiment 1B. In both Experiments 1A and 1B, some displays were generated by randomly choosing each item's color, and some were generated to explicitly contain patterns.

locations but also exhibited some higher-order structure (as in Philips, 1974). As stimuli we created 24 displays that consisted of 5x5 patterns in which each space was filled in by a red or blue circle (Exp. 1A) or the same patterns were filled with black or white squares (Exp. 1B). The patterns could be anything from completely random to vertical or horizontal lines (see Figure 3-2). Our displays were thus simple relative to real scenes but were complex enough that we expected existing models, which encode independent items, would fail to predict what observers remember about these displays. 8 of the 24 displays were generated by randomly choosing the color of each dot. The other 16 were generated to explicitly contain patterns (For details of how we generated the patterned displays, see Chapter Appendix, 3.8.2).

The displays each subtended 150x150 pixels inside a 400pixel by 180pixel black (Exp 1A) or gray (Exp 1B) box. On each trial, the pre-change display appeared on the left of the box, followed by the (potentially) changed version of the display appearing on the right side of the box. Observers' monitor size and resolution was not controlled. However, all observers attested to the fact that the entire stimulus presentation box was visible on their monitor.

3.2.2 Results

For each display we computed a d', measuring how difficult it was to detect the change in that particular display (averaged across observers). The stimuli in Experiments 1A were exactly the same as those in 1B, except that the patterns were constructed out of red and blue dots in Experiment 1A and black and white squares in Experiment 1B. As expected, performance in Experiments 1A and 1B was highly similar: the correlation in the display-by-display d' was r=0.91 between the two experiments. As a result, we collapsed performance across both experiments for the remaining analyses, though the results remain qualitatively the same when considering either experiment alone.

On average, human observers d' was 2.18 (S.E. 0.06), suggesting that observers were quite good at detecting changes on these displays. Since the displays each contain 25 dots, this d' corresponds to a K value of 17.8 dots if the items are assumed to be represented independently and with no summary information encoded (Pashler, 1988).

In addition, observers were highly consistent in which displays they found most difficult to detect changes in (see Figure 3-3). We performed split-half analyses, computing the average d' for each display using the data from a randomly-selected half of our observers, and then comparing this to data from the other half of the observers. The same displays were difficult for both groups (r=0.75, averaged over 200 random splits of the observers). Computing d' separately for each display and each observer is impossible as each observer saw each displays only once. Thus, to compute standard errors on a display-by-display basis we used bootstrapping. This provides a visualization of the display-by-display consistency (Figure 3-3). Some displays, like those on the left of Figure 3-3, are consistently hard for observers. Others, like those on the right of Figure 3-3, are consistently easy for observers to detect changes in.

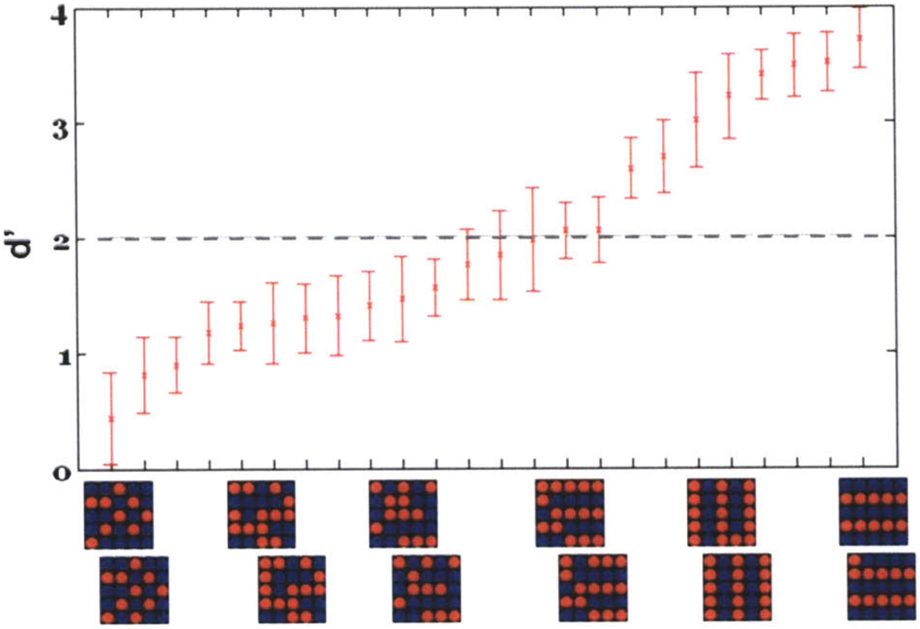


Figure 3-3: Consistency in which displays are most difficult in Exp 1A. The x-axis contains each of the 24 display pairs, rank ordered by difficulty (lowest d' on the left, highest on the right; for visualization purposes, only a subset of pairs are shown on the x-axis). The top display in each pair is the study display; the bottom is the test display with a single item changed. The dashed gray line corresponds to the mean d' across all displays. The error bars correspond to across-subject standard error bars. The consistent differences in d' between displays indicate some displays are more difficult than other displays.

3.2.3 Conclusion

In Experiments 1A and 1B, we assessed observers' visual working memory capacity for structured displays of red and blue dots or black and white squares. We found multiple aspects of human performance in this task which conflict with the predictions of standard models of visual working memory.

First, we find that observers perform much better in detecting changes to these displays than existing working memory models would predict. Under existing models, in which items are assumed to be represented independently with no summary information encoded, observers d' in this task would correspond to memory for nearly 18 dots (Pashler, 1998). This is nearly 5 times the number usually found in simpler displays (Cowan, 2001), and thus presents a direct challenge to existing formal models of change detection and visual working memory capacity.

Furthermore, observers are reliable in which changes they find hard or easy to detect. This consistent difference between displays cannot be explained under a model in which observers treat the items independently. Previous formal models of change detection treat all of the displays as equivalent, since all displays change only a single item's color and all contain an equal number of items. They thus make no predictions regarding differences in difficulty across displays, or regarding which particular changes will be hard or easy to detect.

To account for the high level of performance overall and the consistent differences in performance between displays, it is necessary to posit a more complex memory representation or encoding strategy. We next consider two alternative models for what information observers might be encoding: a model in which observers encode both an overall summary of the display (e.g., "it looked like it contained vertical lines") in addition to information about particular items, and a model in which observers 'chunk' information by perceptually grouping dots of the same color into single units in working memory. These models formalize particular hypotheses about what representations observers encode from these displays. They thus allow us to examine whether observers performance is compatible with a fixed working memory capacity in terms of some format of representation other than a fixed number of independent items.

3.3 Summary-based encoding model

In real-world scenes, observers encode not only information about specific objects but also information about the gist of the scene (e.g., Lampinen, Copeland, Neuschatz, 2001). In addition to this semantic information, observers encode diffuse visual summary information in the form of low-level ensemble statistics, which they make use of even in simple displays of gabors or circles (Alvarez & Oliva, 2009; Brady & Alvarez, 2011). For example, in a landmark series of studies on summary statistics of sets, Ariely (2001) demonstrated that observers extract the mean size of items from a display, and, moreover, store it in memory even when they have little to no information about the size of the individual items on the display (Ariely, 2001; see Alvarez, 2011; Haberman & Whitney, 2011 for reviews). Observers seem to store not only summary information like mean size but also spatial summary information, like the amount of horizontal and vertical information on the top and bottom of the display (Alvarez & Oliva, 2009) and even high-level summary information like the mean gender and emotion of faces (Haberman & Whitney, 2007). Furthermore, observers integrate this summary information with their representation of particular items: for example, Brady and Alvarez (2011) have shown that observers use the mean size of items on a display to modulate their representation of particular items from that display.

To examine whether such summary representations could underlie performance on our patterned displays, we built a model that formalized such a summary-based encoding strategy. We posited that observers might encode both a spatial summary of the display and particular 'outlier' items that did not match this summary. Our modeling proceeded in two stages, mirroring the two stages of the change detection task: view and encode the study display, then view the test display and decide if a change occurred.

More specifically, in the summary-based encoding model we propose that observers use the information in the study display to do two things: first, they infer what summary best describes the display; then, using this summary, they select the subset of the dots that are the biggest outliers (e.g., least well captured by the summary) and encode these items specifically into an item-based memory. As a simplifying assumption, we use a summary representation based on Markov Random Fields which consists of just two parameters: one representing how likely a dot in this display is to be the same or different than its horizontal neighbors and one representing how likely a dot is to be the same or different than its vertical neighbors. This summary representation allows the model to encode how spatially-smooth a display is both horizontally and vertically, thus allowing it to represent summaries that are approximately equivalent to "vertical lines", "checkerboard", "large smooth regions", etc.

After a short viewing, the study display disappears and the observer is left with only what they encoded about it in memory. Then a test display appears and the observer must decide, based on what they have encoded in memory, whether this display is exactly the same as the first display. Thus, at the time of the test display (the change detection stage), the observer has access to the test display and both the item-level and summary information from the study display that they encoded in memory. Using the constraint that at most one item will have changed, it is then possible to use Bayesian inference to put a probability on how likely it is that a given test display is the same as the study display and, using these probabilities, to calculate the likelihood that the display changed.

For example, an observer might encode that a particular display is relatively smooth (horizontal neighbors are similar to each other, and vertical neighbors are also similar to each other), but that the two items in the top right corner violate this assumption, and are red and blue respectively. Then, when this observer sees the test display, they might recognize that while both items they specifically encoded into an item memory are the same color they used to be, the display does not seem as smooth as it initially was: there are some dots that are not like their horizontal or vertical neighbors. This would lead the observer to believe there was a change, despite not having specifically noticed what items changed.

Importantly, when this model encodes no higher-order structure it recaptures the standard slot-based model of change detection. However, when the displays do have higher-order regularities that can be captured by the models' summary representation, the model can use this information to both select appropriate individual items to remember and to infer properties of the display that are not specifically encoded. For a formal model specification, see Chapter Appendix, 3.8.3.

3.3.1 Modeling results and fit to human performance

In Experiment 1, we obtained data from a large number of human observers detecting particular changes in a set of 24 displays. For each display observers saw, we can use the summary-based encoding model to estimate how hard or easy it is for the model to detect the change in that display. The model provides an estimate, for a given change detection trial, of how likely it is that there was a change on that particular trial. By computing this probability for both a 'same' trial and a 'change' trial, we can derive a d' measure for each display in the model.

The model achieves the same overall performance as observers (d'=2.18) with a 'K' value of only 4, thus encoding only 4 specific dots in addition to the display's summary (model d'=1.2, 1.8, 2.05, 2.25 at K=1, 2, 3, 4). This is because the model does not represent each dot independently: instead, it represents both higher-order information as well as information about specific dots.

Furthermore, the model correctly predicts which display observers will find easy and which displays observers will find difficult. Thus, the correlation between the model's d' for detecting changes in individual displays and the human performance on these displays is quite high ($r=0.72$, $p<0.00001$ with $K=4$; averaging observers' results across Exp 1A and 1B, see Figure 3-4). Importantly, this model has no free parameters other than how many specific items to remember, K, which we set to K=4 based on the model's overall performance, not its ability to predict display-by-display difficulty. Thus, the model's simple summary representation captures which changes people are likely to detect and which they are likely to miss.

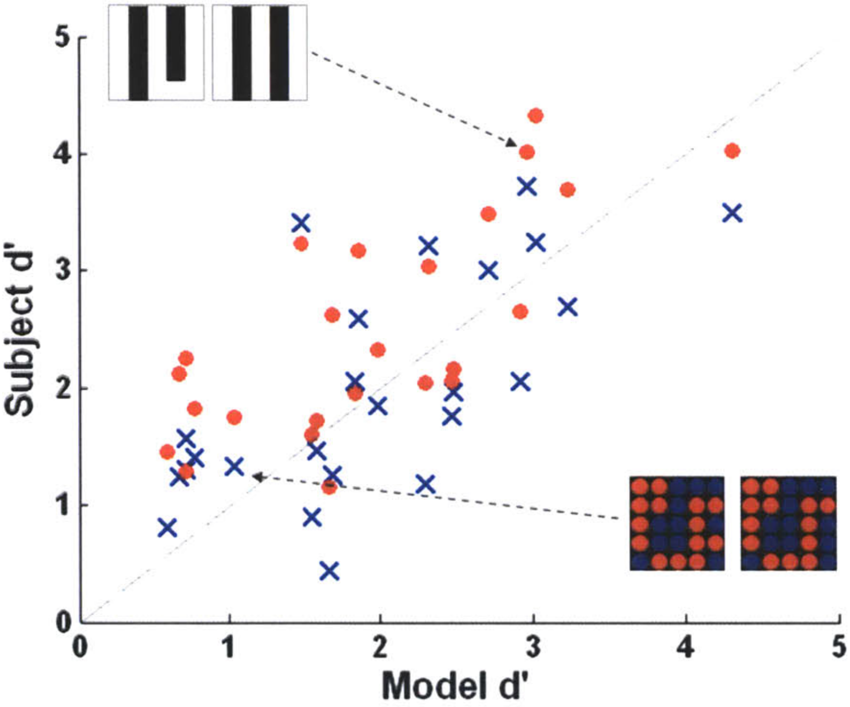


Figure 3-4: The fit of the summary-based encoding model with K=4 to the observers' data for Experiments 1A (blue x's) and 1B (red circles). Each point is the d' for a particular display. Example of both a hard and easy pair of displays is shown.

3.3.2 Necessity of the summary representation

The summary-based encoding model posits that observers encode a summary representation of the display and use this summary to choose outlier items to encode into a specific item memory. However, it is possible that a single one of these processes might account for the model's fit to human data. For example, it is possible that simply choosing outlier items using a summary representation, but not encoding the actual summary representation into memory is sufficient to capture human performance. Alternatively, it is possible that simply encoding a summary representation but not using this representation to encode outlier items is sufficient to explain human performance.

To address this and examine the necessity of each component of the model's representation, we 'lesioned' the model by looking at model predictions without one or the other of these components.

Choosing outlier items but not remembering the summary representation

Is remembering the summary representation helping us to accurately model human performance, or can we predict human performance equally well by using the summary to choose outliers to encode into memory but then discarding the summary representation itself? To examine this, we looked at the fit of a model that did not have access to the summary representation at the time of change detection, and detected changes solely based on the specific objects encoded.

We find that such a model does not fit human performance as well as the full summary-based encoding model. Firstly, to achieve human levels of performance such a model must encode as many objects as a model which encodes objects completely at random (human levels of performance at K=18; d'=0.47, 0.92, 1.30, 1.69, 2.27 at K=4, 8, 12, 16, 20). Furthermore, this model does not accurately predict which specific changes will be noticed, either at K=4 (r=0.30, p=0.15) or at K=18 (r=0.39, p=0.06), accounting for at most 28% of the amount of the variance that is accounted for by the full model.

One reason this model does not fit human performance as well as the full model is that it fails to recognize changes that introduce irregular items: e.g., if the initial display is quite smooth and has no outliers, this model simply encodes items at random. Then, if the 'change' display has an obvious outlier item the model cannot detect it. To recognize this kind of change requires knowing what the summary of the initial display was.

Thus, if we remove the memory representation of the summary from the model, it provides a significantly worse fit to human performance.

Remembering a summary representation but choosing items at random

It is also possible to examine a model that encodes both a summary of the display and specific items, but does not choose which items to specifically encode by selecting outliers from the summary. Rather than preferentially encoding unlikely items, such a model chooses the items to encode at random.

We find that such a model does not fit human performance as well as the full summary-based encoding model. To achieve human levels of performance such a model must encode as many objects as a model which encodes objects completely at random (human levels of performance at K=20; d'=0.26, 0.54, 0.91, 1.39, 2.06 at K=4, 8, 12, 16, 20). Furthermore, it does not do a good job predicting which specific changes will be noticed, either at K=4 (r=0.09, p=0.68) or at K=20 (r=0.40, p=0.05), accounting for at most 31% of the amount of the variance that is accounted for by the full model. One reason this model fails to fit human performance is that it fails to recognize changes that remove irregular items: e.g., if the initial display is quite smooth but has a single outlier, it will be encoded as a relatively smooth display. Then, if the 'change' display removes the outlier item the model cannot detect it. To recognize this kind of change requires maximizing your information about the first display by encoding specific items that are not well captured by the summary.

Thus, if remove the model's ability to encode outlier items, it also provides a significantly worse fit to human performance.

3.3.3 Conclusion

Typically, we are forced to assume that observers are representing independent objects from a display in order to calculate observers' capacity. By using a Bayesian model that allows for more structured memory representations, we can calculate observers' memory capacity under the assumption that observers remember not just independent items, but also a summary of the display. This model provides a reasonable estimate of the number of items observers are remembering, suggesting only 4 specific items in addition to the summary representation must be maintained to match human performance. The model thus aligns with both previous work from visual working memory suggesting a capacity of 3-4 simple items (Luck & Vogel, 1997; Cowan, 2001) and also with data from the literature on real-world scenes and simple dot displays which suggests a hierarchical representation with both gist and item information (Lamplin et al. 2001; Brady & Alvarez, 2011; Chapter 2).

Furthermore, because the summary-based model does not treat each item independently, and chooses which items to encode by making strategic decisions based on the display's summary, this model correctly predicts the difficulty of detecting particular changes. By contrast, a model which assumes we encode each item in these displays as a separate unit and choose which to encode at random can predict none of the display-by-display variance. This model thus represents a significant step forward for formal models of change detection and visual working memory capacity.

3.4 Chunk-based encoding model

Rather than encoding both a summary of the display and specific items, it is possible that observers might use a chunk-based representation. For example, a large number of working memory models assume a fixed number of items can be encoded into working memory (Cowan, 2001; Luck & Vogel, 1997). To account for apparently disparate capacities for different kinds of information, such models generally appeal to the idea of chunking, first explicated by George Miller (Miller, 1956). For example, Miller reports on work which found that observers could remember 8 decimal digits and approximately 9 binary digits. By teaching observers to recode the binary digits into decimal (e.g., taking subsequent binary digits like 0011 and recoding them as '3'), he was able to increase capacities up to nearly 40 binary digits. However, observers remembered these 40 digits using a strategy that required them to remember only 7-8 'items' (recoded decimal digits). Ericsson, Chase and Faloon (1980) famously reported a similar case where a particular observer was able to increase his digit span from 7 to 79 digits by recoding information about the digits into running times from various races he was familiar with, effectively converting the 79 digits into a small number of already-existing codes in long-term memory. More recently, Cowan et al. (2004) have found that by teaching observers associations between randomly chosen words in a cued-recall task, observers can be made to effectively treat a group of two formerly unrelated words as a single 'chunk' in working memory, and that such chunking seems to maintain a fixed capacity in number of chunks even after learning.

In the domain of visual working memory, little work has explicitly examined chunking or what rules apply to grouping of items in visual working memory. In part, this is because visual working memory representations seem to be based on representing objects and features, and so it may not be possible to recode them into alternative formats to increase capacity without using verbal working memory. However, some work has focused on how learning associations impacts which items are encoded into memory (Olson & Jiang, 2004; Olson, Jiang & Moore, 2005) and which items are represented as a single chunk (Orbn, Fiser, Aslin & Lengyel, 2008). Furthermore, it has been shown that learned associations can even result in greater numbers of individual items being encoded into memory (Brady, Konkle & Alvarez, 2009). However, almost no work has formalized the rules behind which items are perceptually grouped and count as a single 'unit' in a slot-model of visual working memory (although see Woodman, Vecera & Luck, 2003; Xu & Chun, 2007; and Xu, 2006 for examples of perceptual grouping influencing capacity estimates in visual working memory).

A simple hypothesis is that the basic Gestalt rules of perceptual grouping, in this case grouping by similarity (Wertheimer, 1938; Koffka, 1935), will determine the perceptual units that are treated as single units in visual working memory. Indeed,

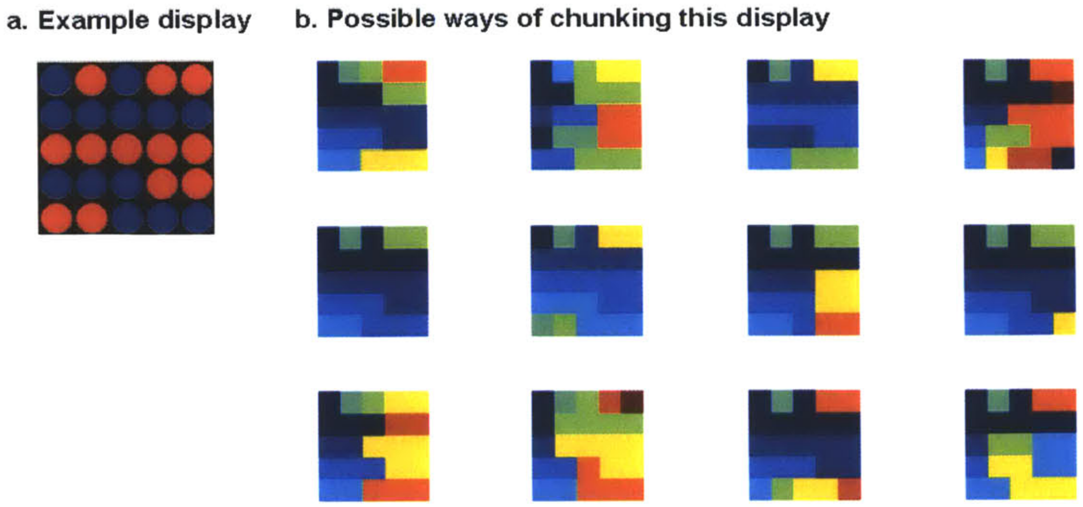


Figure 3-5: (a) An example display, (b) Several possible ways of chunking this display. These are 12 independent samples from our probabilistic chunking model with the smoothness parameter set to 4. Each color represents a particular 'chunk'.

some work has attempted to examine how observers might group adjacent items of similar luminance together in order to remember more dots in displays much like the displays we use in the current task (Halberda et al submitted; see also Hollingworth, Hyun & Zhang, 2005). However, little formal work has been done examining how well such a model accounts for human change detection, and no work has examined whether such a model predicts which displays will be easy or difficult to detect changes in.

To model such a chunking process, we added two components to our basic change detection model. First, rather than encoding K single objects, we encode up to K regions of a display. Second, to select these regions we use two factors, corresponding to the Gestalt principles of proximity and similarity: (1) a spatial smoothness term that encourages the model to put only adjacent items into the same chunk; (2) a likelihood term that forces the model to put only items of the same color into the same chunk. We thus probabilistically segment the display into M regions, and then select which K of these M regions to encode by preferentially encoding larger regions (where chance of encoding is proportional to region size; e.g., we are twice as likely to encode a region of 4 dots as a region of 2 dots). This allows us to examine how likely an observer that encoded a display in this way would be to detect particular changes for different values of K (see Figure 3-5 for a sample of possible region-segmentations for a particular display).

In this model, we examine the possibility that observers use the information in the first display to form K regions of the display following the principles of proximity and similarity, and then encode the shape and color of these K regions into memory. Then, the second display appears and the observer must decide, based on what they have encoded in memory, whether this display is exactly the same as the first display. They do so by independently judging the likelihood of each dot in the second display, given the chunks they have encoded in memory. This model has a single free parameter, a smoothness parameter which affects how likely adjacent items of the same color are to end up in the same chunk. For values >0, this parameter prefers larger chunks to smaller chunks, since it prefers neighboring items to have the same chunk-label. The model is relatively insensitive to the value of this parameter for values >= 1.0. For all simulations, we set this value to 4.0 because this provided a model that created different segmentations of the display fairly often, while still making those segmentations consist of relatively larger chunks. For full model specification, see Chapter Appendix, 3.8.4.

3.4.1 Modeling results and fit to human performance

The chunk-based model provides an estimate, for a given change detection trial, of how likely it is that there was a change on that particular trial. By computing this probability for both the 'same' trial and a 'change' trials that observers saw in Experiment 1, we can derive a d' for each display in the model.

The model achieves the same performance as people (d'=2.18) with a K value of only 4, thus encoding only four chunks of dots (model d'=0.44, 0.93, 1.49, 2.08, 2.69 at K=1, 2, 3, 4, 5). This is because the model does not represent each dot independently: instead, it represents grouped sets of dots as single chunks.

Furthermore, because the chunk-based model does not treat each item independently, the model makes predictions about the difficulty of detecting particu-

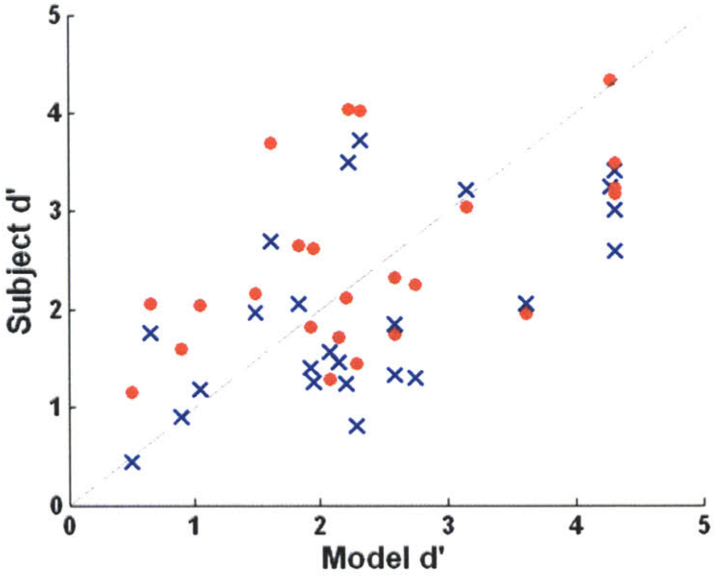


Figure 3-6: The fit of the chunk-based encoding model with K=4 (Smoothness=4) to the observers' data for Experiments 1A (blue x's) and 1B (red circles). Each point is the d' for a particular display.

lar changes. In fact, the correlation between the model's difficulty with individual displays and the human performance on these displays was relatively high (r=0.58, p=0.003; see Figure 3-6).

At K=4, we can examine the effect of different values of the smoothness parameter on this correlation rather than simply setting this parameter to 4. We find that this correlation is relatively robust to the smoothness preference, with r=0.35, r=0.45, r=0.45, r=0.58, r=0.58 for values of 1, 2, 3, 4, and 5 (with smoothness = 5, the model always segments the display into the largest possible chunks). Thus, the model's simple summary representation captures which changes people are likely to detect and which they are likely to miss independently of the settings of the chunk-size free parameter.

3.4.2 Conclusion

The chunk-based model provides a reasonable estimate of the number of items observers are remembering, suggesting only 4 chunks need be remembered to match human performance. The model thus provides evidence that fits with previous work from visual working memory suggesting a capacity of 3-4 simple items (Luck & Vogel, 1997; Cowan, 2001), with the addition of a basic perceptual organization process that creates chunks before the items are encoded into memory. Furthermore, because the chunk-based model does not treat each item independently, this model makes predictions about the difficulty of detecting particular changes. These predictions coincide well with observers difficulty in detecting particular changes in particular displays. Together with the summary-based encoding model, this chunk-based model thus provides a possible representation that might underly human change detection performance in more structured displays.

3.5 Combining the summary-based and chunk-based models

Both the chunking model and summary-based encoding model capture a significant amount of variance, explaining something about which displays observers find difficult to detect changes in. Do they explain the same variance? Or do both models provide insight into what kinds of representations observers use to represent these displays? To assess this question, we examined whether combining these two models resulted in a better fit to the data than either model alone.

The summary-based encoding model and chunk-based model's display-by-display d' predictions are almost totally uncorrelated with each other (r=0.03), despite both doing a reasonable job predicting which displays people will find difficult. We thus averaged the predictions of the two models and looked at whether this provides a better fit to human performance than either model alone. We find that the average of the two models together results in an impressive fit to human performance (r=0.90, p<0.00001; see Figure 3-7). In fact, the two models together account for 81% of the variance in observers' d-prime across displays without any free parameters set to maximize this correlation. This is compatible with the idea that observers' represen-

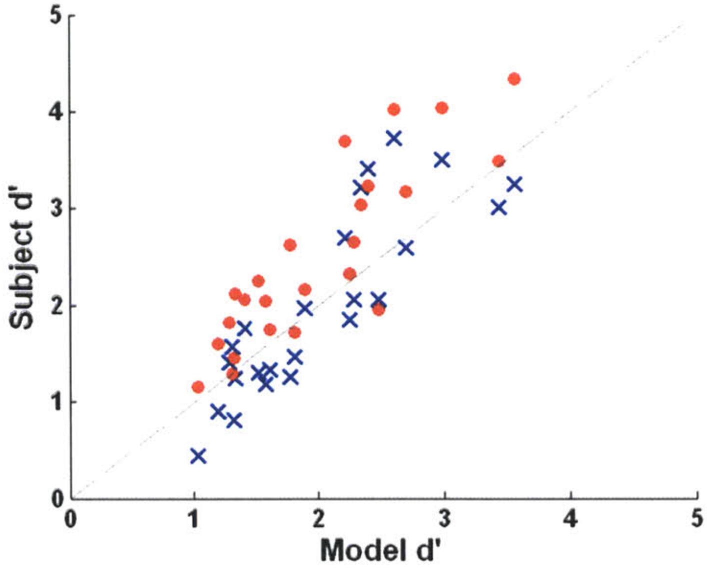


Figure 3-7: The fit of combined model with K=4 in both models to the observers' data for Experiments 1A (blue x's) and 1B (red circles). Each point is the d' for a particular display. Combining the predictions of the summary-based and chunking models results in a better fit to the human data than either model alone.

tations might sometimes be more chunk-based and sometimes be more hierarchical, perhaps depending on their explicit strategy or perhaps because different displays lend themselves to different styles of encoding.

Rather than simply averaging the model's predictions, we can also introduce a parameter that weights the two models unequally. A linear model producing the best fit weights for the predictions of the summary-based and chunk-based models yields a best fit of r=0.92, with weights of 0.67 for the summary-based encoding model and 0.45 for the chunk-based encoding model (intercept: -0.13). While weighting the summary-based encoding model more produces a better fit, such a model does not significantly enhance our ability to fit observers' display-by-display difficulty.

3.5.1 Conclusion

We examined whether a Bayesian change detection model with more structured memory representations can provide a window into observers' memory capacity. We find that both a summary-based encoding model that encodes specific items and also a summary representation, and a chunking-based model in which observers first use basic principles of perceptual organization to chunk the display before encoding a fixed number of items provide possible accounts for how observers encode patterned displays. These models can match human levels of accuracy while encoding only 3-4 items or chunks, and moreover, provide a good fit to display-by-display difficulty, accurately predicting which changes observers will find most difficult. Furthermore, the two models seem to capture independent variance, indicating that perhaps observers use both kinds of representations when detecting changes in patterned displays. Taken together, the two models account for 81% of the variance in observers' d-prime across displays. By contrast, the simpler models of change detection typically used in calculations of visual working memory capacity do not predict any reliable differences in difficulty between displays because they treat each item independently. These models thus represent a significant step forward for formal models of change detection and visual working memory capacity.

3.6 Experiment 2: Randomly colored displays

Using a Bayesian model of change detection together with more structured memory representations allows us to examine observers' working memory capacity in displays with explicit patterns. Can these models also predict which displays are hard or easy on displays without explicit patterns, as in most typical visual working memory experiments (e.g., Luck & Vogel, 1997)? If so, what are the implications for standard K values and for simple models of working memory capacity based on these values?

While most working memory experiments generate displays by randomly choosing colors and placing those color at random spatial locations, this does not mean that there are no regularities present in any given display. In fact, any particular working memory display tends to have significant structure and regularities present even though on average the displays are totally random.

Variance in observers encoding or storage in particular displays can have a significant influence on models of memory capacity. For example, Zhang and Luck (2008)

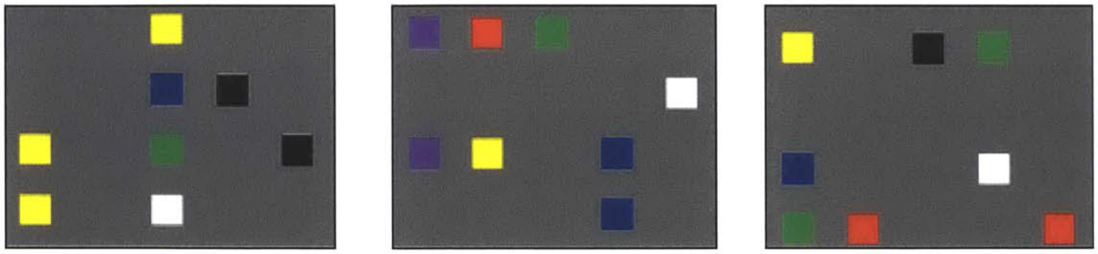


Figure 3-8: Example displays from Experiment 2. These displays were generated randomly by sampling with replacement from a set of 7 colors, as in Luck & Vogel (1997).

used a continuous report task (based on Wilken & Ma, 2004) in which observers are briefly shown a number of colored dots and then asked to report the color of one of these dots by indicating what color it had been on a color wheel. They then modeled observers' responses to partial out observers' errors into two different kinds (noisy representations and random guesses), arriving at an estimate of the number of colors observers remember, on average, across all of the displays. They found evidence that supported the idea that observers either remember the correct answer or completely forget it, and used this to argue for a model of working memory model in which observers can encode at most three items at a time (a quantized resource model).

Importantly, however, by fitting their model only to the results across all displays rather than taking into account display-by-display variability, they failed to model factors that influence the overall capacity estimate, but average out when looking at many different displays. For example, Bays, Catalao and Husain (2009) showed that many of observers' 'random guesses' are actually reports of an incorrect item from the tested display. Reports of the incorrect item tend to average out when looking at all displays, but for each individual display make a large difference in how many items we should assume observers' were remembering. Once these incorrect reports are taken into account, Bays, Catalao and Husain (2009) find that the model of Zhang and Luck (2008) no longer provides a good fit to the data. This suggests that display-by-display factors can sometimes significantly influence the degree to which a particular model of working memory is supported, despite a good fit to the average across all displays.

In the current experiment, we sought to examine whether display-by-display variance in encoding particular working memory displays could be formalized using our Bayesian model of observers' memory representations. We applied the same models used in the patterned displays in Experiment 1 the summary-based encoding model and chunk-based model to displays like those used in typical visual working memory experiments. We find evidence that observers use such structured representations when encoding these displays, and are able to predict which particular displays observers will find easy or difficult to detect changes in. This indicates that simple models of working memory which encode a small number of independent objects at random do not match the representation observers' use even in relatively simple working memory displays.

3.6.1 Methods

Observers

100 observers were recruited and run using Amazon Mechanical Turk. All were from the U.S., gave informed consent, and were paid 30 cents for approximately 4 minutes of their time.

Procedure and Stimuli

We randomly generated 24 pairs of displays by selecting 8 colors with replacement from a set of 7 possible colors (as in Luck & Vogel, 1997) and placing them randomly on a 5 x 4 invisible grid (see Figure 3-8). While it is standard to jitter the items in such displays to avoid co-linearities, to facilitate modeling and comparison with the previous experiments we allowed the items to be perfectly aligned.

The displays each subtended 320x240 pixels, with the individual colored squares subtending 30x30 pixels. On each trial, the pre-change display appeared on the left, followed by the (potentially) changed version of the display appearing on the right. Observers' monitor size and resolution was not controlled. However, all observers attested to the fact that the entire stimulus presentation box was visible on their monitor.

The methods were otherwise the same as Experiment 1.

3.6.2 Results

For each display we computed a d', measuring how difficult it was to detect the change in that display (averaged across observers). The mean d' was 1.5 across the displays, corresponding to a K value of 4.0 if we assume all of the items are represented independently (Pashler, 1988).

However, as in Experiment 1, observers were consistent in which displays they found easy or difficult (see Figure 3-9). For example, if we compute the average d' for each display using the data from half of our observers and then do the same for the other half of the observers, we find that to a large degree the same displays were difficult for both groups (r=0.68, averaged over 200 random splits of the observers). By bootstrapping to estimate standard errors on observers' d-prime for each individual display we can visualize this consistency (Figure 3-9). Some displays, like those on the left of Figure 3-9, are consistently hard for observers. Others, like those on the right of Figure 3-9, are consistently easy for observers to detect changes in. Contrary to the assumption of standard working memory models, observers do not appear to treat items independently even on randomly generated displays like those typically used in working memory experiments.

Model fits

We next fit the summary-based encoding model and the chunk-based model to these data to examine whether these models capture information about observers representations in these displays. We find that the summary-based model provides a good fit to the data, and in addition correlates with observers' display-by-display difficulty (see Figure 3-10). The summary-based encoding model equals human performance (d'=1.5) at K=4 (d'=1.47 at K=4), and at this K value correlates with display-by-display difficulty well (r=0.60; p=0.003). Furthermore, this correlation is

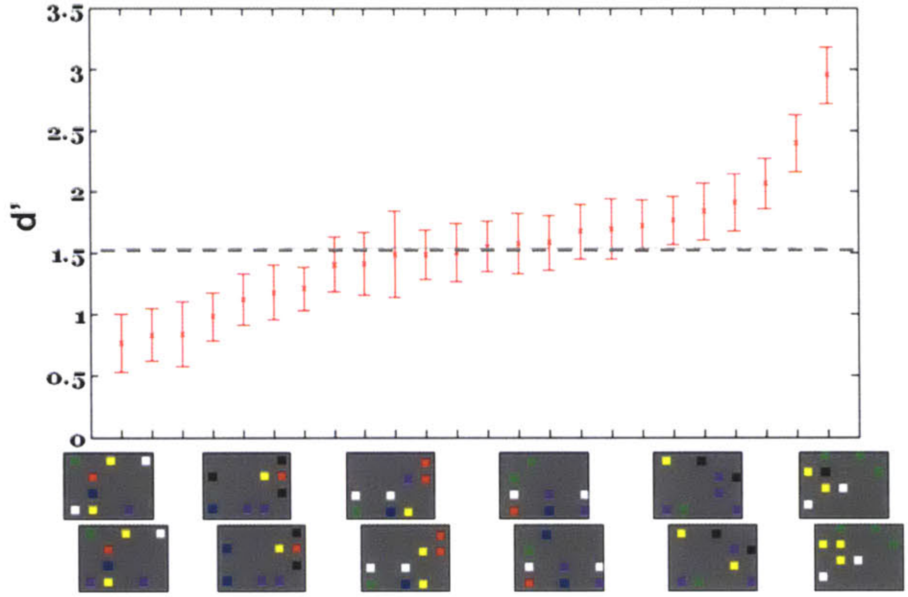


Figure 3-9: Consistency in which displays are most difficult in Exp 2. The x-axis contains each of the 24 display pairs, rank ordered by difficulty (lowest d' on the left, highest on the right; for visualization purposes, only a subset of pairs are shown on the x-axis). The dashed gray line corresponds to the mean d' across all displays. The error bars correspond to across-subject standard error bars. The consistent differences in d' between displays indicate some displays are more difficult than other displays.

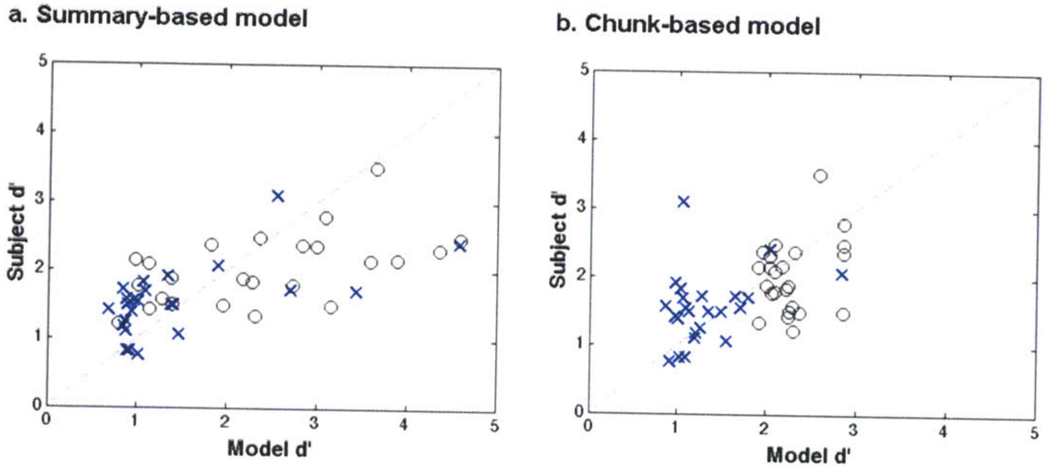


Figure 3-10: (a) Fit of the summary-based model with K=4. The blue xs represent the data from Experiment 2, using randomly generated displays as in typical visual working memory experiments (fit: r=0.60). The black circles represent data from the control experiment where displays were generated to purposefully contain patterns (fit: r=0.55). (b) Fit of the chunk-based model with K=4. The blue xs represent the data from Experiment 2, using randomly generated displays as in typical visual working memory experiments (fit: r=0.32). The black circles represent data from the control experiment where displays were generated to purposefully contain patterns (fit: r=0.28).

not driven by the outliers: the Spearman rank-order correlation is also high (r=0.53, p=0.009) and if we exclude displays where the model predicts an excessively high d, the correlation remains high despite the decreased range (excluding displays with model d>3, r=0.61). The chunk-based model does not provide as good a fit, equaling human performance at K=4 (d'=0.88 at K=3, d'=1.32 at K=4, d'=1.81 at K=5) but only marginally correlating with display-by-display difficulty (r=0.33 at K=3, r=0.32 at K=4, r=0.41 at K=5). In addition, combining the chunking model with the summary-based model does not significantly improve the summary-based model, with the average of the two models giving a slightly worse fit than the summary-based model alone (with K=4 for both models, r=0.56).

Generating the displays used in Experiment 2 completely at random means that few displays contained significant enough pattern information to allow for chunking or summary information to play a large role. This allowed us to quantify exactly how well our model representations explained data from truly random displays, as used in most working memory studies (e.g., Luck & Vogel, 1997). However, while we find that even with a sample of just 24 displays some displays are easier than others and this is well-explained by our summary-based model, the limited range in d' prevents any strong conclusions about the particular memory representations observers make use of in displays of colored squares (for example, do observers' representations truly resemble the summary-based model more than the chunk-based model?). We thus ran another experiment (N=100 observers that did not participate in Experiments 1 or 2) using 24 new displays we generated to purposefully contain patterns[2]. Within these new displays, we found that the summary-based model once again provided a strong fit to the data (r=0.55) whereas the chunk-based model provided a considerably worse fit (r=0.32). In addition, when combining the displays from this control experiment with the displays from Experiment 2, we find that the summary-based model provides a better fit (r=0.64) than the chunk-based model (r=0.50) and averaging the two models does not improve the fit of the summary-based model significantly (r=0.66).

This suggests that the summary-based model's representation provides a better fit to how observers encode these working memory displays than the chunk-based model does. This could be because the distance between the items prevents low-level perceptual grouping from occurring (Kubovy, Holcombe & Wagemans, 1998).

3.6.3 Conclusion

Even in standard working memory displays, observers are consistent in which displays they detect changes in and which displays they do not detect changes in. This suggests that the assumption of independence between items does not hold even in these relatively simple displays of segmented shapes. Thus, we need models that take into account basic perceptual grouping and higher-order summary representations in order to understand the architecture of visual working memory even when our displays are impoverished relative to real scenes.

[2] We generated the displays by creating displays at random and retaining only displays where either the chunk-based model or the summary-based predicted the display would have a d' greater than 2.

Interestingly, even in displays chosen to minimize the presence of patterns, our summary-based model's representation captures which changes people are likely to detect and which they are likely to miss. By contrast, a model which assumes we encode each item in these displays as a separate unit and choose which to encode at random can predict none of the display-by-display variance. This suggests that observers' representation are more structured than standard models based on independent items would suggest, even in simple working memory displays.

3.7 General Discussion

We presented a formal model of change detection which relies upon Bayesian inference to make predictions about visual working memory architecture and capacity. This model allows us to take into account the presence of higher-order regularities, while making quantitative predictions about the difficulty of particular working memory displays.

In Experiment 1, we found that observers are able to successfully detect changes to displays containing spatial patterns with much greater accuracy than would be expected if they were remembering only 3-4 individual items from these displays. Furthermore, we found that observers are highly reliable in which particular changes they find easiest and hardest to detect. We posited two memory representations that might underlie observers performance on these displays: a summary-based representation, where observers encode both a spatial summary of the display (items tend to be the same color as their horizontal neighbors) and outlier items; and a chunk-based representation, where observers group individual items into chunks before encoding them into memory. Using our change detection model, we demonstrated that both observers' high performance and a significant amount of the variance in display-by-display difficulty can be predicted by a model that uses either of these representations. Furthermore, we showed that a model that combines both forms of representation explains nearly all of the variance in change detection performance in patterned displays.

In Experiment 2, we examined the memory representations that underlie standard working memory displays composed of colored squares with no explicit spatial patterns. We again found significant consistency in display-by-display difficulty, suggesting that even in simple displays observers are not treating items independently. In addition, our summary-based encoding model successfully predicted which changes observers found hard or easy to detect.

We thus show that it is necessary to model both more structured memory representations as well as observers encoding strategies to successfully understand what information observers represent in visual working memory. We provide a framework for such modeling Bayesian inference in a model of change detection and show that it can allow us to understand the format of observers memory representations. Interestingly, our models converge with the standard visual working memory literature on an estimate of 3-4 individual objects remembered, even in the patterned displays where simpler formal models massively underestimate observers performance.

3.7.1 Predicting display-by-display difficulty

Because each individual item in a typical working memory display is randomly colored and located at a random spatial position, formal models of working memory have tended to treat the displays themselves as interchangeable. Thus, existing models of visual working memory have focused on average memory performance across many different displays. For example, the standard "slot" model used to calculate K values takes into account only the number of items present and the number of items that change between study and test, ignoring any display-by-display variance in which items are likely to be encoded and how well the items group or how well they can be summarized in ensemble representations. Even modeling efforts that do not focus on slots have tended to examine only performance across all displays (for example, Wilken and Mas (2004) signal detection model where the performance decrement with increasing numbers of items encoded results only from internal noise and noise in the decision process).

However, even when the items themselves are chosen randomly, each display may not itself be random: instead, any given display may contain significant structure.

Furthermore, by focusing on average performance across displays, existing models have necessarily assumed that each individual item is treated independently in visual working memory. In the current work we find that this assumption of independence between items may not hold even in simple displays, but perhaps more importantly, requiring independence between items leaves little room to scale up formal models of working memory to displays where items are clearly not random, as in real-world scenes or even the patterned displays in Experiment 1.

There are two examples of work that fit a formal model which takes into account information about each individual display in working memory, although neither examines model fits for each particular display as we do in the current work. In the first, Bays, Catalao and Husain (2009) showed that taking into account information about particular displays may be critical to distinguishing between slot models and resource models in continuous report tasks (Bays, Catalao & Husain, 2009; Zhang & Luck, 2008). Zhang and Luck (2008) found evidence that observers seem to frequently randomly guess what color an item was, perhaps suggesting a limit on how many items can be encoded. However, Bays, Catalao and Husain (2009), using data taking which takes into account display-by-display differences, have argued that many of these random guesses are actually reports of an incorrect item from the study display. Reports of the incorrect item tend to average out when looking at all displays, but for each individual display make a large difference in how many items we should assume observers' were remembering. Bays et al. (2009) argue that once these trial-by-trial variations are taken into account, the data support a resource model rather than a slot model of working memory (although see Anderson et al. 2011).

The second example of fitting a working memory model to each individual display is work done by Brady, Konkle and Alvarez (2009; Chapter 4) on how statistical learning impacts visual working memory. By creating displays where the items were not randomly chosen (particular colors appear in a pair together more often than chance), they showed that observers can successfully encode more individual colors as they learn regularities in working memory displays. Furthermore, using an information-theoretic model to predict how "compressible" each display was based on how predictable the pairings of colors are, Brady et al. (2009) were able to explain how well observers would remember particular displays. For example, displays that have a large number of highly predictable color pairs were remembered better than displays with less predictable pairs.

In the current work, we introduce the encoding of summary statistics and perceptual grouping as possible factors in observers memory representations. Since the influence of these factors differs on each display, we are able to separately predict the difficulty of each individual visual working memory display. We thus collected data from large numbers of observers performing the same change detection task on exactly the same displays. This allowed us to examine how well our model predicted performance for each individual display for the first time. This display-by-display approach could potentially open up a new avenue of research for understanding the representations used in visual working memory, because it allows clear visualizations of what factors influence memory within single displays.

3.7.2 The use of ensemble statistics for summary-based encoding

In our summary-based encoding model, we suggested that observers might store two distinct kinds of memory representations: a set of individual objects, plus summary statistics which encode an overall gist of the display. We found evidence that such summary-based encoding can explain human change detection in both patterned displays and in simple displays. In addition, we found evidence that a crucial role of summary-based encoding is to guide attention to outlier items.

Our model of summary-based encoding links to both a rich literature on how we encode real-world scenes (e.g., encoding both scene information and specific objects: Hollingworth, 2006; Oliva, 2005) and to an emerging literature on the representation of visual information using ensemble statistics (e.g., encoding mean size of a set of items or the distribution of orientations on a display: Alvarez, 2011; Haberman & Whitney, 2011). When representing a scene observers encode not only specific objects but also semantic information about a scenes category as well as its affordances and other global scene properties (e.g., Greene & Oliva, 2009a, 2009b, 2010b). Observers also represent some scene-based summary statistics that are encoded visually rather than semantically. For example, Alvarez and Oliva (2009) have shown that observers are sensitive to some global patterns of orientation but not others, possibly based on how meaningful such patterns are in the statistics of natural scenes (Oliva & Torralba, 2001). This visual ensemble information seems to be linked to the way we process texture (Haberman & Whitney, 2011).

There is also existing evidence that the representation of such scene and ensemble information influences our encoding of specific objects. In real-world scenes, much of the influence of such gist representations on the representation of individual objects seems to be semantic. For example, observers are better at remembering the spatial position of an object when tested in the context of a scene (Mandler & Johnson, 1976; Hollingworth, 2007), and this effect is stronger when the scene information is meaningful and coherent (Mandler & Johnson, 1976; Mandler & Parker, 1976). In addition, gist representations based on semantic information seem to drive the encoding of outlier objects. Thus objects are more likely to be both fixated and encoded into memory if they are semantically inconsistent with the background scene (e.g., Friedman, 1979; Hollingworth & Henderson, 2000, 2003).

Visual information from scenes also influences our encoding of objects. Thus, observers encoding real-world scenes not only preferentially encode semantic outliers but also visual outliers ("salient" objects) (Wright, 2005; Fine & Minnery, 2009, J Neuro; although see Stirk & Underwood, 2007). In addition, when computing ensemble visual representations in simpler displays observers discount outlier objects from these representations (Haberman & Whitney, 2010), and combine their representations of the ensemble statistics with their representation of individual items (Brady & Alvarez, 2011; Chapter 2).

Taken together, this suggests that observers representations of both real-world scenes and simpler displays consists of not only information about particular objects but also scene-based information and ensemble visual information. Furthermore, this summary information is used to influence the choice of particular objects to encode and ultimately influences the representation of those objects.

In the current work, we formalized a simplified version of such a summary-based encoding model. Rather than representing semantic information, we use displays that lack semantic information and used a summary representation based on Markov Random Fields (Geman & Geman, 1984). This summary representation represents only local spatial continuity properties of the display (e.g., the similarity between items that are horizontal and vertical neighbors). Interestingly, however, a very similar representation seems to capture observers impression of the subjective randomness of an image patch (Schreiber & Griffiths, 2007), a concept similar to Garners (1974) notion of pattern goodness. Pattern goodness is an idea that has been difficult to formalize but qualitatively seems to capture which images are hard and easy to remember (Garner, 1974).

Nevertheless, our summary representation is likely too impoverished to be a fully accurate model of the summaries encoded in human memory, even for such simple displays. For example, if letters or shapes appeared in the dot patterns in our displays, observers would likely recall those patterns well by summarizing them with a gist-like representation. Our model cannot capture such representations. Additional visual summary information is also likely present but not being modeled: for example, if we changed the shape of one of the items in Experiment 1 from a red circle to a red square observers would almost certainly notice despite the large number of individual items on the display (e.g., see Brady, Konkle & Alvarez, 2011; Chapter 1). However, we believe that our model nevertheless represents a step forward in understanding how people make use of such summary information during change detection. Despite the relative simplicity of the summary representation, the model seems to capture a large amount of variance in how well observers remember not only patterned displays but also simple visual working memory displays.

3.7.3 Chunking

In our chunk-based encoding model, we suggested that observers might make use of the Gestalt principle of similarity to form perceptual units out of the individual items in our displays and encode these units into memory as chunks. We found evidence that such chunk-based encoding can explain part of human change detection in patterned displays.

This idea that memory might encode chunks rather than individual objects relates to two existing literatures. One is the literature on semantic, knowledge-based chunk formation. For example, a large amount of work has been done to understand how form chunks based on knowledge, both behaviorally (e.g., Chase & Simon, 1973; Cowan et al. 2004; Brady et al 2009; Gobet et al. 2001) and with computational models of what it means to form such chunks, how all-or-nothing chunk formation is and what learning processes observers undergo (e.g., Brady et al 2009; Gobet et al. 2001). The other literature on chunk formation is based on more low-level visual properties, as examined under the headings of perceptual grouping and pattern goodness (e.g., Wertheimer, 1938; Koffka, 1935; Garner, 1974). In the current work we use non-semantic stimuli and do not repeat stimuli to allow for learning, and thus it is likely we are tapping a form of chunk formation that is based on grouping properties of low-level vision rather than based on high-level knowledge.

Some previous work has focused on how to formalize this kind of perceptual grouping (Kubovy & Van den Berg, 2008; Rosenholtz, Twarog, Schinkel-Bielefeld, & Wattenberg, 2009). For example, Kubovy and Van den Berg (2008) have proposed a probabilistic model of perceptual grouping with additive effects of item similarity and proximity on the likelihood of two objects being seen as a group. In the current experiments, our items differ only in color, and thus we make use of a straightforward model of grouping items into chunks, where items that are adjacent and same-colored are likely but not guaranteed to be grouped into a single unit. This grouping model is similar in spirit to that of Kubovy and Van den Berg (2008), and in our displays seems to explain a significant portion of the variance in observers memory performance. This provides some evidence that perceptual grouping may occur before observers encode items into memory, allowing observers to encode perceptual chunks rather than individual items per se.

Similar models of perceptual grouping have been proposed to explain why observers are better than expected at empty-cell localization tasks using patterned stimuli much like ours (Hollingworth, Hyun & Zhang, 2005) and why some displays are remembered more easily than others in same/different tasks (Howe and Jung, 1986; Halberda et al. submitted) . However, this previous work did not attempt to formalize such a model of perceptual grouping. This is important because in the current experiments we find that summary-based encoding provides another possible explanation for the benefits observed in patterned displays, and in fact may provide a more general solution since it helps explain performance in simpler displays better than perceptual grouping. Thus, we believe it is an important open question the extent to which summary-based encoding rather than perceptual grouping could explain improved performance for patterned displays in previous experiments (Hollingworth, Hyun & Zhang, 2005; Howe & Jung, 1986; Halberda et al., submitted).

3.7.4 Fidelity in visual working memory

In line with the previous literature on working memory, the current modeling effort largely treats working memory capacity as a fixed resource in which up to K items may be encoded with little noise. While expanding on what counts as an item (in the chunk-based model) or suggesting a hierarchical encoding strategy (in the summary-based model), nevertheless we do not investigate in detail the fidelity stored in the representations or the extent to which encoding is all-or-none (e.g., slot-like) versus a more continuous resource.

There are several important caveats to the simplistic idea of all-or-none slots that we use throughout the current modeling effort. The first is that for complex objects, observers are able to represent objects with greater detail when they are encoding only a single object or only a few objects than when they are encoding many such objects (Alvarez & Cavanagh, 2004; Awh et al 2007). In fact, the newest evidence suggests this is true even of memory for color (Zhang & Luck, 2008). For example, Zhang and Luck (2008) find that observers have more noise in their color reports when remembering 3 colors than when remembering only a single color. It has been proposed this is due to either a continuous resource constraint with an upper-bound on the number of objects it may be split between (Alvarez & Cavanagh, 2004), a continuous resource with no upper bound (Bays & Husain, 2008; Bays, Catalao & Husain, 2009), a continuous resource that must be divided up between a fixed number of slots (Awh et al. 2007), or because observers store multiple copies of an object in each of their slots when there are fewer than the maximum number of objects (Zhang & Luck, 2008). In either case, the simplistic model in which several items are perfectly encoded needs to be relaxed to incorporate these data.

Furthermore, in real-world displays which contain many real objects in a scene, observers continually encode more objects from the display the more time they are given (Hollingworth, 2004; Melcher, 2001, 2006). In fact, even on displays with objects that are not in a coherent scene, if those objects are semantically rich real-world objects, observers remember more detailed representations for a larger number of objects as they are given more time to encode the objects (Brady, Konkle, Oliva & Alvarez, 2009; Melcher, 2001).

Despite these complications, in the current modeling we focus on expanding a basic all-or-none slot model to the case of dealing with higher-order regularities and perceptual organization. We use such a model as our basic architecture of working memory because of its inherent simplicity and because it provides a reasonable fit to the kind of change detection task where the items to be remembered are simple and the changes made in the change detection task are large, as in the current studies (e.g., categorical changes in color, Luck & Vogel, 1997). Future work will be required to explore how perceptual grouping and summary-based encoding interact with memory fidelity.

3.7.5 Conclusion

Memory representations of real-world scenes are complex and structured: observers encode both scene-based semantic and visual information as well as specific objects, and the objects they encode are chosen based on the scene-based information. By contrast, formal models of working memory have typically dealt with only simple memory representations that assume items are treated independently and no summary information is encoded.

In the current work we presented a formal model of change detection that uses Bayesian inference to make predictions about visual working memory architecture and capacity. This model allowed us to take into account the presence of summary information and perceptual organization, while making quantitative predictions about the difficulty of particular working memory displays. We found evidence that observers make use of more structured memory representations not only in displays that explicitly contain patterns, but also in randomly-generated displays typically used in working memory experiments. Furthermore, we provided a framework to model these structured representations Bayesian inference in a model of change detection and showed that it can allow us to understand how observers make use of both summary information and perceptual grouping.

By treating change detection as inference in a generative model, we make contact with the rich literature on a Bayesian view of low-level vision (Knill & Richards, 1996; Yuille & Kersten, 2006) and higher-level cognition (e.g., Griffiths & Tenenbaum, 2006; Tenenbaum, Griffiths, & Kemp, 2006). Furthermore, by using probabilistic models we obtain the ability to use more complex and structured knowledge in our memory encoding model, rather than treating each item as an independent unit (e.g., Kemp & Tenenbaum, 2008; Tenenbaum, Griffiths & Kemp, 2006). Our model is thus extensible in ways that show promise for building a more complete model of visual working memory: within the same Bayesian framework, it is possible to integrate existing models of low-level visual factors with existing models of higher-level conceptual information (e.g., Kemp & Tenenbaum, 2008), both of which will be necessary to ultimately predict performance in working memory tasks with real-world scenes.

3.8 Chapter Appendix

3.8.1 Basic Change Detection Model Details

Walkthrough

We wish to model a simple change detection task of the form frequently used in visual working memory (see Figure 3-11). Observers are presented with a display (the study display) consisting of 8 colored squares. The display then disappears, and a short time later another display reappears (the test display) that is either identical to the study display or in which a single square has changed color. Observers must decide whether the test display is identical to the study display or there has been a change. Observers are told that at most a single item will change color.

To create a model of this change detection task, we first specify how observers' encode the study display. For the simplest case, in order to most closely match the models used to quantify "K-values" in standard displays (Cowan, 2001), we assume that memory takes the form of a discrete number of slots, K, each of which stores which color was present on the display in a particular location (using seven colors: black, white, red, green, blue, yellow and purple). Also in line with standard slot models, we initially assume that observers choose which K of the N dots to encode at random. To model the change detection task, we then formalize how observers make a decision about whether there was a change when the test display is presented. When observers must decide if there has been a change, observers have access to the test display and to the items they encoded in memory from the study display. Using the information that at most a single item can change color, we model the observer as performing an optimal inference to arrive at a judgment for whether the display has changed.

For this simple model, this inference is straightforward. When an observer is looking at a particular test display, there are 49 (1 + 8*6) possibilities for what might have been the study display: There is the possibility that the study display was exactly the same as the test display (1 +), plus the possiblility the study display had any one of the eight items as a different color then in the test display (since there are 6 other colors for each of the 8 items, this gives 8*6). Assuming 50% of trials are 'change' and 50% 'no-change', and thus observers use 50% as the prior probability of a change, this means that after observing the test display, observers start with the belief that there is a 50% chance that the study display was the same as the test display, and a 1/(8*6), or 1.04% chance that the study display was any particular one of the possible changed displays.

To arrive at the final inference, this information must then be updated based on the K colors encoded in memory from the study display. Using the colors of these K items, observers can rule out (e.g., assign 0 likelihood) any of the hypothesized study displays that have a color that differs from their memory. For example, if an observer remembers a particular item was blue in the study display, and it remains blue in the test display, this observer can rule out all 6 possible changes in color for that item, and thus reduce the possible changed displays from 48 to 42.

After ruling out the displays that are incompatible with their memory representation, observers can then calculate the final posterior probability of a change: this is the percentage of the remaining probability that is part of the 8*6 "change" displays as opposed to the 1 "no-change" display. Note that if one of the remembered items differs from the second display, this will rule out 'no change' and observers will be sure there was a change (all remaining probability will be on the "change" displays). If none of the remembered items differ from the test display, then observers will be more likely to say no change, and how likely they will be to say no change will depend on how many of the possible study displays can be ruled out based on the items they remember. The larger the value of K, the more possible study displays will be ruled out and the more sure observers will become that there was no change (see Figure 3-11).

If we cue which item to check for a change (e.g., by putting only a single square up when we show the test display and asking whether only this item is the same

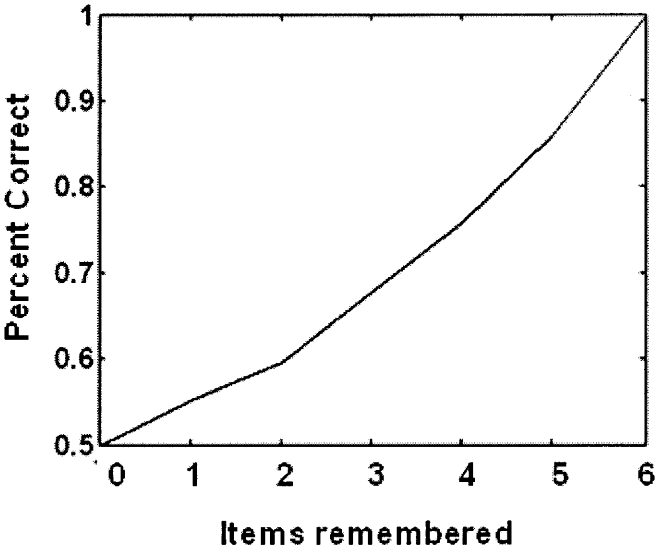


Figure 3-11: Results of probabilistic change detection model on detecting a single change in a display with 8 colored squares as a function of the number of items remembered. As more items are encoded, the model gets a larger percentage of displays correct.

or different), then our Bayesian change detection model becomes even simpler. In the full Bayesian change detection model, if we do not encode an item and that item changes, the likelihood we say 'change' is a function of our capacity K. As K increases, we get greater implicit evidence for 'no-change', since the more items we encode the less likely it is that we would have failed to encode the item that changed.

However, if we are cued to which item may have changed, then we need only weigh the probability the color is the same as the first display (50% prior) against the probability it used to be one of the 6 other colors (50% * 1/6 = 8.3% chance of each other color). If we encoded the item in memory, which happens on K/N out of each N trials because we randomly sample which items to encode, then we get the trial correct. If we don't encode the item in memory, then there is a 50% chance it is same and 50% chance it is different, given our prior probability about whether there was a change (regardless of how many items we encoded). Thus, the chance of getting a trial correct reduces to:

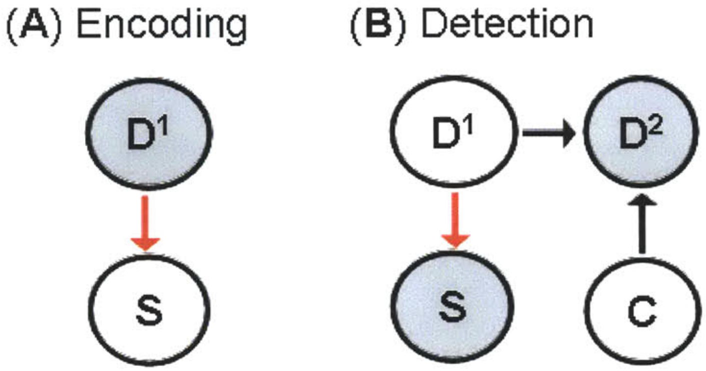


Figure 3-12: Graphical model notation for the basic multinomial change-detection model at (A) encoding and (B) detection. Shaded nodes are observed. The red arrows correspond to observers' memory encoding strategy; the black arrows correspond to constraints of the task (e.g., at most 1 dot will change between display 1 (D^1) and display 2 (D^2)).

$$P.C.(\%) = \frac{K}{N} * 1 + \frac{N-K}{N} * 0.50$$

This is the same as the formula used to calculate Cowan's K (Cowan, 2001). Thus, the Bayesian model of change detection has Cowan's K as a special case where the item that may have changed is cued.

Formalization

Formally, in this model with no higher-order information, we treat each of the N items on the study display as a multinomial random variable D_i^1 (i from 1, 2,... N), where the set of possible values of each D_i^1 is 1, 2, ..., 7, representing the seven possible colors. We choose whether to remember each object from the display by sampling K specific objects from the display without replacement. S denotes the set of K specific objects encoded: $S = S_1, ..., S_K$. For each item encoded in S, we store a multinomial distribution with 100% of the mass on the color of that item in the display (D_i^1). Nothing is stored about all other items.

At detection, observers have access to the information encoded in S and also the items on the test display (D^2). This test display is generated by taking the study display (D^1) and either modifying no colors (C=0, prior probability 0.5) or modifying a single color at random (C=1, prior probability 0.5). Observers must infer, using S and D^2, whether a change was made (the value of C). To do so, they must use S and D^2 to put a probability distribution on all possible values of D^1 (e.g., calculate $p(D^1|D^2, S)$. In other words, to know if there was a change, observers must know what the second display and the information in memory suggests about the first display. In this case, $p(D^1|D^2)$ is independent of $p(D^1|S)$, and so $p(D^1|D^2, S)$ can be calculated as simply $p(D^1|D^2) * p(D^1|S)$.

Because of the constraint that at most one item will change between D^1 and D^2, the value of D^2 rules out all but a small number of possible displays for consideration as the possible study display, D^1. In other words, $p(D^1|D^2)$ puts non-zero probability on only 1 + N*6 of the N^6 possible study displays. These correspond to the study display that is exactly the same as the test display (1 +), plus the study displays that correspond to any one of the N items changing from the color they are in the test display to one of the 6 other colors (N*6).

The value of S rules out a number of displays proportional to the number of items encoded in S (K items). $p(D^1|S)$ assigns zero probability to all displays where an item D^1_i is a different color than S_i, for all K items encoded in S. All other displays are given equal likelihood.

Taken together with an equal prior probability of all possible displays D^1, these two distributions, $p(D^1|D^2)$ and $p(D^1|S)$, provide the posterior function for $p(D^1)$. As the final step in the inference, this posterior over possible first displays must be converted to a posterior on whether there was a change in the display (C). This is based on the prior probability of a change (0.5) and whether D^1 is equal to D^2 (e.g., whether there is a change between the two displays). Thus:

$$p(C = 1|D^2, S) = 0.5 * \frac{p(D^1 = D^2|D^2, S)}{0.5p(D^1 = D^2|D^2, S) + 0.5p(D^1 \neq D^2|D^2, S)}$$

3.8.2 Patterned Display Generation

To generate the patterned displays used in Experiment 1A and 1B, we sampled a set of 16 displays from a Markov Random Field (MRF) smoothness model like those used in our summary-based encoding model and our chunking model (see Chapter Appendix, Sections 3.8.3 and 3.8.4). Using an MRF with separate parameters for horizontal and vertical smoothness, we used Gibbs sampling to generate a set of four displays from each of 4 possible parameter settings. These parameters encompassed a wide range of possible patterns, with horizontal/vertical smoothness set to all combinations of +/- 1 [e.g., (1, 1), (-1, -1), (1, -1), (1, -1)]. This gave us 16 displays with noticeable spatial patterns.

In addition, we generated 8 displays by randomly and independently choosing each dots color (50%/50%). In Experiment 1A, these 24 displays consisted of red and blue dots. In Experiment 1B they were exactly the same displays, but composed of black and white squares instead.

3.8.3 Summary-Based Encoding Model Details

Encoding

The graphical model representation of the encoding model (shown in Figure A3) specifies how the stimuli are initially encoded into memory. We observe the study display (D1), and we use this to both infer the higher-order structure that may have generated this display (G) and to choose the specific set of K items to remember from this display (S).

In the model, any given summary representation must specify which displays are probable and which are improbable under that summary. Unfortunately, even in simple displays like ours with only 2 color choices and 25 dots, there are 2^{25} possible displays. This makes creating a set of possible summary representations by hand and specifying the likelihood each summary gives to each of the 2^{25} displays infeasible. Thus, as a simplifying assumption we chose to define the summary representation using Markov Random Fields, which allow us to specify a probability distribution

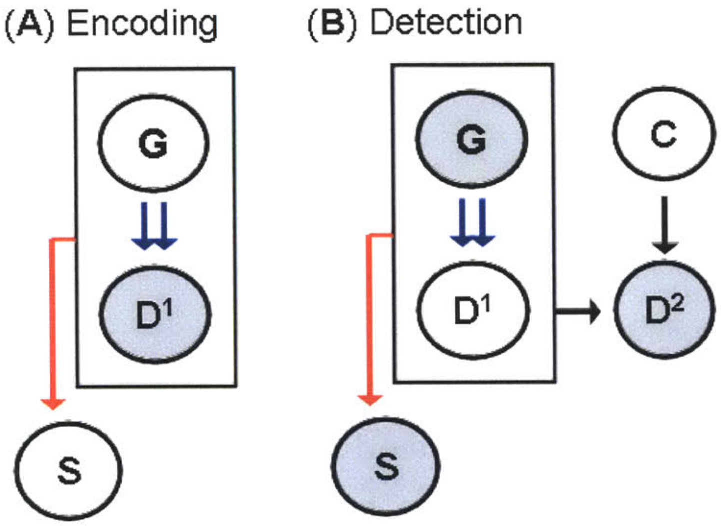


Figure 3-13: Graphical model notation for the summary-based encoding model at (A) encoding and (B) detection. Shaded nodes are observed. The red arrows correspond to observers' memory encoding strategy; the black arrows correspond to constraints of the task (e.g., at most 1 dot will change between the study display (D^1) and test display (D^2)). The blue arrows correspond to our model of how a display is generated; in this case, how the summary or gist of a display relates to the particular items in that display.

over all images by simply defining a small number of parameters about how items tend to differ from their immediate neighbors. Such models have been used extensively in computer vision (Geman & Geman, 1984; Li, 1995). We use only two summary parameters, which specify how often items are the same or different color than their horizontal neighbors (G_h) and how often items are the same or different color than their vertical neighbors (G_v). Thus, one particular summary representation ($G_h = 1, G_v = -1$) might specify that horizontal neighbors tend to be alike but vertical neighbors tend to differ (e.g., the display looks like it has horizontal stripes in it). This summary representation would give high likelihood to displays that have many similar horizontal neighbors and few similar vertical neighbors.

We treat each item in these change detection displays as a random variable D^1_i, where the set of possible values of each D^1_i is -1 (color 1) or 1 (color 2). To define the distribution over possible displays given the gist parameters, $P(D|G)$, we assume that the color of each dot is independent of the color of all other dots when conditioned on its immediate horizontal and vertical neighbors.

We thus have two different kind of neighborhood relations (clique potentials) in our model. One two parameters (G_h and G_v) apply only to cliques of horizontal and vertical neighbors in the lattice (N_h and N_v) respectively. Thus, $P(D^1|G)$ is defined as:

$$P(D^1|G) = \frac{exp\left(-En(D^1|G)\right)}{Z(G)} \tag{3.1}$$

$$En(D^1|G) = G_v \sum_{(i,j)\in N_v} \psi(D^1_i, D^1_j) + G_h \sum_{(i,j)\in N_h} \psi(D^1_i, D^1_j)$$

where the partition function:

$$Z(G) = \sum_{D^1} exp\left(-E(D^1|G)\right)$$

normalizes the distribution. $\psi(D^1_i, D^1_j)$ is 1 if $D^1_i = D^1_j$ and -1 otherwise. If $G > 0$ the distribution will favor displays where neighbors tend to be similar colors, and if

$G < 0$ the distribution will favor displays where neighbors tend to be different colors.

The summary of the display is therefore represented by the parameters G of an MRF defined over the display. Our definition of $p(D^1|G))$ thus defines the probability distribution $p(display|summary)$. To complete the encoding model we also need to define $p(items|display, summary)$ $(p(S|D^1, G))$. To do so, we define a probability distribution that preferentially encodes outlier objects (objects that do not fit well with the summary representation).

We choose whether to remember each object from the display by looking independently at the conditional probability of that object under the summary representation, assuming all of its neighbors are fixed $p(D^1_i|G, D^1_{/i})$. S denotes the set of K specific objects encoded: $S = s_1, ..., s_k$. To choose S, we rank all possible sets of objects of size 0, 1, 2, ... to K objects based on how unlikely they are under the encoded summary representation. Thus, the probability of encoding a set of objects (S) is:

$$p(S|G, D^1) = \prod_{j:s_j \in S} [1 - p(D^1_j|G, D^1_{/j})] \prod_{j:s_j \notin S} p(D^1_j|G, D^1_{/j}) \tag{3.2}$$

This defines $p(S|D^1, G)$, which provides the probability of encoding a particular set of specific items in a given display, $p(items|display, summary)$, in our model.

To compute the model predictions we use exact inference. However, due to the computational difficulty of inferring the entire posterior distribution on MRF parameters for a given display (e.g., the difficulty of computing $Z(G)$), and because we do not wish to reduce our summary representation to a single point estimate, we do not compute either the maximum posterior MRF parameters for a given display or the full posterior on G. Instead, we store the posterior in a grid of values for G in both horizontal and vertical directions $(G_h = -1.5, -1, -.5, 0, .5, 1, 1.5, G_v = -1.5, -1, -.5, 0, .5, 1, 1.5)$. We compute the likelihood of the display under each of these combinations of G_h and G_v and then choose the items to store (S) by integrating over the different choices of G (we store the full posterior over S)). We choose a uniform prior on the summary representation (e.g., a uniform prior on MRF parameters G).

In summary, to encode a display we first treat the display as an MRF. We then calculate the posterior on possible summary representations by calculating a posterior on G at various (pre-specified) values of G. We then use this G and the original display to compute a posterior on which set of $\leq K$ items to encode into item memory (S). At the completion of encoding we have both a distribution on summary representations (G) and a distribution on items to remember (S), and these are the values we maintain in memory for the detection stage.

Detection

At the detection stage, we need to infer the probability of a change to the display. To do so, we attempt to recover the first display using only the information we have in memory and the information available in the second display. Thus, using the probabilistic model, we work backwards through the encoding process, so that, for example, all the possible first displays that don't match the specific items we remembered are ruled out because we would not have encoded a dot as red if it were in fact blue.

More generally, to do this inference we must specify $P(D^1|S)$, $P(D^1|D^2)$, $P(D^1|X)$, $P(S|G, D^1)$. Almost all of these probabilities are calculated by simply inverting the model we use for encoding the display into memory initially with a uniform prior on possible first displays. Thus, $P(D^1|G)$ and $P(S|G, D^1)$ are given by the same equations described in the Encoding section.

Those probabilities not specified in the forward model represent aspects of the change detection task. Thus, $P(D^1|S)$ is a uniform distribution over first displays that are consistent with the items in memory and 0 for displays where one of those items differs. This represents our simplifying assumption (common to standard "slot" models of visual working memory) that items in memory are stored without noise and are never forgotten (it is possible to add noise to these memory representations by making $P(D^1|S)$ a multinomial distribution over possible values of each item, but for simplicity we do not model such noise here). $P(D^1|D^2)$ is uniform distribution over all displays D^1 such that either $D^1 = D^2$ or at most one dot differs between D^1 and

D^2. This represents the fact that the task instructions indicate at most one dot will change color.

Together these distributions specify the probability of a particular first display given the information we have about the second display and information we have in memory, $P(D^1|G, S, D^2)$. Given the one-to-one correspondence between first displays and possible changes, we can convert this distribution over first displays to a distribution over possible changes. Our prior on whether or not there is a change is 0.5, such that 50% of the mass is assigned to the "no change" display and the other 50% is split among all possible single changes. Thus:

$$P(C|G, S, D^2) = \frac{0.5P(D^1 = D^2|G, S, D^2)}{0.5P(D^1 = D^2|G, S, D^2) + 0.5\sum P(D^1 \neq D^2|G, S, D^2)}$$

This fully specifies the model of change detection.

Model with no summary information at time of detection

Is remembering the summary representation helping us to accurately model human performance, or can we predict human performance equally well by using the summary to choose outliers to encode into memory but then discarding the summary representation itself? To examine this, we looked at the fit of a model that did not have access to the summary representation at the time of change detection, and detected changes solely based on the specific objects encoded.

Formally, this model was identical to the model described, but without conditioning on G when doing change detection. Thus, detection was based only on the probabilities $P(D^1|S)$ and $P(D^1|D^2)$, which are once again calculated by using the same equations as used in the encoding model.

Model with objects chosen at random

It is also possible to examine a model that encodes both a summary of the display and specific items, but does not choose which items to specifically encode by selecting outliers from the summary. Rather than preferentially encoding unlikely items, such a model chooses the items to encode at random.

Formally, we use S to denote the set of K specific objects encoded: $S = s_1, ..., s_k$. In the full model, it is calculated by choosing objects that are outlier with respect to G:

$$p(S|G, D^1) = \prod_{j:s_j \in S} [1 - p(D_j^1|G, D_{/j}^1)] \prod_{j:s_j \notin S} p(D_j^1|G, D_{/j}^1) \tag{3.3}$$

To lesion the model and encode objects at random, we instead choose the set S of objects to encode at random. Thus, to choose S, we no longer consider all possible sets of objects up to size K based on how unlikely they are, but simply sample K of the N objects in the display.

Model as applied to random dot displays

To apply the model to the displays from Exp. 2, we use the same model and treat the items that are adjacent in the grid as neighbors. Blank spots on the display are ignored, such that the MRF is calculated only over pair of items (cliques, N_v and N_h) that do not contain a blank location.

To do inference in this model, we can no longer use exact inference, since calculating the partition function $Z(G)$ for these displays is computationally implausible. Instead, to calculate the likelihood of a given display under a particular summary representation, we use the pseudolikelihood, which is the product, for all of the items, of the conditional probability of that item given its neighbors (Li, 1995; Besag, 1975, 1977). Thus, $P(D^1|G)$ is calculated as:

$$p(D^1|G) = \prod_i \frac{exp(-En(D_i^1|G))}{exp(-En_i(0|G)) + exp(-En_i(1|G))} \tag{3.4}$$

$$En_i(D^1|G) = G_v \sum_{j \in N_v(i)} \psi(D_i^1, D_j^1) + G_h \sum_{j \in N_h(i)} \psi(D_i^1, D_j^1) \tag{3.5}$$

Such an estimate of the likelihood is computationally straightforward, and in MRFs has been shown to be a reasonable approximation to the true underlying likelihood function (Besag, 1977). We can calculate how good an approximation it is for our particular change detection model by examining how closely predictions using the pseudolikelihood approximate the exact likelihood computations in Experiment 1. In that model (with K=4), the change detection estimates (how likely each test display is to be the same as the study display) correlate r=0.98 between the model that uses exact inference and the model that relies on the pseudolikelihood to estimate the likelihood. This suggests the pseudolikelihood provides a close approximation of the true likelihood in our displays.

3.8.4 Chunking Model Details

To model chunk-based encoding, we add two components to our basic change detection model. First, rather than encoding K single objects, we encode up to K chunks from a display (S now encodes chunks rather than individual items). Second, to select these chunks we use two factors, corresponding to the Gestalt principles of proximity and similarity: (1) a spatial smoothness term that encourages the model to put only adjacent items into the same chunk; (2) a likelihood term that forces the model to put only items of the same color into the same chunk.

We probabilistically segment the display into chunks, and then select which K of these chunk to encode into our chunk memory, S, by preferentially encoding larger chunks (where chance of encoding is proportional to chunk size; e.g., we are twice as likely to encode a chunk of 4 dots as a chunk of 2 dots). This allows us to examine how likely an observer that encoded a display in this way would be to detect particular changes for different values of K (see Figure 5 in the main text for a sample of possible chunk-segmentations for a particular display).

Our formalization of the chunk-based model has three stages. First, we compute a distribution over all possible ways of segmenting the study display into chunks, R. Then, for each value of R, we compute a distribution over all possible ways of choosing K chunks from R to encode into our chunk memory, S. Finally, we calculate how likely the display is to be the same for each possible value of R and each possible value of S given this R. Due to the huge number of possible values of R, we use Gibbs sampling to sample possible segmentations rather than doing a full enumeration. For any given segmentation R, however, we do a full enumeration of assignments of S and thus likelihoods of the display being the same or different.

To compute a distribution over R, we treat the chunk-assignment of each item D_i^1 as a random variable R_i. Thus, R_i corresponds to which region D_i^1 is considered a part of, and each R_i can take on any value from 1...25 (the total number of items present in the display, and thus the maximum number of separate regions).We then compute a distribution over possible assignments of R_i using a prior that encourages smoothness (such that items D_i^1 that are either horizontal or vertical neighbors are likely to have the same region assignment), and using a likelihood function that is all-or-none, simply assigning 0 likelihood to any value of R where two items assigned the same chunk differ in color (e.g., likelihood zero to any R where $R_i{=}R_j$, $D_i^1 \neq D_j^1$) and uniform likelihood to all other assignments of R.

We sample from R using Gibbs Sampling. We thus start with a random assignment of values for each R_i, and then sample each R_i repeatedly from the distribution $p(R_i|R_{\sim i}, D^1)$ to generate samples from the distribution

$$P(R_i|R_{\sim i}) \propto exp(-En(R_i|Sm)) \tag{3.6}$$

$$En(R_i|Sm) = Sm \sum_{i,j \in N} \psi(R_i, R_j) \tag{3.7}$$

Where, again, $\psi(R_i, R_j) = 1$ if $R_i = R_j$ and -1 otherwise.

For values of Sm >>0, we prefer larger chunks to smaller chunks, since we more strongly prefer neighboring items to have the same chunk-label. As discussed in the main text, the model is relatively insensitive to the value of this parameter for values >= 1.0. For all simulations, we set this value to 4.0 because this provided a model that created different segmentations of the display fairly often, while still making those segmentations consist of relatively larger chunks.

The likelihood function, $P(R|D^1)$, is simply defined such that all chunks/regions must have only a single color within them. Thus, if for any $R_i = R_j$, $D_i^1 \neq D_j^1$, then $P(R|D^1) = 0$, otherwise $P(R|D^1) \propto 1$. Taken together, this likelihood and the MRF smoothness prior specify the distribution over R.

To compute a distribution over S for a given value of R, we enumerate how many unique chunk assignments are present in R (total number of chunks, M), labeling each of these chunks $L = 1, 2...M$. We then choose K chunks from this set of M possible chunks for our chunk memory, S, by choosing without replacement and giving each chunk label a chance of being chosen equal to the % of the items in the display that belong to that chunk. Thus:

$$P(S|R, D^1) \propto \prod_{i:s_i \in S} \frac{\sum_{j=1...25}(R_j = L_i)}{25} \prod_{i:s_i \ni S} (1 - \frac{\sum_{j=1...25}(R_j = L_i)}{25}) \tag{3.8}$$

To calculate the chance of the display being the same given a value of R and S, we use the following logic (similar to Pashler, 1988). The set of items encoded is all the items assigned to any chunk that is encoded. Thus if $D_i^1 \neq D_i^2$, and i is part of a chunk encoded in S, we notice the change 100% of the time. If no such change is detected, we get more and more likely to say 'same' in proportion to how many items we have encoded from the total set of items: thus, probability of the display having changed is:

$$P(C|R, S) = 1 - (0.5 + 0.5 * \frac{\sum_{j=1...25}(R_j \in S)}{25}) \tag{3.9}$$

Chapter 4

Compression in visual working memory: Using statistical regularities to form more efficient memory representations[1]

The information we can hold in working memory is quite limited, but this capacity has typically been studied using simple objects or letter strings with no associations between them. However, in the real world there are strong associations and regularities in the input. In an information theoretic sense, regularities introduce redundancies that make the input more compressible. In this chapter we show that observers can take advantage of these redundancies, enabling them to remember more items in working memory. In two experiments, we introduced covariance between colors in a display so that over trials some color pairs were more likely than other color pairs. Observers remembered more items from these displays than when the colors were paired randomly. The improved memory performance cannot be explained by simply guessing the high probability color pair, suggesting that observers formed more

[1] Parts of this chapter were published as Brady, T.F, Konkle, T. & Alvarez, G.A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General.*

efficient representations to remember more items. Further, as observers learned the regularities their working memory performance improved in a way that is quantitatively predicted by a Bayesian learning model and optimal encoding scheme. We therefore suggest that the underlying capacity of their working memory is unchanged, but the information they have to remember can be encoded in a more compressed fashion.

4.1 Introduction

Every moment, a large amount of information from the world is transmitted to the brain through the eyes, ears, and other sensory modalities. A great deal of research has examined how the perceptual and cognitive system handles this overwhelming influx of information (Neisser, 1967). Indeed, this information overload is the motivating intuition for why we need selective attention: to actively filter out irrelevant input to allow specific processing of the intended stimuli (Broadbent, 1958). However, since the world is filled with regularities and structure, the information transmitted to the brain is also filled with regularities (Barlow, 1989). In quantitative terms, there is significant redundancy in the input (Huffman, 1952; Shannon, 1948). An intuitive example of the redundancy in the visual input is to consider all the possible images that could be made from an 8 x 8 grid where any pixel can be any color. Most of the images will look like noise, and only a very tiny percentage of these images will actually look like a picture of the real-world (Chandler & Field, 2007). This indicates that real-world images are not randomly structured, and in fact share many structural similarities with each other (e.g., Burton and Moorehead, 1987; Field, 1987; Frazor and Geisler, 2006). Interestingly, computationally efficient representations of image-level redundancy produce basis sets that look remarkably like primary visual cortex, providing evidence that our visual perceptual system takes advantage of this redundancy by tuning neural response characteristics to the natural statistics of the world (Olshausen & Field, 1996).

Being sensitive to the statistics of the input has direct consequences for memory as well as for perception (Anderson & Schooler, 2000). Recent work on the rational analysis of memory, for example, suggests that the power laws of forgetting and practice approximate an optimal Bayesian solution to the problem of memory retrieval given the statistics of the environment (Anderson & Schooler, 1991; see also Shiffrin & Steyvers, 1997; Shiffrin, & Steyvers, 1998). Here we apply similar principles of rational analysis (Chater & Oaksford, 1999) to the capacity of the working memory system. We focus on the abstract computational problem being solved by the working memory system: the storage of as much information as possible in the limited space available.

4.1.1 Working memory capacity and redundancy

According to information theory, in an optimal system more content can be stored if there are redundancies in the input (Cover & Thomas, 1991). In other words, if the input contains statistical structure and regularities, then each piece of information we encode limits the likely possibilities for the remaining information (e.g. given a 'q', the next letter is likely to be 'u'). This makes it possible to encode more items in less space. If the human working memory system approximates an optimal memory system, it should be able to take advantage of statistical regularities in the input in order to encode more items into working memory.

However, while the capacity of short-term and working memory has been extensively studied (e.g., Alvarez & Cavanagh, 2004; Baddeley, 1986; Cowan, 2001, 2005; Zhang & Luck, 2008), little formal modeling has been done to examine the effects of redundancy on the system. Nearly all studies on visual working memory have focused on memory for arbitrary pairings or novel stimuli. While some studies have investigated the effects of associative learning on visual working memory capacity (Olson & Jiang, 2004; Olson, Jiang & Moore, 2005), they have not provided clear evidence for the use of redundancy to increase capacity. 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).

4.1.2 Chunking

However, the effects of redundancy on working memory capacity have been well studied through the phenomenon of "chunking", particularly in verbal working memory (Cowan, 2001; Miller, 1956; Simon, 1974). Cowan (2001) defines a chunk as a group of items where the intra-chunk associations are greater than the inter-chunk associations. In other words, in the sequence FBICIA the letters F, B, and I are highly associated with each other and the letters C, I, and A are highly associated with each other, but the letters have fewer associations across the chunk boundaries. Thus, observers are able to recall the sequence using the chunks 'FBI' and 'CIA', effectively taking up only 2 of the 4 'chunks' that people are able to store in memory (Cowan, 2001; Cowan, Chen, Rouder, 2004). By comparison, when the letters are random, say HSGABJ, they are more difficult to remember, since it is more difficult to chunk them into coherent, associated units.

Chunking is not usually framed as a form of compression analogous to information theoretic views. In fact, in the seminal work of Miller (1956), chunking and information theoretic views of memory were explicitly contrasted, and the most nave information theoretic view was found lacking in its ability to explain the capacity of working memory. However, at its root chunking approximates a form of compression: it replaces highly correlated items (which are therefore highly redundant with each other), with a single chunk that represents all of the items. Thus, it is possible to frame the strategy of chunking as a psychological implementation of a broader computational idea: removal of redundancy to form compressed representations and allow more items to be stored in memory. At this level of description, chunking is compatible with information theoretic analyses. In fact, information theory and Bayesian probability theory may be able to explain exactly when human observers will form a chunk in long-term memory (e.g., Orban, Fiser, Aslin & Lengyel, 2008), in addition to how useful that chunk will be to subsequent working memory tasks. Thus, information theory may be not only compatible with chunking, but in fact may provide useful constraints on theories of chunking.

In the present experiments we asked whether human observers learn and use regularities in working memory in a way that is compatible with an information theoretic compression analysis. In two experiments we present observers with displays of colors that are either random or patterned. By presenting regularities in the display over the course of the experiment, we examine if and how observers take advantage of these regularities to form more efficient representations. We then present a quantitative model of how learning occurs and how the stimuli are encoded using the learned regularities. We show that more items can be successfully stored in visual working memory if there are redundancies (patterns) in the input. We also show that this learning is compatible with the compressibility of the displays according to information theory.

4.2 Experiment 1: Regularities Within Objects

In classic visual working memory experiments, the stimuli used are generally colored oriented lines, shapes, and circles with colors, and the aim is to quantify how many objects or features can be remembered. In one of the seminal papers in this field, Luck and Vogel (1997) proposed that people can remember four objects no matter how many features they contain. This view has since been tempered, with some arguing for independent storage of different feature dimensions (Magnussen, Greenlee & Thomas, 1996; Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002) and others arguing for more graded representations, in which information load determines how many objects can be stored (Alvarez & Cavanagh, 2004; Bays & Husain, 2008). However, nearly all current work emphasizes that at best 3 or 4 features from a given stimulus dimension can be encoded successfully.

Here we modify the standard paradigm by introducing regularities in the displays for some observers. One group of participants was presented with colors drawn randomly, as in classical visual working memory tasks, such that all possible pairs of colors were equally likely to occur. A second group of participants were presented with colors that occurred most often paired with another color. For example, a par-

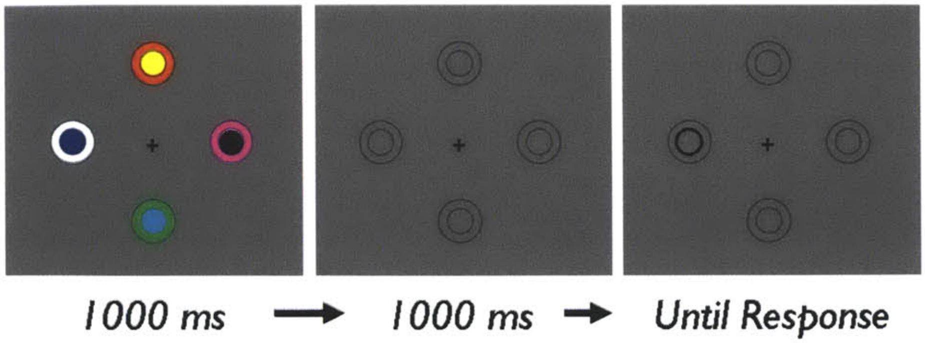


Figure 4-1: A sample trial from Experiment 1. Eight colors were presented within four objects. The colors disappeared for one second and then either the inside or outside of an object was cued by making it darker. Observers had to indicate what color was at the cued location.

ticular observer might see red most often around yellow, white most often around blue, while a smaller percentage of the time these colors appear with any other color. Because this manipulation introduces redundancy into the displays, in information-theoretic terms these displays contain less information. An information theoretic view of memory therefore predicts that the observers presented with regularities should be able to encode more items into memory.

4.2.1 Method

Observers

Twenty naive observers were recruited from the MIT participant pool (age range 18-35) and received 10 dollars for their participation. All observers gave informed consent.

Procedure

Observers were presented with displays consisting of four objects around the fixation point (see sample display in Figure 4-1). Each object was made up of two different colored circles, with one circle inside the other. Observers were informed that their task was to remember the locations of each of the eight colors. At the start of a trial, the colors appeared and remained visible for 1000ms. Then the colors disappeared, with placeholder circles present for the next 1000ms (long enough to prevent observers from relying on iconic memory; Sperling, 1960), and then either the inside or outside circle on a random object was darkened.

The task was to indicate which of the eight colors had been presented at the indicated location, by pressing one of eight color-coded keys. Observers completed 600 trials, presented in 10 blocks of 60 trials each. Afterward, they completed a questionnaire, reporting the strategies they employed and whether they noticed the presence of patterns in the displays.

The stimuli were presented using MATLAB with the Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997). The eight colors used were red, green, blue, magenta, cyan, yellow, black and white.

Manipulation

Observers were randomly assigned to two groups, patterned and uniform, which differed in how the colors for each trial were chosen. For observers in the uniform condition, the locations of the colors in each trial were chosen randomly, with only the constraint that each color had to appear exactly once in a display.

For observers in the patterned condition, the stimuli for each trial were not chosen randomly. First, for each subject a joint probability matrix was constructed to indicate how likely each color was to appear inside and outside of each other color. This matrix was made by choosing four high probability pairs at random (probability = 0.2151), and then assigning the rest of the probability mass uniformly (probability = 0.0027). As in the uniform condition, all eight colors were present in each display. In order to achieve this, the diagonal of the joint probability matrix was set to zero in order to prevent the same color from appearing twice in the same display.

The pairs were constrained so that each color was assigned to exactly one high probability pair. For example, if (Blue-outside, Red-inside) was a high probability pair in this joint probability matrix, the observer would often see blue and red appear together, in that configuration. However, blue and red each would also sometimes appear with other colors, or in a different configuration. So, for example, (Blue-outside, Yellow inside) and (Red-outside, Blue-inside) could also appear with low probability. High probability pairs accounted for approximately 80% of the pairs shown during the experiment, and low probability pairs constituted the other 20%.

In the final block of the experiment in the patterned condition, the distribution from which the displays were drawn was changed to a uniform distribution. This eliminated the regularities in the display, and allowed us to assess whether observers had used the regularities to improve their performance. Further, this manipulation gives a quantitative measure of learning: the difference in performance between block 9 and block 10.

4.2.2 Results

We estimated the number of colors observers could successfully hold in memory using the following formula for capacity given an eight-alternative forced choice (see the Appendix for a derivation of this formula):

$$K = ((PC * 8 * 8) - 8)/7$$

By correcting for chance we can examine exactly how many colors from each display observers would have had to remember in order to achieve a given percent correct (PC). It should be noted that K is a way of quantifying the number of colors remembered that does not necessarily reflect what observers actually represent about the displays. For instance, observers may have all 8 colors with uncertainty rather than some subset of the colors with perfect certainty (see, for example, Wilken & Ma, 2004; Bays & Husain, 2008; however, see Rouder et al. 2008, Zhang & Luck, 2008, for evidence of discrete fixed-resolution representations).

Performance Across Groups

Observers in the uniform condition remembered 2.7 colors on average throughout the experiment (see Figure 4-2). This is consistent with previous results on the capacity

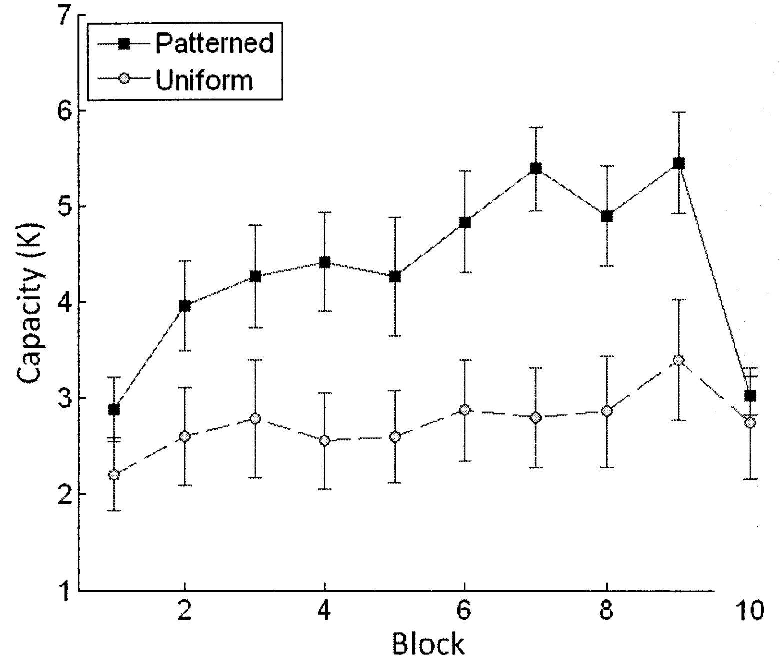


Figure 4-2: Results of Experiment 1. Error bars correspond to $+/-1$ s.e.m.

of visual working memory for colors (e.g. Vogel & Awh, 2008, in which the K values varied from less than 1 to more than 6 across 170 individuals, M = 2.9, SD = 1).

Critically, we found that observers in the patterned condition could successfully remember K = 5.4 colors after learning the regularities in the displays (block 9). This memory capacity is significantly higher than the K = 3.0 colors they were able to remember when the displays were changed to be uniformly distributed in block 10 (See Figure 4-2; two-tailed t-test, t(9) = 4.90, p=0.0009; note that this is a within-subjects test, and so the between-subject error bars on Figure 4-2 underestimate the reliability of this effect). In addition, capacity for colors increased significantly across the first nine blocks of the experiment: one-way repeated measures ANOVA, F(8,72) = 12.28, p<0.0001. There was also a significant interaction between color capacity in the uniform condition and color capacity in the patterned condition across blocks, with observers in the patterned condition increasing their capacity more over time: F(8,144) = 2.85, p=0.006.

Seven of ten observers in the patterned condition reported noticing regular patterns in the display. The magnitude of the decrease in memory performance from block 9 to 10 was the same for observers who explicitly noticed the regularities (M=26%) and those who did not (M=27%), and 9 of 10 observers showed such decreases (Mean decrease across all observers = 26%). In addition, one of ten observers in the uniform condition reported noticing regular patterns in the configuration of colors, although no such patterns existed.

Post-perceptual inference

One concern is that observers might simply have remembered one color from each pair and then inferred what the other colors were after the display was gone. This would suggest that observers were actually only remembering three or four colors and were using a post-perceptual guessing strategy to achieve a higher performance in the memory test. This makes two predictions. First, when a color from a low probability pair is tested (20% of the time), observers should guess wrong and thus should show worse performance on these pairs over time. Second, on these trials they should guess wrong in a specific waythat is, they should guess the high-probability color of the item in the adjacent location. For example, if an observer only remembers the outside of color of an object was blue, and the inside color is tested, they should wrongly infer and report the high probability color that is often paired with blue.

To test these two predictions, we separated out trials where the tested item was from a high probability pair from those where the tested item was from a low probability pair. In other words, if blue often appeared inside red, we considered only the 20% of trials where blue appeared with another color or in another configuration. On these trials, an explicit inference process would cause observers to report the wrong color. However, we still find that performance improved over blocks (See Figure 4-3). Capacity (K), the number of colors remembered, is significantly greater in block 9, when the low-probability pairs are in the context of high probability pairs, than block 10, than when all the pairs are low-probability ($t(9) = 4.08$, $p=0.003$).

We next analyzed trials in the first 9 blocks where a color from a low probability pair was tested and observers answered incorrectly (on average there were 35 such

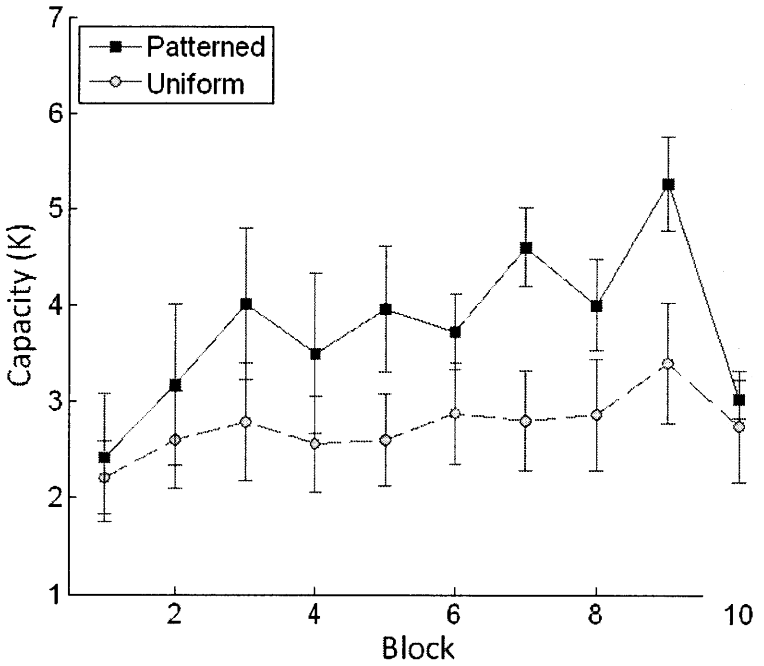


Figure 4-3: Results of Experiment 1 when only considering cases where the colors appeared in low probability pairings. Error bars correspond to +/-1 s.e.m. The dark squares represent data from observers in the patterned condition for the 20% of trials where a low probability pair was tested; the gray circles represent the data from observers in the uniform condition. The gray circle in block 10 corresponds to 100% of trials, since all pairs were low-probability.

trials per observer, for a total of 350 such trials across all 10 observers in the first experiment). If observers do not know what color was present and are explicitly inferring what was on the display using the high-probability pairings, then their responses should more often reflect the high-probability color of the adjacent item. However, on these trials, observers reported the high probability color of the adjacent item only 9% of the time (where chance is 1/7, or 14%). Further, observers wrongly report the high probability color of the tested color only 2% of the time. In fact, the only systematic trend on these low-probability error trials is that observers tend to swap the inner and outer colors much more often than chance: 41% of the time when observers were incorrect, they mistakenly reported the adjacent color. Interestingly, the rate of swaps with the adjacent color was lower in the high probability pairs: on trials where a high probability pair was tested, only 27% of error trials were explained by observers incorrectly reporting the adjacent color. This could be taken to suggest that the high probability pairs tend to be encoded as a single perceptual unit or chunk.

This analysis strongly argues against a post-perceptual account of increased memory capacity, where unencoded items are inferred during the testing stage. Not only do observers mostly get trials with the low probability pairs correct - suggesting they are not performing post-perceptual inference - but even on the trials where they do make mistakes, they do not tend to report the associated high probability colors, as would be predicted by an inference account.

Instead we suggest that observers learned to encode the high probability pairs using a more efficient representation. For example, suppose a display contains two high probability pairs and two low probability pairs. Over time, the high probability items are encoded more efficiently, leaving more memory resources for the low probability items. Such an account explains why even colors presented in low probability pairs show improved memory performance relative to the uniform group, but only when they are on the same displays as high probability pairs. In addition, an analysis across trials demonstrates that, on trials with more high probability pairs in the display, more items were successfully encoded (K = 3.2, 3.2, 3.6, 4.0, 4.7 for 0, 1, 2, 3, and 4 high probability pairs in the display, averaged across the entire experiment).

This increase in capacity as a function of the number of high probability pairs was significant, F(4,36) = 4.25, p=0.0065. Furthermore, the only difference between displays containing 3 or 4 high probability pairs is whether the remaining pair's colors are presented in the proper inner-outer configuration. Nevertheless, there was a trend for performance in these two conditions to differ, suggesting that learning may have been specific to the spatial configuration (t(9) = 1.78; p = 0.11). Together with the fact that observers did not often flip the inner and outer color of the high probability pairs, this suggests that observers may have been encoding the inner and outer colors as a single bound unit or chunk.

4.2.3 Discussion

The present results indicate that, if we consider working memory capacity in terms of the number of colors remembered, observers were able to use the regularities in the displays to increase their capacity past what has been assumed to be a fixed limit of approximately three or four colors. When colors are redundant with each other (i.e., are correlated with each other), then observers can successfully encode more than simply 3 or 4 colors. This suggests that the information content of the stimuli is incredibly important to determining how many can be successfully stored (see Alvarez & Cavanagh, 2004 for converging evidence of fewer high-information-load items being stored).

These data can also be interpreted with respect to current psychological constructs for analyzing the capacity of visual working memory ('slots') and working memory more broadly ('chunks'). In visual working memory, it has been argued that objects with multiple features (e.g. color and orientation) can be stored in a single slot as effectively as objects with only a single feature (Luck & Vogel, 1997, Vogel, Woodman, & Luck, 2001). In these models, the unit of memory is thus considered an 'object', a collection of features that are spatiotemporally contiguous (Luck & Vogel, 1997; see Scholl, 2001, for evidence pertaining to the definition of objects in mid-level vision). However, it has been found that memory for objects with two values along a single feature dimension does not show the expected within-object advantage, suggesting that what can be stored in a slot is a single value along each feature dimension, rather than an entire object (e.g. a single object with two colors on it, as in the present experiment, is not represented in a single slot; see Olson & Jiang, 2002; Wheeler & Treisman, 2002; Xu, 2002 for further discussion). This is consistent with the present data from the uniform group, where capacity was 3 colors rather than 3 multi-color objects (6 colors).

The data from the patterned group represent a challenge to this view. The ability of the patterned group to remember up to 6 colors represents a capacity of more than a single color per object, suggesting that capacity cannot be fixed to 3-4 objects with a single value along each feature dimension. Instead, the present data can be framed in terms of a slot-model only if slots can hold not just one color, but multiple colors from the same object as the objects are learned over time. In this sense, slots of visual working memory become more like 'chunks' in the broader working memory literature (Cowan, 2001). We return to this issue in Experiment 2, when we explore whether these regularities can be used when they are present across objects.

We next performed an information theoretic analysis of the current data to examine if observers have a fixed working memory capacity when measured in bits. We can estimate the amount of redundancy in the displays to test the hypothesis that observers actually have the same amount of resources to allocate in both uniform and patterned conditions. On this account, the difference in memory performance comes from the fact that the patterned displays allow observers to allocate their memory space more effectively. This allows us to make quantitative predictions about working memory capacity given a specific amount of redundancy in the display.

4.3 Modeling

Modeling provides a formal framework for theories of compression, and allows us to test the hypothesis that there is a limit of visual working memory capacity not in terms of the number of colors that can be remembered, but in terms of the amount of information required to encode those colors. The modeling has four stages. First, we model how observers might learn the color regularities based on the number of times they saw each pair of colors. The probability of each color pair is estimated with a Bayesian model that accounts for the frequency with which each color pair appeared, plus a prior probability that the colors will be paired uniformly. Second, we assess how these learned statistics translate into representations in bits, using Huffman coding (Huffman, 1952). Huffman coding is a way of using the probabilities of a set of symbols to create a binary code for representing those symbols in a compressed format. This allowed us to estimate the number of bits required to encode each item on the display. Third, we show that the information theoretic model successfully predicts observers data, suggesting they perform near optimal compression. Finally, we show that a discrete chunking model can also fit the data. Importantly, the best fitting chunking model is one that closely approximates the information theoretic optimal. MATLAB code implementing the model can be downloaded from the first authors website.

4.3.1 Learning the color pairs

We used a Dirichlet-multinomial model (Gelman, Carlin, Stern, & Rubin, 2003) to infer the probability distribution that the stimuli were being drawn from, given the color pairs that had been observed. We let d equal the observations of color pairs. Thus, if the trial represented in Figure 4-1 is the first trial of the experiment, after this trial $d = 1$ Yellow-Green, 1 Black-White, 1 Blue-Red, 1 Magenta-Cyan. We assume that d is sampled from a multinomial distribution with parameter θ. In other words, we assume that at any point in the experiment, the set of stimuli we have seen so far is a result of repeated rolls of a weighted 64 sided die (one for each cell in the joint probability matrix; i.e., one for each color pair), where the chance of landing on the ith side of the 64 sided die is given by θ_i. Note that this is a simplification, since the experiment included the additional constraint that no color could appear multiple times in the same display. However, this constraint does not have a major effect on the expected distribution of stimuli once a large number of samples has been obtained, and was thus ignored in our formalization.

We set our a priori expectations about θ using a Dirichlet distribution with parameter α. The larger α is, the more strongly the model starts off assuming that the true distribution of the stimuli is a uniform distribution. The alpha parameter can be approximately interpreted as the number of trials the observers imagine having seen from a uniform distribution before the start of the experiment. Using statistical notation, the model can be written as:

$$\begin{aligned} \theta &\sim Dirichlet(\alpha) \\ d &\sim Multinomial(\theta) \end{aligned}$$

To fit the model to data we set a fixed α and assume that the counts of the pairs that were shown, d, are observed for some time period of the experiment. Our goal is to compute the posterior distribution $p(\theta|d, \alpha)$. The mean of this posterior distribution is an observers best guess at the true probability distribution that the stimuli are being drawn from, and the variance in the posterior indicates how certain the observer is about their estimate. The posterior of this model reduces to a Dirichlet posterior where the weight for each color pair is equal to the frequency with which that color pair appears in d, plus the prior on that pair, α_i.

4.3.2 Encoding the color pairs

Any finite set of options can be uniquely encoded into a string of bits. For example, if we wished to encode strings consisting of the four letters A, B, C, and D into strings of bits, we could do so by assigning a unique two bit code to each letter and then concatenating the codes. Imagine we had assigned the following codes to the letters: A = 00, B = 01, C = 10, D = 11. The string ACAABAA could then be written as 00100000010000 (14 bits), and uniquely decoded to retrieve the original string.

Importantly, however, this nave method of generating a code performs quite badly in the case where some letters are much more likely to appear than others. A better method gives items that occur most frequently the shortest codes, while less frequent items are assigned longer codes. So, for example, if P(A) = 0.5, and P(B) = 0.2, P(C) = 0.2, and P(D) = 0.1, then we can achieve a great deal of compression by representing strings from this language using a different code: A = 0, B = 10, C = 110, D = 111. Using this code, the string from above, ACAABAA, would be represented as 0110001000 (10 bits), a significant savings even for such a short string (29%). Note that it can still be uniquely decoded, because no items code is the same as the beginning of a different items code.

Huffman coding (Huffman, 1952) is a way of using the probabilities of a set of symbols to create a binary code for representing those symbols in a compressed format. (as in the example of A, B, C, D above). Here, we used Huffman coding to estimate how much savings observers should show as a result of the fact that the color pairs in our experiment were drawn from a non-uniform distribution. In the Appendix, we demonstrate that the same results also hold for another way of assessing compression using self-information.

We used the probabilities of each color pair, as assessed by the Bayesian model described above, to generate a unique bit string encoding the stimuli on each trial, averaged for each block of the experiment. We supposed that if observers were using some form of compression to take advantage of the redundancies in the display, the length of the code that our compression algorithm generates should be inversely proportional to how many objects observers were able to successfully encode. In other words, if there were many low frequency color pairs presented (as in block 10), these items should have longer codes, and observers should be able to successfully remember fewer of them. Alternatively, if there are many high frequency color pairs presented, the better they should be able to compress the input, and the more colors they remember.

4.3.3 Information theory

With these learning and coding models, we can compute a prediction about the memory performance for each subject for each block. In order to assess the fit between the model and the behavioral data, we used the following procedure. For each display

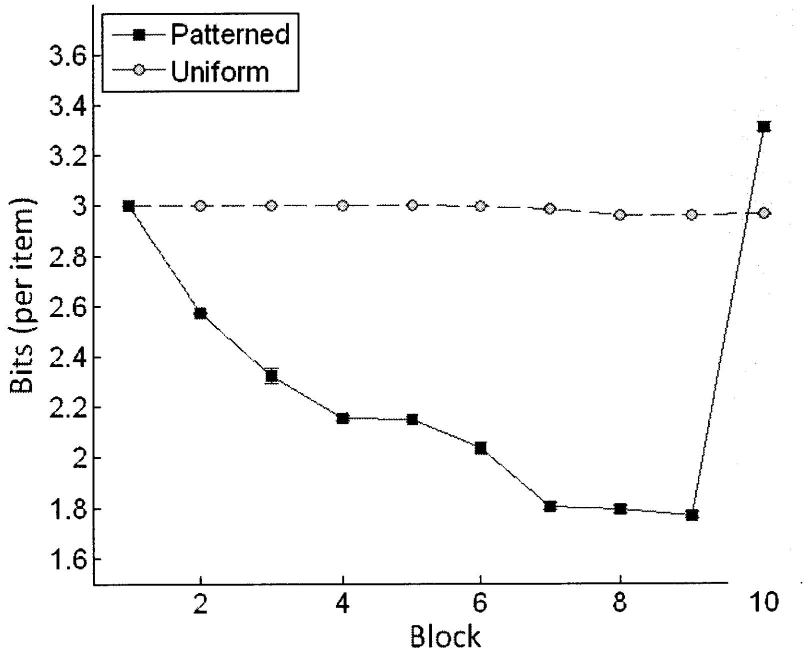


Figure 4-4: The average length of the Huffman code for a single color, by block.

in a block, we calculated the number of bits required to encode that display based on the probabilities from the learning model. Next, we correlated the average number of bits per display from the model with the memory performance of the observers. We expect that the fewer bits/display needed, the better observers memory performance, and thus we expect a negative correlation.

This prediction holds quite well, with the maximum fit between this Huffman code model and the human data at a = 34, where r, the correlation coefficient between the human and model data, is -0.96 (See Figure 4-4; p<0.0001). This large negative correlation means that when the model predicts there should be long bit strings necessary to encode the stimuli, human visual working memory stores a low number of items. This is exactly as you would expect if visual working memory took advantage of a compression scheme to eliminate redundant information. In addition, this modeling suggests that if observers encoded the displays completely optimally, they would be able to remember approximately 6.1 colors. By block 9, observers are remembering 5.4 colors on average, significantly better than with no compression at all, but not

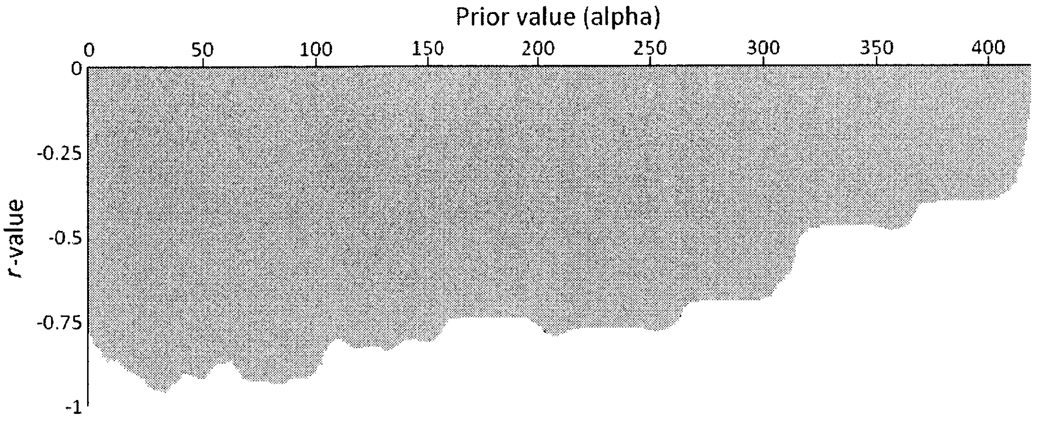


Figure 4-5: The correlation between the information theoretic model and the human behavioral data as a function of the value of the prior, α.

quite at the theoretically maximal compression.

The fit between the human data and the model is reasonably good across a broad range of values for the prior probability of a uniform distribution (see Figure 4-5). The fit is not as high where the prior is very low, since with no prior there is no learning curve the model immediately decides that whatever stimuli it has seen are completely representative of the distribution (as a non-Bayesian model would do). The fit is also poor where the prior is very high, because it never learns anything about the distribution of the stimuli, instead generating codes the entire time as though the distribution was uniform. However, across much of the middle range, the model provides a reasonable approximation to human performance.

Importantly, this model allows us to examine if there is a fixed information limit on memory capacity. The Huffman codes provide a measure of the average number of bits per object, and the memory performance gives a measure in number of colors remembered. Thus, if we multiply the average bits / item specified by the Huffman code times the number of items remembered, we get an estimate of the number of bits of information a given set of observers recalled in a given block (Figure 4-6). Notice first that both groups of observers in the uniform condition and the patterned condition show roughly the same total capacity in bits, despite the overall

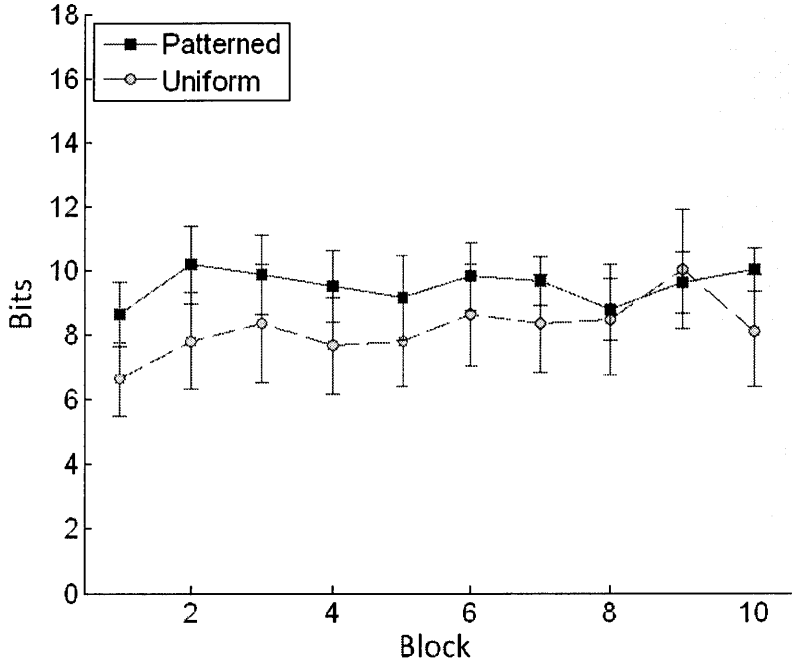


Figure 4-6: The size of memory estimated in bits, rather than number of colors (using the Huffman coding model). Error bars represent +/-1 s.e.m.

difference in the number of items remembered between the groups. Second, the total bit estimate remains remarkably constant between block 9 and block 10 in the patterned group, even though the memory performance measured in number of colors showed a significant cost when the statistical regularities were removed. Thus, while the patterned group was able to remember more colors throughout the experiment, this increase was completely explained in the model by the fact that the items to be remembered were more redundant and presumably took less space in memory.

4.3.4 Chunking model

The information theoretic modeling gives a way of formally specifying how compressible a set of input is given the accumulated statistics about the previous input. Huffman coding and self-information are ways to formalize this, and are thus a form of rational analysis or computational theory, specifying the optimal solution to the computational problem facing the observer (Anderson, 1990; Marr, 1982). Interestingly, we find that observers closely approximate this optimum. However, Huffman coding and self-information are not meant as serious candidates for the psychological mechanism people use for implementing such compression. Indeed, it is a different level of analysis to understand what psychological algorithms and representations are actually responsible for allowing more items to be encoded in memory when those items are redundant with each other. For instance, is the nature of the compression graded over time, or all-or-none?

In the information theoretic models (Huffman coding, self-information), the 'cost' of encoding each color pair is equal to the log of the chance of seeing that pair relative to the chance of seeing any other pair. This is the optimal cost for encoding items if they appear with a given probability, and provides for graded compression of a color pair as the items probability of co-occurrence increases. However, the actual psychological mechanism that people use to remember more items could be either graded as in the rational analysis, or could function as a discrete approximation to this optimum by sometimes encoding highly associated items into a single representation. Chunking models are one way of approaching this kind of discrete approximation (e.g., Cowan et al. 2004). They show increased memory capacity for highly associated items, but convert the compressibility to a discrete form: either a single chunk is encoded or the two colors are separately encoded into two chunks. This distinction between graded compression and all-or-none compression is important because it predicts what is actually encoded by an observer in a single trial. The current results do not address this distinction directly, however, since we do not examine the representational format of the color pairs on each trial. However, there is a broad literature with a preference for viewing compression in working memory as based on discrete chunking (e.g., Cowan, 2005; Chase & Simon, 1973; Miller, 1956; however, see Alvarez & Cavanagh, 2004; Bays & Husain, 2008; and Wilken & Ma, 2004, for support for a graded view). Thus, we sought to examine whether our data could be accurately modeled using this kind of approximation to the information theoretic analysis presented above.

To implement a simple chunking model, one needs to determine a threshold at which associated items become a chunk. The most naive chunking model is one in which observers reach some fixed threshold of learning that a pair of colors co-occur and treat them as a chunk thereafter (perhaps after this new chunk enters long-term memory). However, this simple model provides a poor fit to the current data. In such a model, each subject will have a strong step-like function in their graph, and the graded form of the group data will arise from averaging across observers. However, in the present data, single observers showed a graded increase in performance by block, suggesting this kind of model does not accurately represent the data.

A more sophisticated class of chunking models have a probabilistic threshold, allowing for a single observer to treat each color pair as one chunk more often if they strongly believe it is a chunk, and less often if they are unsure if it is a chunk. In the case where the chance of chunking in such a model is logarithmically proportional to the association between the items, this chunking model is exactly equivalent to a thresholded version of the information theoretic compression model and therefore makes the same predictions across large numbers of trials. However, a chunking model could also assume that the possibility of chunking is linearly proportional to the association between the items ($p_{chunk(i,j)} = \beta * \theta_{i,j}$), in which case it would be possible that the chunking models fit would differ significantly from that of the more ideal compression algorithms. We do not find this to be the case for the current experiment.

The graph from the best fit linear chunking model is shown in Figure 4-7. The best fit constant of proportionality was 15, which provided a fit to the data of r=-0.90 (e.g., for each pair, the chance of being chunked on any given trial was equal to $15 * \theta_{i,j}$, such that once the probability of seeing a given color pair was greater than 1/15th, that color pair was always encoded as a single chunk). Interestingly, this constant of proportionality, because it causes such a steep increase in the chance of chunking even at very low associations and plateaus at a 100% chance of chunking by the time the association reaches 1/15, or 0.067, approximates the shape of a logarithmic curve. The correlation between the probability of chunking under this linear model and the optimal cost function derived via information theory (using self-information) is therefore approximately r=-0.73. This model thus provides an

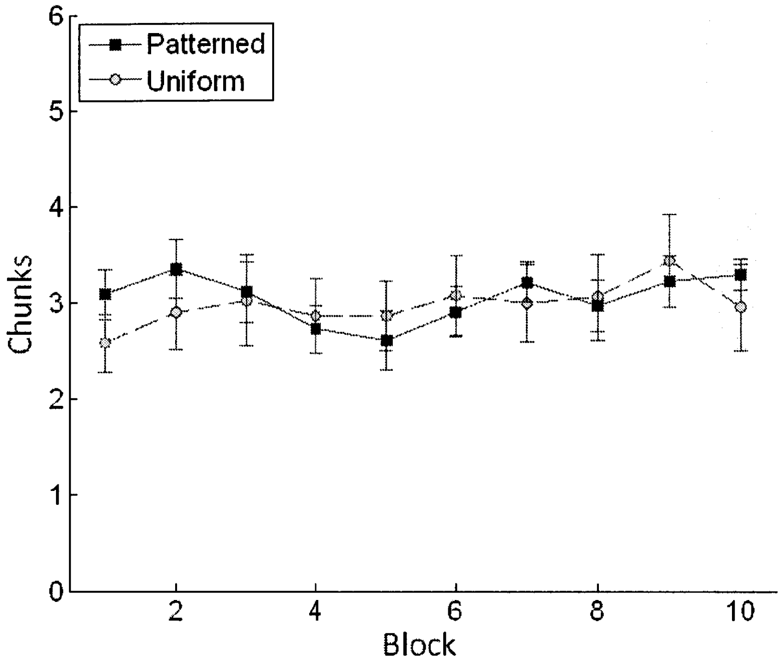


Figure 4-7: The size of memory (in chunks) for Experiment 1 estimated using the probabilistic linear chunking model. Error bars represent +/-1 s.e.m.

excellent approximation to the ideal compression algorithm as well.

Thus, we find that the best chunk-model matches the data well and generates a flat estimate of the number of chunks needed across the entire experiment. Importantly, however, the expected probability of chunking in this model closely matches the optimal information-theoretic cost function (higher cost = lower probability of chunking). This is to be expected because the information theoretic model predicted 92 percent of the variance in the behavioral data. This suggests that chunking can be usefully thought of as a discrete approximation to an ideal compression algorithm, and therefore can be thought of as a possible psychological implementation of compression.

It is important to note that, despite the assumptions we make in this modeling section, it unlikely that the degree of association between items determines when they form chunks in long-term memory. Instead, it may be that human chunk learning depends on how useful a particular chunk would be in describing the world while avoiding suspicious coincidences (see, for example, Orban et al. 2008, which provides an elegant Bayesian analysis of this problem). Our analysis of chunking here is meant only as a proof of concept that chunking models in general implement a form of compression that approximates the true information theoretic optimum.

4.3.5 Discussion

The modeling work we present illustrates two main conclusions: First, compression of redundancies must be taken into account when quantifying human visual working memory capacity; Second, this compression can be modeled either in a graded fashion, or in an all-or-none fashion ('probabilistic chunking') which closely approximates ideal compression algorithms.

The fact that the estimate of the amount of information observers are able to store is constant across the entire experiment, whereas the estimate in terms of number of colors varies a great deal, suggests that compression of redundancies must be taken into account when quantifying human visual working memory capacity. In addition, it is important to note that fitting our information theoretic model by minimizing the correlation to the data is not guaranteed to provide a fit that results in a flat line in terms of the total information remembered. In fact, in most instances a negative correlation will not lead to a flat estimate across the experiment, since a flat line additionally depends on the proportional amount of the decrease at each step. The information theoretic modeling results provide significant evidence that the capacity of working memory is a fixed amount of information. Because the chunking model is the discrete version of an optimal compression scheme, this model leads to a fixed capacity measured in discrete units ('chunks') just as the information theoretic model let to a fixed capacity measured in continuous information ('bits').

While our model suggests a working memory capacity of 10 bits, this number should not be taken as indicative of limits on human performance. The exact number 10 bits depends critically on assumptions about how the colors are encoded (3 bits/-color in our model, given the 8 possible color choices). Importantly, however, the fact that the estimate of memory size is constant across the experiment and across conditions does not depend on our choice of encoding scheme, but only on the redundancy inherent in the associations between colors. If observers actually required 100 bits to encode each color then our estimate of capacity in bits would change to be about 300 bits but the estimate would still remain consistent across the experiment, since each color still provides the same proportional amount of information about each other color. Thus, it is safe to conclude that our results are compatible with a fixed amount of information limiting memory performance, but it is difficult to quantify the exact number of bits without specifying the true coding model (see the General Discussion for further discussion of the problem of specifying an encoding scheme).

4.4 Experiment 2: Regularities Between Objects

The aim of Experiment 2 was to examine if compression can affect encoding across objects as well as within objects. This experiment was very similar to Experiment 1, with the only difference being how the colors were presented on the display. In Experiment 2, colors were presented side-by-side as separate objects, in close proximity but not spatially contiguous.

While there are many possible definitions of 'object', we use the term to refer to a specific well-defined definition of what counts as an object for mid-level vision. Specifically, an object is a spatiotemporally contiguous collection of visual features (Scholl, 2001; Spelke, 1990). This definition is motivated by both neuropsychological and behavioral evidence (Behrmann & Tipper, 1994; Egly, Driver, & Rafal, 1994; Mattingley, Davis, & Driver, 1997; Scholl, Pylyshyn, & Feldman, 2001; Watson & Kramer, 1999). For example, simply connecting two circles with a line to form a dumbbell can induce 'object-based neglect,' in which the left half of the dumbbell is neglected regardless of the half of the visual field in which it is presented (Behrmann & Tipper, 1994). If these two circles are not connected, neglect does not operate in an object-based manner. Thus, on this definition of what counts as an object, the displays of Experiment 1 contained 4 objects while the displays of Experiment 2 contained 8 objects (see Figure 4-8).

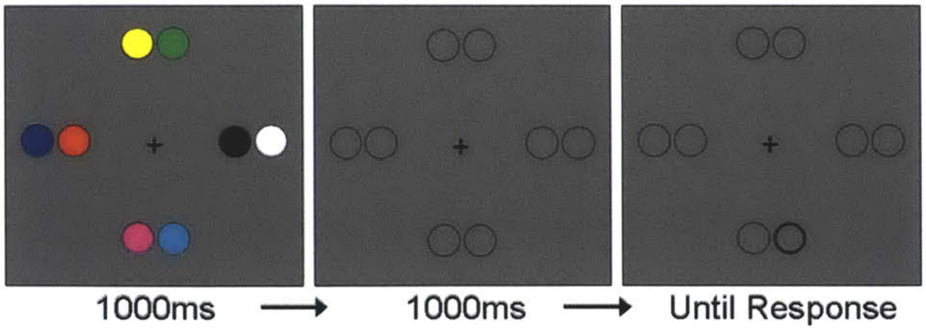


Figure 4-8: A sample trial from Experiment 2. Eight colors were presented and then disappeared, and after 1 second, one location was cued by making it darker. Observers had to indicate what color was at the cued location.

In the present experiment, we examined whether or not working memory capacity can take advantage of the statistics between objects. If visual working memory capacity limits are 'object-based', i.e. if capacity is constrained by mid-level visual objects, then observers will not be able to take advantage of regularities across objects. However, if multiple visual objects can be stored together, (akin to 'chunks' of letters, as in FBI-CIA), then people will be able to remember more colors from the display as they learn the statistics of the input.

4.4.1 Method

Observers

Twenty naive observers were recruited from the MIT participant pool (age range 18-35) and received 10 dollars for their participation. All gave informed consent.

Procedure

Observers were presented with displays consisting of eight objects arranged in four pairs around the fixation point (see sample display in Figure 4-8). Each object was made up of only one colored circle. Here the two associated colors appeared on separate objects, but we provided a grouping cue in order to not significantly increase the difficulty of the learning problem. All other aspects of the stimuli and procedure were identical to those of Experiment 1.

4.4.2 Results

Performance Across Groups

Observers in the uniform condition remembered K = 3.4 colors on average throughout the experiment (see Figure 4-9), consistent with previous results on the capacity of visual working memory for colors (Vogel & Awh, 2008), and the results of Experiment 1.

We found that observers in the patterned condition could successfully remember K = 5.4 colors after learning the regularities in the displays (block 9). This memory capacity is significantly higher than the K = 3.3 colors they were able to remember when the displays were changed to be uniformly distributed in block 10 (See Figure 4-9; t(9) = 9.72, $p<0.0001$). In addition, capacity increased significantly across the first nine blocks of the experiment: F(8,72) = 7.68, $p<0.0001$. There was a significant interaction across blocks between capacity in the uniform condition and capacity in the patterned condition, with observers in the patterned condition remembering more colors over time: F(8,144) = 2.27, p=0.025.

Eight of ten observers reported noticing regular patterns in the display. The magnitude of the decrease in memory performance from block 9 to 10 was the same for observers who explicitly noticed the regularities (M=22%) and those who did not (M=23%), and 9 of 10 observers showed such decreases (mean decrease across all observers 23%). Three of ten observers in the uniform condition reported noticing regular patterns in the configuration of colors, although no such patterns were present.

We once again separated out trials where the tested item was from a high probability pair from those where the tested item was from a low probability pair. When we examine only the low probability trials, we still find that capacity in block 9 is significantly higher than in block 10 (4.9 colors in block 9 and 3.4 colors in block 10; t(9)=4.84, p=0.0009). Thus, as with Experiment 1, we do not find evidence that people are remembering more items from the display by using post-perceptual inference.

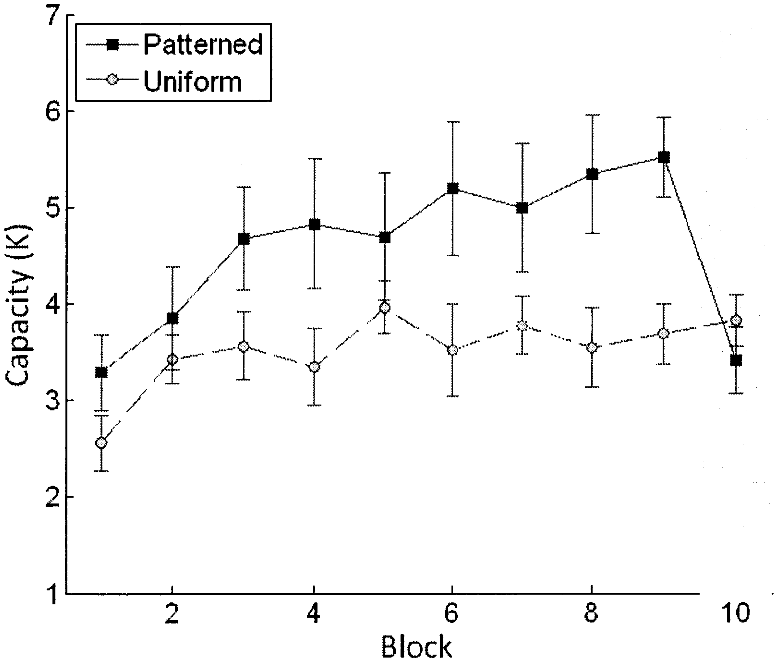


Figure 4-9: Results of Experiment 2. Error bars correspond to +/-1 s.e.m.

Performance Across Experiments

We compared the first 9 blocks in the patterned condition to the first 9 blocks in the patterned condition of Experiment 1. There were no main effects or interactions, all F < 1. Furthermore, we compared the drop in performance between block 9 and block 10 across the two experiments. The size of the drop was not significantly different, t(9) = 0.58, p=0.58, suggesting that learning was of a comparable magnitude in both experiments. Verbal Interference

One potential concern is that observers could have used some verbal memory capacity to augment their visual working memory in either the current experiment or Experiment 1. Many past studies have found that estimates of visual working memory capacity are similar with and without verbal interference (e.g., Luck & Vogel, 1997; Vogel et al. 2001). However, because of the added element of learning regularities in our experiments, we decided to test the effects of verbal interference on our paradigm. Because of the similarities between Experiment 1 and Experiment 2, we ran a control experiment using only the paradigm of Experiment 2.

This control experiment was conducted with 7 observers using an identical paradigm to Experiment 2s patterned condition, but with the addition of a verbal interference task (remembering 4 consonants throughout the duration of the trial, with a new set of 4 consonants every 10 trials). Observers successfully performed both the verbal interference task and the visual working memory task, with a capacity of 4.5 colors in block 9 but only 3.2 colors in block 10 ($t(6) = 2.1$; $p=0.08$). Capacity in block 9 under verbal interference was not significantly different than that obtained in block 9 of Experiment 2 ($t(9) = 1.07$, $p=0.31$). These data show that observers are still capable of learning the regularities to remember more colors, when subject to verbal interference in a challenging dual-task setting.

Modeling

We once again modeled these results to see if they were compatible with a model in which compression is explained via information theory. The maximum fit between the Huffman code model and the human data occurred at $a = 31$ where r, the correlation coefficient between the human and model data, is -0.96 ($p < 0.0001$). This large negative correlation means that when the model predicts there should be long bit strings necessary to encode the stimuli, observers memory capacity in terms of the number of colors remembered is low. This is exactly what one would expect if visual working memory had a fixed size in bits and took advantage of a compression scheme to eliminate redundant information.

In addition, this model allows us to once again examine if there is a fixed-bit limit on memory capacity. The Huffman codes gives a measure of average bits per object, and the memory performance gives a measure in number of objects remembered. As in Experiment 1, multiplying the average size of the Huffman code times the number of items remembered gives us an estimate of the number of bits of information a given set of observers recalled in a given block (Figure 4-10). Notice that once again both the groups of observers in the uniform condition and the patterned condition show the same total capacity in bits, despite the overall difference in the number of items

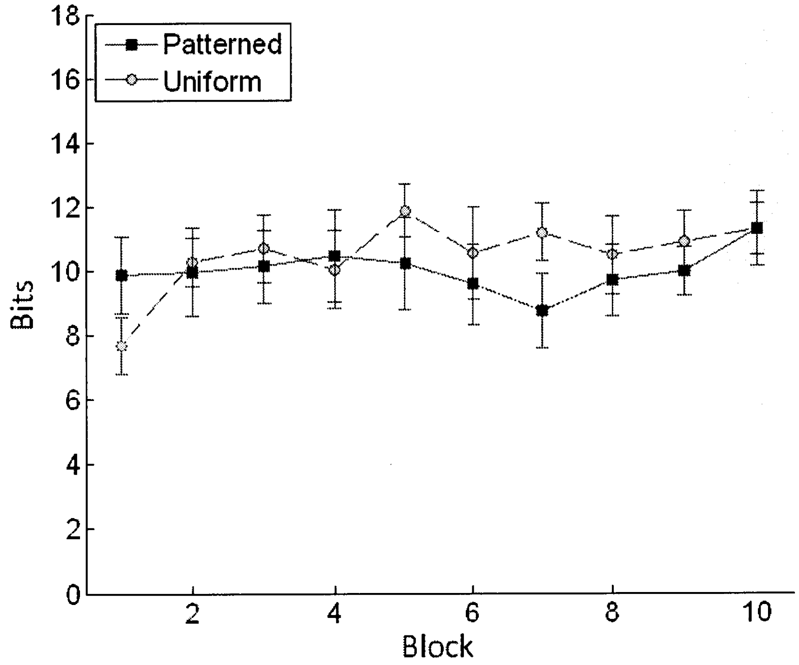


Figure 4-10: The size of memory estimated in bits, rather than number of objects (using the Huffman coding model). Error bars represent +/-1 s.e.m.

remembered between the groups. Second, the total bit estimate remains remarkably constant between block 9 and block 10 in the patterned group, even though the memory performance measured in number of items showed a significant cost when the statistical regularities were removed.

One interesting prediction of the model is that the patterned group should actually be worse at block 10 than the uniform group, since the patterned group now has a set of statistics in mind that are no longer optimal for the displays. Indeed, the pattern in the behavioral data trends this way, but the difference between both groups in block 10 is not significant (See Figure 4-9; t(9) = 0.64; p=0.47). One possible explanation for why performance for the patterned group does not fall completely below the uniform group is that observers notice that their model has become inappropriate after several trials in block 10, and begin using a relatively local estimate of the probability distribution (e.g., across the last few trials), or revert to a uniform model. This suggests a good deal of flexibility in the model observers use to encode the

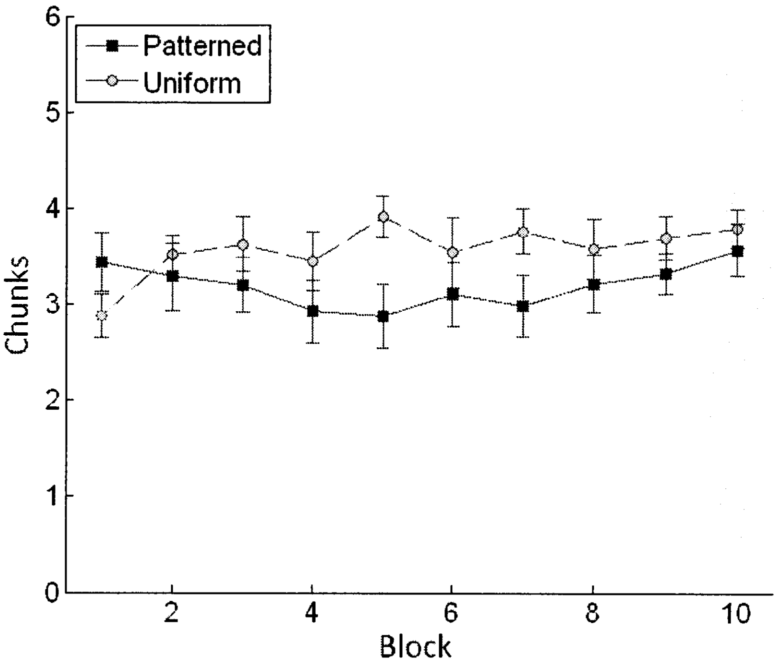


Figure 4-11: The size of memory (in chunks) for Experiment 2 estimated using the probabilistic linear chunking model. Error bars represent +/-1 s.e.m.

display.

In addition, we modeled these results using a probabilistic chunking model where the chance of chunking was linearly proportional to the probability of the color pair. Using the same parameters as in Experiment 1, this model too provided a good fit to the data (r=0.94; see Figure 4-11), and it produced an almost flat estimate of the number of chunks over time in both groups.

4.4.3 Discussion

Observers in the patterned group were able to successfully take advantage of the redundancy in the displays, as their capacity increased significantly over time. These data, as well as the estimated capacity in bits from the modeling, reveal strikingly similar patterns between Experiment 1 and Experiment 2. This suggests that observers were equally able to take advantage of the redundancy in the displays when the redundancies were present between adjacent mid-level visual objects rather than within such objects.

This experiment has some implications for the classic slot-model of visual working memory (Luck & Vogel, 1997; Zhang & Luck, 2008). Specifically, a strict interpretation that one slot can hold only one mid-level visual object from the display does not account for the present data. The patterned group was able to remember more objects over time, so the capacity of working memory cannot be a fixed number of mid-level visual objects. However, if multiple objects can fit into one slot, then the present data can be accounted for. Indeed, this suggests that 'slots' in visual working memory should be viewed similarly to 'chunks' in verbal working memory (Cowan, 2001). Thus, in the present experiment 'visual chunks' could be formed that consist of pairs of colored objects (see also Orban et al., 2008 for evidence of statistical learning of chunks of multiple objects). Of course, another possibility is that working memory capacity should be thought of as graded, rather than based on chunks or slots at all (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Bays & Husain, 2009). This would make it more closely approximate the information theoretic ideal and would account for the present data directly.

An important factor in the present experiment is that we provided a grouping cue for observers, by putting the two colors that will co-vary in closer proximity to each other than to the other colors. We expect that learning would still be possible even if the items were not specifically grouped, as others have demonstrated that statistical learning can operate across objects, even in cases when the display is unparsed, and that such learning results in the formation of visual chunks (Fiser & Aslin, 2001; Fiser & Aslin, 2005; Orban et al., 2008; see also Baker et al., 2004). However, our aim in this experiment was not to create a difficult learning situation; rather, our aim was to demonstrate that visual working memory can take advantage of these learned statistics to remember more of the display even when the statistics relate the co-occurrence of different objects, as in the work of Fiser and Aslin (2001). It is an avenue of future research to explore what kinds of statistics can be gleaned from the input and where the statistical learning mechanisms fail. It will also be important to discover if there exist situations in which observers can successfully learn statistical regularities but are unable to take advantage of those regularities to efficiently store items in memory.

Finally, as in Experiment 1, the modeling showed that even though people are remembering more items, their working memory capacity is actually constant when quantified by the amount of information remembered (or the number of chunks stored). In general, this suggests that the tools of information theory combined with Bayesian learning models enable us to take into account the compressibility of the input information, and provide clear, testable predictions for how many items observers can remember. This suggests that compression must be central to our understanding of visual working memory capacity.

4.5 General Discussion

We presented two experiments contrasting memory capacity for displays where colors were presented in random pairs with memory capacity for displays where colors were presented in recurring patterns. In the first experiment, the colors which formed a pattern were presented as part of the same object. In the second experiment, the colors which formed a pattern were presented on two different but spatially adjacent objects. For both experiments we found that observers were successfully able to remember more colors on the displays in which regularities were present. The data indicate that this is not due to post-perceptual inference but reflects an efficient encoding. We proposed a quantitative model of how learning the statistics of the input would allow observers to form more efficient representations of the displays, and used a compression algorithm (Huffman coding) to demonstrate that observers performance approaches what would be optimal if their memory had a fixed capacity in bits. In addition, we illustrated that a discrete model of chunking also fits our data. The degree of compression possible from the display was highly correlated with behavior, suggesting that people optimally take advantage of statistical regularities to remember more information in working memory.

We thus show that information theory can accurately describe observers working memory capacity for simple colors that are associated with each other, since such a capacity depends on compression. By using a statistical learning paradigm, we control the statistics of the input, allowing is to measure the possible compression in this simple task. Since in the world almost all items we wish to remember are associated with other objects in the environment (Bar, 2004), using information theory to quantify the limits of working memory capacity is likely of more utility for natural viewing conditions than measuring the number of independent items that people can remember.

4.5.1 Resolution versus number

One interesting factor to consider is whether the increase in percent correct we observe during training in the patterned group could be due to an increase in the resolution at which observers store the colors, rather than an increase in the number of colors remembered per se (similar to the claims of Awh, Barton, & Vogel, 2007).

We believe several factors speak against such an account. In particular, if a fixed number of items are remembered and only the resolution of storage is increasing, then the fixed number of items remembered would have to be at least 6 (the number of colors remembered by the patterned group in the 9th block of trials). This seems very unlikely, given that previous estimates of the fixed number are on the order of 3 or 4 (Luck & Vogel, 1997; Cowan, 2001), even for studies that explicitly address this issue of the resolution with which items are stored (Awh, Barton, & Vogel, 2007; Zhang and Luck, 2008). In addition, while Awh et al (2007) provide some evidence that for complex objects there may be a difference of resolution between different object classes, both Rouder et al. (2008) and Zhang and Luck (2008) have recently argued for discrete fixed-resolution representations in the domain of color (although see Bays & Husain, 2009, for a critique of this work). These papers provide evidence that for simple features like color, the colors are either remembered or not remembered, rather than varying in resolution. Finally, it is not clear why the covariance introduced in the colors would affect the resolution of a single color, and what the proper relationship would be between the resolution and the association strength. For these reasons we believe it is unlikely that the current data reflect changes in the resolution of the items rather than the quantity of items stored.

4.5.2 The relationship between slots and objects

Much of the work on visual working memory has emphasized the privileged role of objects in memory capacity. For example, there is often an advantage to representing two features from the same object as opposed to two of the same features from two different objects (Luck & Vogel, 1997; Xu, 2002). In fact, visual working memory is often conceptualized as containing 3-4 'slots', in which one object, and all its features, can be stored into one slot (e.g., Luck & Vogel, 1997; Zhang & Luck, 2008) with some degree of fidelity. In this literature, 'objects' typically are assumed be the units of mid-level vision, specifically a spatiotemporally contiguous collection of features.

Our data suggest that at least the simplest version of an object-based capacity limit, in which one object in the world is stored in one slot in the mind, is not sufficient. If observers have a fixed working memory capacity of 3-4 objects on average, then both the uniform and patterned groups should show the same memory performance in Experiment 2. Similarly, if observers can remember at most 3-4 values along a single feature dimension (like color), then both the uniform and patterned groups should show the same memory performance in Experiment 1. However, in both Experiment 1 and Experiment 2, the patterned groups were able to remember almost twice as many objects by the end of the experiment. Thus, if there are slots in the mind, they must be able to hold more than one mid-level visual object, much like 'chunks' can contain multiple digits or words in the verbal working memory literature. The critical point here is that visual working memory should not be said to hold only 3-4 mid-level visual objects or 3-4 values along a single feature dimension, but instead needs to allow for 'visual chunking'. Alternatively, visual working memory capacity may be characterized in a more graded fashion rather than using slots or chunks as a unit of measure (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wilken & Ma, 2004).

4.5.3 Chunking

The current behavioral data cannot directly address whether the proper way to characterize the capacity of the system is in terms of a continuous measure of information or in terms of a model in which items are stored discretely in chunks or slots (Cowan, 2001; Luck & Vogel, 1997; Miller, 1956; Simon, 1974). Our information theoretic analysis puts a theoretical limit on how compressible this information should be to a learner. However, exactly how this compression is implemented in psychological constructs remains an open question. One possibility is that associated items become more and more compressible over time (i.e., start to take up less space in memory). Another possibility is that pairs of items take up either two chunks or one, depending on a probabilistic chunking threshold. Importantly, a continuous model of compression can be closely approximated by a discrete model, as long as the threshold for forming a chunk is related to the cost of the items in information theoretic terms. In fact, any chunking model which will account for our data will need to form chunks in a way that is compatible with our information theoretic analysis. In this sense, information theory allows us to constrain chunking models significantly, and has the potential to break us out of the circular dilemma of determining what ought to count as a single chunk (Simon, 1974).

4.5.4 Coding model

It is important to emphasize that compression must be defined with respect to a coding model. Naive information theoretic models (e.g., Kleinberg & Kaufman, 1971), which simply assume that all items are coded with respect to the possible choices for a particular task, are not adequate ways of characterizing the capacity of the memory system. For example, using such a coding scheme it takes 1 bit to represent a binary digit and 3.3 bits to represent a decimal digit. However, as described clearly in Miller (1956), if observers can remember a fixed amount of information, then based on the number of decimal digits they can remember, they ought to be able to remember many more binary digits.

Some hints at what the psychological coding model might be like in this case comes from evidence that shows observers tend to store digits phonetically (Baddeley, 1986). Thus, perhaps a proper information theoretic model would encode both binary and decimal digits with respect to the entire set of phonemes. Of course, even the phoenetic coding scheme is not sufficient for capturing how much information is in a string, as the conceptual content matters a great deal. For example, people can remember many more words if they make a coherent sentence than if they are randomly drawn from the lexicon (Simon, 1974). This is also true in memory for visual information: sparse cartoon drawings are remembered better when given a meaningful interpretation (Bower, Karlin, & Dueck, 1975; see also Wiseman & Neisser, 1974). Presumably abstract line drawings have much longer 'coding strings' than when those same line drawings can be encoded with respect to existing knowledge.

In the current experiment, we specifically avoided having to discover and specify the true coding model. By exploring compression within the domain of associations between elements (colors in the current study), we only need to specify the information present in their covariance. Specifying how long the bit string is for a display of eight colored circles would require a complete model of the visual system, and how it encodes the dimensions of colored circles. Since the true coding model is likely based in part on the natural statistics of the visual input, and given the frequency of gray screen with eight colored circles on them in our everyday visual experience, the bit strings for such a display are likely quite long. Instead we used a paradigm that builds associations between elements over time, allowing us to control the coding model that could be learned from the regularities in the displays. This method avoids many of the pitfalls traditionally associated with information theoretic models (e.g., those examined by Miller, 1956). Importantly, our results demonstrate that in this simplified world of associated colors, visual working memory is sensitive to the incoming statistics of the input. This approach opens the door for future work to apply information theoretic models to human cognition without first solving for the perceptual coding schemes used by the brain.

Moving beyond simple pairwise associations between colors, for more complex stimuli and in more real-world situations, observers can bring to bear rich conceptual structures in long-term memory and thus achieve much greater memory performance (e.g., Ericsson, Chase, & Faloon, 1980). These conceptual structures act as internal models of the world, and therefore provide likelihoods of different items appearing in the world together. For example, observers know that computer monitors tend to appear on desks; that verbs follow subjects; that kitchens tend to be near dining rooms. Importantly, our information theoretic framework can, at least in principle, scale up to these more difficult problems, since it is embedded in a broader Bayesian framework which can make use of structured knowledge representations (Kemp & Tenenbaum, 2008; Tenenbaum, Griffiths, & Kemp, 2006).

4.5.5 Relation to learning and long-term memory

Compressibility and chunking are rarely formalized outside the literature on expertise (e.g., chunking models: Gobet et al., 2001), and thus the relation between visual working memory capacity and the learning of relations between items has received little attention in the literature (although see Cowan et al. 2004 for an analysis in the verbal domain). However, there are several interesting data points about the role of learned knowledge in working memory capacity more broadly: for example, adults have a greater working memory capacity than children (Simon, 1974). In addition, there is a large literature on expertise and chunking (Chase & Simon, 1973; Gobet et al., 2001), where there is significant appreciation of the fact that long-term knowledge is a significant factor in working memory capacity (see also Kurby, Glazek & Gauthier, 2009; Olsson & Poom, 2005; Scolari, Vogel & Awh, 2008).

By relating working memory capacity and chunking strongly to information theory, our results suggest a broad purpose for a particular kind of long-term knowledge acquisition: statistical learning. In particular, a great deal of recent work has focused on a set of statistical learning mechanisms which are capable of extracting many different regularities with only minutes of exposure and appear to be relatively ubiquitous, occurring in the auditory, tactile and visual domains, and in infants, adults, and monkeys (Brady & Oliva, 2008; Conway & Christiansen, 2005; Fiser & Aslin, 2002; Kirkham, Slemmer & Johnson, 2002; Hauser, Newport & Aslin, 2001; Saffran, Aslin & Newport, 1996; Turk-Browne, Junge & Scholl, 2005). The present results suggest that one of the primary reasons for being sensitive to such regularities might be that it allows us to remember more in working memory by eliminating redundancy in our representations. They also emphasize how quickly such long-term memories can be built and can start to influence capacity measures observers in the present studies demonstrated significant improvements in working memory capacity by block 2, only a few minutes into the experiment. In addition, it is important to keep in mind that statistical learning mechanisms need not be limited to learning simple associations between items. Both the learning process and the representations that are learned can be, and likely are, much richer than simple associations (see, for example, Orban et al. 2008 and Frank, Goldwater, Mansinghka, Griffiths & Tenenbaum, 2007).

4.5.6 Conclusion

The information we can hold in working memory is surprisingly limited. However, in the real world there are strong associations and regularities in the input, and our brain is tuned to these regularities in both perception and memory (Field, 1987; Anderson, 1990). In an information theoretic sense, such regularities introduce redundancies that make the input more compressible.

We have shown that observers can take advantage of these redundancies, enabling them to remember more colors in visual working memory. In addition, while we showed this using simple associations between colors, the Bayesian modeling framework we used has the potential to scale up to learning over more complex representations. Thus, we believe that the tools of probabilistic modeling and information theory can help in understanding how observers form long-term memory representations and use them in working memory. More generally, our data support the view that perceptual encoding rapidly takes advantage of redundancy to form efficient codes.

4.6 Chapter Appendix

4.6.1 Model Results using Self-Information

The self-information at seeing a given item, i, expressed in bits, is:

$$S = -log_2(p_i)$$

More bits of information (S) are gained by seeing items that are of low probability (small pi) than items that are of high probability (large p_i). The number of bits of self-information is the mathematical optimum for how many bits must be required to encode particular stimuli from a given distribution (Shannon, 1948).

In practice, it is difficult or impossible to achieve codes that are exactly equal in length to the self-information for an item, simply because codes must be discrete. Hence, throughout the paper we focused on a particular coding scheme Huffman coding that is both simple and approximates optimal compression. However, it is worthwhile to ask whether we find similar results looking not at the length of the Huffman codes for all the items in a given block, but instead looking at the number of bits of surprise for those items. Thus, we modeled our experiment using surprise to calculate the number of bits for each item rather than the length of the code generated by Huffman coding.

We used the same values for the priors as the Huffman code results in the main text: $\alpha = 34$, and $\alpha = 31$, respectively, for the two experiments. The number of bits of self-information correlate r = -0.94 (Experiment 1) and r = -0.95 (Experiment 2) with human memory performance. Figures 12 and 13 show the results of multiplying the number of bits of surprise with the number of colors remembered by observers for Experiments 1 and 2, respectively. The results once against support the idea of compression as a major factor in visual working memory: observers are able to remember an approximately fixed number of bits, remembering more colors when the items are more redundant.

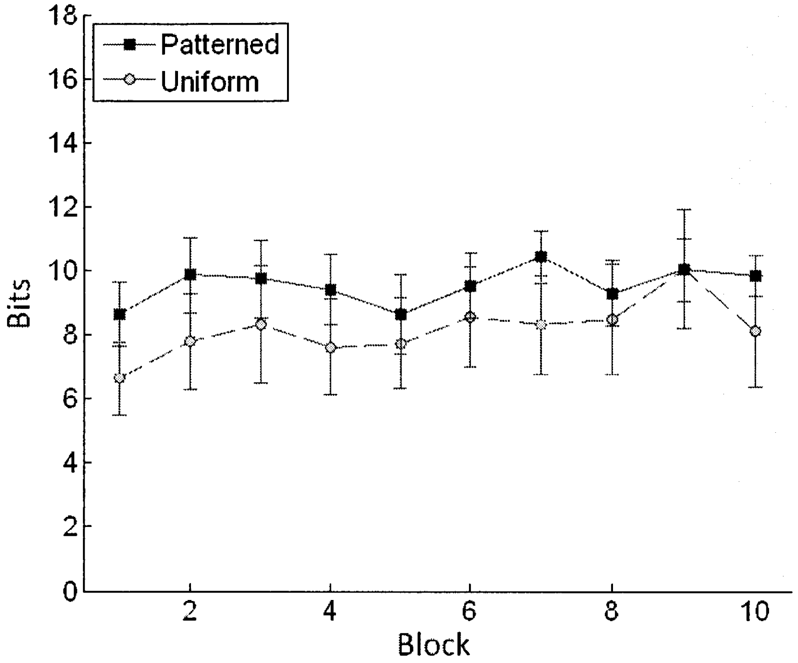


Figure 4-12: The size of memory for Experiment 1 estimated using self-information. Error bars represent +/-1 s.e.m.

4.6.2 Derivation of K formula

In an eight-alternative forced-choice, observers may choose the correct answer for one of two reasons: (1) they may know the correct answer, or, (2) they may guess the correct answer by chance. In order to estimate capacity (the number of items remembered out of the 8 items in the display), we need an estimate of the first kind of correct answers (knowing the colors), discounting the second kind of correct answers (guesses).

To begin deriving such a formula we write percent correct (PC) as a function of the two different kinds of answer answers for those items that observers remember, which they get right 100% of the time, and answers for those items that observers do not remember, which they get right 1/8th of the time. If observers successfully remember K items from a display of 8 items, percent correct (PC) may thus be formulated as:

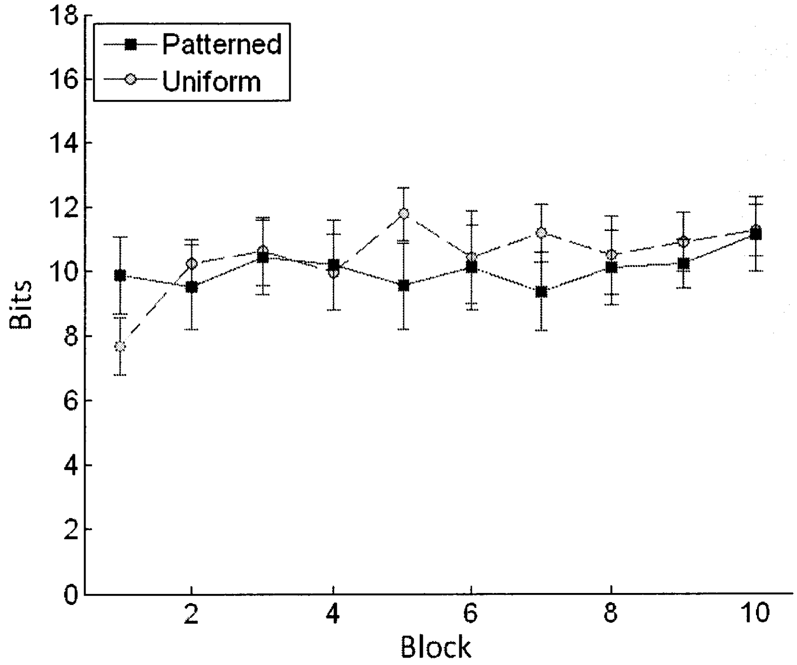


Figure 4-13: The size of memory for Experiment 2 estimated using self-information. Error bars represent +/-1 s.e.m.

$$PC = (\frac{K}{8} * 1) + (\frac{8-K}{8} * \frac{1}{8})$$

Where the first term accounts for items correctly remembered and the second term accounts for items on which the observer guesses. For example, if an observer remembers 2 items (K=2), then for 2/8ths of the items they choose the right answer 100% of the time, whereas the other 6/8ths of the time, they guess and choose the right answer 1/8th of the time. Simplifying and solving for K, we get:

$$(PC * 8 * 8) = 8 * K + 8 - K$$

$$(PC * 8 * 8) - 8 = 8 * K - K$$

$$(PC * 8 * 8) - 8 = K * (8 - 1)$$

$$K = ((PC * 8 * 8) - 8)/7$$

This equation then allows us to directly calculate the capacity of an observer (K) as a function of percent correct (PC).

Chapter 5

General Discussion

How much visual information can a person hold in mind at once? From William James (1890) "primary memory" to current studies of visual working memory capacity, researchers have struggled to understand how much about the visual world we can maintain in memory. The present thesis provides a novel perspective on this question by proposing that observers form rich, structured memory representations, and proposing a framework for modeling such representations.

Chapter 2 demonstrated that observers form hierarchical memory representations in simple working memory displays. In the same way that observers looking at real scenes encode both scene-based information (e.g., it is a blue kitchen) as well as specific items (e.g. that refrigerator), the experiments presented in Chapter 2 scaled up traditional working memory displays to contain patterns, where items are perceptually related to one another. We found that even within a single display observers do not encode items in isolation; instead, they encode both the individual items and the summary statistics of the display, and use the summary statistics to adjust their representation of each individual item. Thus the remembered size of each individual item is biased toward both the mean size of the set of items in the same color, and the mean size of all items in the display. This suggests that visual working memory is constructive, encoding the display at multiple levels of abstraction and integrating across these levels rather than maintaining a veridical representation of each item independently. Furthermore, this pattern of data is compatible with a simple hierarchical Bayesian model where memory representations are stored at multiple levels of abstraction, suggesting that even such structured memory representations can be usefully formalized despite the fact that observers representations are more complex than a simple list of independent items.

Chapter 3 showed that, in addition to using simple summary information like the mean size to help encode specific items, observers may also encode spatial patterns from a display. This kind of higher-order summary information is incompatible with traditional formal models of change detection, like those used to estimate visual working memory capacity (e.g., Cowan, 2001), which assume observers encode only a simple memory representation which includes no higher-order structure and treats items independently from each other. Thus, Chapter 3 presented a probabilistic model of change detection that attempted to bridge this gap by formalizing the role of perceptual organization and allowing for richer, more structured memory representations. Using either standard visual working memory displays or displays in which the dots are purposefully arranged in patterns, we showed that models which take into account perceptual grouping between items and the encoding of higher-order summary information are necessary to account for human change detection performance. Such models can account for observers' performance even on individual displays, whereas models which assume independence between items fail to capture performance even in the simplest displays of colored dots. This demonstrates that items are not encoded independently of each other, and provides a formal framework for understanding this integrative encoding and retrieval.

Chapter 4 examined the influence of learned regularities on visual working memory performance. In the standard visual working memory paradigm, observers are asked to remember a set of arbitrary colored objects, and it is usually found that they can remember only 3 or 4 of the colors (e.g., Luck & Vogel, 1997). This is surprisingly impoverished, and raises the question of how we are able to successfully function in everyday memory tasks. One of the major distinctions between this standard paradigm and real-world tasks is that in the real-world we often have prior knowledge that informs what features we expect to see where in a given scene. Chapter 4 showed that if we introduce regularities into standard visual working memory stimuli, observers not only learn these regularities, but are able to encode the learned items more efficiently in working memory, representing twice as many colors. Furthermore, using an information-theoretic model, we that observers' memory for colors is compatible with having a fixed capacity in terms of information (bits). This provides a theoretical explanation for memory capacity in terms of how compressible the information is in a given display, rather than how many objects can be remembered.

Ultimately, working memory representations in the real-world contain information about scenes that is not purely in the form of a list of independent items contained in those scenes: we make use of our prior knowledge about what items go together, we encode texture, surfaces and other ensemble statistics from scenes, and we make use of the relationships between items to provide information on items we did not specifically encode. Models of working memory need to be capable of dealing with these phenomena in order to provide true insight into the structure of the working memory system and its capacities. In this thesis we have proposed that information is represented at the individual item level as hierarchical feature bundles (Chapter 1), across individual items in terms of ensemble or scene context (Chapter 2; Chapter 3), and that our prior knowledge about regularities between items is crucial to determining the structure of our memory representations (Chapter 4). This thesis thus provides empirical evidence that observers use structured knowledge to represent displays in working memory, and, in addition, provides a set of computational models to formalize these structured memory representations.

References

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19 (6), 716723

Allen, R.J., Baddeley, A.D., & Hitch, G.J. (2006). Is the binding of visual features in working memory resource-demanding? Journal of Experimental Psychology: General, 135, 298313.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. Journal of Experimental Child Psychology, 106(1), 20-9.

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

Alvarez, G.A. (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences, 15(3), 122-131.

Alvarez, G.A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. Psychological Science, 19(4), 392-398.

Alvarez, G.A., & Oliva, A. (2009). Spatial Ensemble Statistics: Efficient Codes that Can be Represented with Reduced Attention. Proceedings of the National

Academy of Sciences, 106, 7345-7350.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in Visual Working Memory Reaches a Stable Plateau When Individual Item Limits Are Exceeded. Journal of Neuroscience, 31(3), 1128-1138.

Anderson, J. R. (1990). The Adaptive Character of Thought. Hillsdale, NJ: Erlbaum.

Anderson, J. R. & Schooler, L. J. (1991). Reflections of the environment in memory. Psychological Science, 2(6), 396-408.

Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving (Ed.), The Oxford handbook of memory (pp. 557-570). New York: Oxford University Press.

Ariely, D. (2001), Seeing sets: Representation by statistical properties. Psychological Science, 12 (2), 157- 162.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and working memory. Trends in Cognitive Sciences, 5(3), 119-126

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

Baddeley, A. D. (1986). Working memory. New York: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? Trends in Cognitive Sciences, 4(11), 417-423

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. Neuropsychologia, 49(6), 1393-1400.

Baddeley, A.D. & Scott, D. (1971) Short-term forgetting in the absence of proactive inhibition. Q. J. Exp. Psychol, 23, 275283.

Baker, C., Olson, C. & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. Psychological Science, 15, 7, 460-466.

Bar, M. (2004). Visual objects in context. Nature Reviews Neuroscience, 5, 617-629.

Barlow, H. B. (1989). Unsupervised Learning. Neural Computation, 1(3), 295-311.

Bartlett, F.(1932). Remembering: A study in experimental and social psychology. NewYork: Macmillan.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. Neuropsychologia, 49(6), 1622-1631.

Bays, P.M. & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. Science, 321 (5890), 851.

Bays, P.M. & Husain, M. (2009). Response to Comment on Dynamic Shifts of Limited Working Memory Resources in Human Vision. Science, 323 (5916), 877d.

Bays, P.M., Catalao, R.F.G. & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(10), 1-11.

Behrmann, M. & Tipper, S.P. (1994). Object-based visual attention: Evidence from unilateral neglect. In C. Umilta & M Moscovitch (Eds), Attention and performance XIV: Conscious and nonconscious processing and cognitive functioning. Cambridge, MA: MIT Press.

Besag, J. (1975). Statistical analysis of non-lattice data. The Statistician, 24(3):179–195.

Besag, J. (1977). Efficiency of pseudo-likelihood estimation for simple Gaussian fields. Biometrika, 64:616–618.

Bower, G. H., Karlin, M. B., & Dueck A. (1975). Comprehension and memory for pictures. Memory and Cognition, 3, 216-220.

Brady, T. F. & Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. Psychological Science, 19(7), 678-685.

Brady, T. F. & Tenenbaum, J.B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. Proceedings of the Cognitive Science Society.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual short-term memory: using statistical regularities to form more efficient memory representations. Journal of Experimental Psychology: General, 138(4), 487-502.

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

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

Brady, T.F., Konkle, T. & Alvarez, G.A. (2011). A review of visual memory capacity: Beyond individual items and towards structured representations. Journal of Vision.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10, 433-436.

Brewer, W.F. & Treyens, J.C. (1981). Role of schemata in memory for places. Cognitive Psychology, 13, 207-230.

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.

Burton, G. J., & Moorehead, I. R., (1987). Color and spatial structure in natural scenes. Applied Optics, 26, 157-170.

Chandler, D.M., and Field, D.J. (2007). Estimates of the Information Content and Dimensionality of Natural Scenes From Proximity Distributions. Journal of the Optical Society of America A, 24(4), 922-941

Chase, W.G., & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 5581.

Chater, N. & Oaksford, M. (1999). Ten years of the rational analysis of cognition. Trends in Cognitive Sciences, 3, 57-65.

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

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. Vision Research, 43, 393-404.

Chong, S. C., & Treisman, A. (2005a). Statistical processing: computing the average size in perceptual groups. Vision Research, 45, 891-900.

Chong, S. C., & Treisman, A. (2005b). Attentional spread in the statistical processing of visual displays. Perception & Psychophysics, 67, 1-13.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology. Learning, Memory, and Cognition, 31(1), 24-39.

Cover, T. M., & Thomas, J.A. (1991). Elements of Information Theory. John Wiley, New York.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24, 87185.

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

Cowan, N., & AuBuchon, A.M. (2008). Short-term memory loss over time without retroactive stimulus interference. Psychonomic Bulletin & Review, 15(1), 230-235. Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. Cognitive Psychology, 51, 42-100.

Cowan, N., Chen, Z., & Rouder, J.N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). Psychological Science, 15, 634-640.

Curby, K. M., & Gauthier. I. (2007). A visual short-term memory advantage for faces. Psychonomic Bulletin and Review, 14(4), 620-628.

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

Daneman, M. & Carpenter, P.A. (1980) Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450466.

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

Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? Visual Cognition, 11(1), 127.

Delvenne, J. F., & Bruyer, R. (2006). A configural effect in visual short-term memory for features from different parts of an object. The Quarterly Journal of Experimental Psychology, 59(9), 15671580.

Droll, J.A., Hayhoe, M.H., Triesch, J. & Sullivan, B.T. (2005). Task demands control acquisition and storage of visual information. Journal of Experimental Psychology: Human Perception and Performance. 31 (6), 1416-1438.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. New York: Teachers College, Columbia University. (Original work published 1885).

Egly, R., Driver, J. & Rafal, R.D. (1994). Shifting Visual Attention Between Objects and Locations: Evidence From Normal and Parietal Lesion Subjects. Journal of

Experimental Psychology: General, 123 (2), 161-177.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. Science, 208, 1181-1182.

Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. Cognition, 107(1), 1-18.

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

Field, D. J. (1987). Relations between the statistics of natural images and response properties of cortical cells. Journal of the Optical Society of America A, 4, 2379-2394.

Fine, M.S. & Minnery, B.S. (2009). Visual salience affects performance in a working memory task. The Journal of Neuroscience, 29(25), 8016.

Fiser, J. Z. & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scene. Psychological Science, 12: 499-504.

Fiser, J. Z., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28(3), 458-467.

Fiser, J. Z., & Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. Journal of Experimental Psychology: General, 134: 521-537.

Fougnie, D. & Alvarez, G.A. (submitted). Breakdown of object-based representations in visual working memory.

Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. Visual Cognition, 17, 48-66.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? Journal of Vision, 10(12), 1-11.

Frank, M. C., Goldwater, S., Mansinghka, V., Griffiths, T., & Tenenbaum, J. (2007). Modeling human performance in statistical word segmentation. Proceedings of the 29th Annual Meeting of the Cognitive Science Society.

Frazor, R.A., Geisler, W.S. (2006) Local luminance and contrast in natural images. Vision Research, 46, 1585-1598.

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General, 108, 316355.

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

Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. Psychonomic Bulletin & Review, 13, 581-587.

Garner, W.R. (1974). The processing of information and structure. Lawrence Erlbaum Associates.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). Bayesian Data Analysis (2nd ed.), London: CRC Press.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Machine Intel., 6, 721741.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C.-H., Jones, G., Oliver, I., & Pine, J.M. (2001). Chunking mechanisms in human learning. Trends in Cognitive Sciences, 5, 236243.

Greene, M.R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. Psychological Science, 20(4), 464.

Greene, M.R., & Oliva, A. (2009). Recognition of Natural Scenes from Global Properties: Seeing the Forest Without Representing the Trees. Cognitive Psychology, 58(2), 137-179.

Greene, M.R., & Oliva, A. (2010). High-Level Aftereffects to Global Scene Property. Journal of Experimental Psychology: Human Perception & Performance, 36(6), 1430-1442

Griffiths, T.L. & Tenenbaum, J.B. (2006). Optimal predictions in everyday cognition. Psychological Science, 17(9), 767.

Haberman, J. & Whitney, D (2007). Rapid extraction of mean emotion and gender from sets of faces. Current Biology, 17(17), R751-53.

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

Haberman, J., & Whitney, D. (2009). The visual system ignores outliers when extracting a summary representation [Abstract]. Journal of Vision, 9(8), 804.

Haberman, J. & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. Attention, Perception, & Psychophysics, 72, 1825-38.

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

Halberda, J., Simons, D. J. & Whetherhold, J. (submitted). Superfamiliarity affects perceptual grouping but not the capacity of visual working memory.

Halberda, J., Sires, S.F., & Feigenson, L. (2006). Multiple spatially-overlapping sets can be enumerated in parallel. Psychological Science, 17 (7), 572-576.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. Cognition, 78(3), B53B64.

Hemmer, P. & Steyvers, M. (2009). Integrating Episodic Memories and Prior Knowledge at Multiple Levels of Abstraction. Psychonomic Bulletin & Review, 16, 80-87.

Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. Journal of Experimental Psychology: Human Perception and Performance, 30, 519537.

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

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. Journal of Experimental Psychology: Human Perception and Performance, 33, 31-47.

Hollingworth, A. (2008). Visual memory for natural scenes. In S. J. Luck & A. Hollingworth (Eds.), Visual Memory (pp. 123-162). New York: Oxford University Press.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. Visual Cognition, 7, 213-235.

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. Memory & Cognition, 31, 930-940.

Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: the relationship between object files and visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 36(3), 543-64.

Hollingworth, A., Hyun, J., & Zhang, W. (2005). The role of visual short-term memory in empty cell localization. Perception & Psychophysics, 67, 1332-1343.

Howe, E. & Jung, K. (1986). Immediate memory span for two-dimensional spatial arrays: Effects of pattern symmetry and goodness. Acta Psychologica, 61(1), 37-51.

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

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

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the Limits of Human Visual Awareness. Science, 317 (5839), 823-825.

Huffman, D.A. (1952). A Method for Construction of Minimum-Redundancy Codes. Proceedings of the I.R.E., 40(9), 1098-1101.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment?. Journal of Experimental Psychology: General, 129(2), 220-241.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization ofVisual-Short Term Memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 26, 683-702.

Johnson, J. S., Hollingworth, A., & Luck, S. J. (2008). The role of attention in the maintenance of feature bindings in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance, 34, 41-55.

Johnson, J. S., Spencer, J., Luck, S., Schner, G. (2009) A Dynamic Neural Field

Model of Visual Working Memory and Change Detection. Psychological Science 20: 568-577.

Kane, M.J., Bleckley, M.K., Conway, A.R.A. & Engle, R.W. (2001) A controlled-attention view of working-memory capacity. Journal of Experimental Psychology: General, 130, 169183.

Kemp, C. and Tenenbaum, J. B. (2008). The discovery of structural form. Proceedings of the National Academy of Sciences. 105(31), 10687-10692.

Kirkham, N. Z., Slemmer, J. A., Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. Cognition, 83, 35-42.

Kleinberg, J. & Kaufman, H. (1971). Constancy in short-term memory: bits and chunks. Journal of Experimental Psychology, 90(2), 326-333.

Knill, D.C. & Richards, W. (1996). Perception as Bayesian inference. Cambridge University Press.

Koffka, K. (1935). Principles of Gestalt Psychology. Hartcourt: New York.

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

Kubovy, M. & Van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. Psychological Review, 115(1), 131154 .

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the Lawfulness of Grouping by Proximity. Cognitive Psychology. 35:7198

Lampinen, J.M., Copeland, S. & Neuschatz, J.S. (2001). Recollections of things schematic: Room schemas revisited. Journal of Experimental Psychology: Learning, Memory and Cognition, 27, 1211-1222.

Li, S.Z. (1995). Markov Random Field Modeling in Computer Vision. Springer: Secaucus, NJ.

Lin, P.-H., & Luck, S. J. (2008). The influence of similarity on visual working memory representations. Visual Cognition, 17, 356-372.

Logie, R. H., Brockmole, J. R., & Vandenbroucke, A. R. E. (2009). Bound feature combinations are fragile in visual short-term memory but form the basis for long-term learning. Visual Cognition, 17, 375-390.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. Nature, 390, 279281.

Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. Journal of Experimental Psychology: Human Perception and Performance, 22(1), 202-212.

Mandler, J. M., & Johnson, N.S. (1976). Some of the thousand words a picture is worth. Journal of Experimental Psychology: Human Learning and Memory, 2, 529-540.

Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. Journal of Experimental Psychology: Human Learning and Memory, 2, 38-48.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: W. H. Freeman and Company.

Mattingley, J.B., Davis, G. & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. Science, 275 (5300), 671-674.

Melcher, D. (2001). Persistence of visual memory for scenes. Nature, 412, 401.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. Journal of Vision, 6(1), 8-17.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 63, 81-97.

Miller, M. B. & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. Neuropsychologia, 36, 513520.

Miyake, A. & Shah, P. (1999). Models of working memory: Mechanisms of active maintenance and executive control. Cambridge University Press.

Neisser, U. (1967). Cognitive psychology. Englewood Cliffs, NJ: Prentice-Hall.

Oliva, A. (2005). Gist of the scene. In the Encyclopedia of Neurobiology of Attention. L. Itti, G. Rees, and J.K. Tsotsos (Eds.), Elsevier, San Diego, CA (pages 251-256).

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

Olshausen, B.A., and Field, D.J. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. Nature, 381, 607-609

Olson, I.R. & Marshuetz, C. (2005). Remembering "What" Brings Along "Where" in visual working memory. Perception & Psychophysics, 67(2): 185-194.

Olson, I.R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the strong-object hypothesis. Perception & Psychophysics, 64, 10551067

Olson, I.R., & Jiang, Y. (2004). Visual short-term memory is not improved by training. Memory and Cognition, 32, 13261332.

Olson, I.R., Jiang, Y., Moore, K.S. (2005). Associative learning improves visual working memory performance. Journal of Experimental Psychology: Human Perception and Performance, 31, 889900.

Olsson H., Poom L. (2005). Visual memory needs categories. Proceedings of the National Academy of Sciences, USA, 102, 87768780.

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

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. Nature Neuroscience, 4, 739744.

Pashler, H. (1988). Familiarity and the detection of change in visual displays. Perception & Psychophysics, 44, 369378.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437-442.

Phillips, W.A. (1974). On the distinction between sensory storage and short-term visual memory. Perception & Psychophysics, 16(2), 283290.

Quinlan, P. T., & Cohen, D. J. (2011). Object-based representations govern both the storage of information in visual short-term memory and the retrieval of information from it. Psychonomic Bulletin & Review, 18(2), 316-323.

Rosenholtz, R. & Alvarez, G. A. (2007) How and why we perceive sets: What does modeling tell us?. Perception, 36, ECVP Abstract Supplement.

Rosenholtz, R., Twarog, N.R., Schinkel-Bielefeld, N. & Wattenberg, M. (2009). An intuitive model of perceptual grouping for HCI design. Proceedings of the 27th International Conference on Human Factors in Computing Systems, 1331-1340.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory.

Proceedings of the National Academy of Sciences, USA, 105(16), 5975-5979.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274, 19261928.

Sanocki, T., Sellers, E., Mittelstadt, J., & Sulman, N. (2010). How high is visual short term memory capacity for object layout? Attention, Perception, & Psychophysics.

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

Scholl, B.J., Pylyshyn, Z.W. & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. Cognition, 80(1/2), 159 - 177.

Schreiber, E., & Griffiths, T. L. (2007). Subjective randomness and natural scene statistics. In D. S. McNamara & J. G. Trafton (Eds.), Proceedings of the 29th Annual Conference of the Cognitive Science Society (pp. 1449-1454). Mahwah, NJ: Erlbaum.

Scolari, M., Vogel, E.K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. Psychonomic Bulletin & Review, 15 (1), 215-222

Sebrechts, M.M. & Garner, WR (1981). Stimulus-specific processing consequences of pattern goodness. Memory & Cognition, 9(1), 41-49.

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379423.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving Effectively from Memory. Psychonomic Bulletin & Review, 4 (2), 145-166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.). Rational models of cognition. (pp. 73-95), Oxford, England: Oxford University Press.

Simon, H.A. (1974). How Big Is a Chunk? Science, 183, 482-488

Spelke, E.S. (1990). Principles of object perception. Cognitive Science, 14, 29-56.

Spencer, J.P. & Hund, A.M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. Journal of Experimental Psychology: General, 131(1), 16-36.

Sperling, G. (1960). The information available in brief visual presentations. Psychological Monographs, 74, 1-29.

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

Stirk, J.A. & Underwood, G. (2007). Low-level visual saliency does not predict change detection in natural scenes. Journal of Vision, 7(10), 3.

Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10(7), 309-318.

Triesch, J J. Ballard, D. Hayhoe, M. & Sullivan, B. (2003). What you see is what you need. Journal of Vision, 3, 86-94.

Turk-Browne, N. B., Jung, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. Journal of Experimental Psychology: General, 134, 552-564.

Victor, J.D. & Conte, M.M. (2004) Visual working memory for image statistics. Vision Research, 44(6), 541-556.

Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., & O'Regan, J. K. (2005). Relational information in visual short-term memory: The structural gist. Journal of Vision, 5(3):8, 244-256.

Viswanathan, S., Perl, D.R., Visscher, K.M., Kahana, M.J., & Sekuler, R. (2010) Homogeneity computation: How inter-item similarity in visual short term memory alters recognition. Psychonomic Bulletin & Review, 17, 59-65.

Vogel, E.& Awh, E. (2008). How to exploit diversity for scientific gain: Using individual differences to constrain cognitive theory. Current Directions in Psychological Science.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2001). Storage of features, conjunctions, and objects in visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 27, 92114.

Watson, S.E., & Kramer, A.F. (1999). Object-based visual selective attention and perceptual organization. Perception & Psychophysics , 61, 31-49.

Wertheimer, M. (1938). Laws of organization in perceptual forms. Harcourt Brace Jovanovich: London.

Wheeler, M.E., & Treisman, A.M. (2002). Binding in short-term visual memory. Journal of Experimental Psychology: General, 131, 4864

Wilken, P., Ma, W.J. (2004) A detection theory account of change detection. Journal of Vision, 4, 11201135.

Wiseman, S. & Neisser, U. (1974). Perceptual Organization as a Determinant of Visual Recognition Memory. The American Journal of Psychology, 87 (4), 675-681.

Wood, J. N. (2009). Distinct visual working memory system for view-dependent and view-invariant representation. PLoS One, 11, 4(8), e6601.

Wood, J. N. (2011a). A core knowledge architecture of visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 37(2), 357-81.

Wood, J.N. (2011b). When do spatial and visual working memory interact? Attention, Perception and Psychophysics, 73(2), 420-39.

Woodman, G.F. & Vogel, E.K. (2008). Top-down control of visual working memory consolidation. Psychonomic Bulletin & Review, 15, 223-229.

Woodman, G.F. Vecera, S.P., & Luck, S.J. (2003). Perceptual organization influences visual working memory. Psychonomic Bulletin & Review, 10, 80-87.

Wright, M.J. (2005). Saliency predicts change detection in pictures of natural scenes. Spatial Vision, 18(4), 413-430.

Xu, Y. & Chun, M. M. (2006). Encoding objects in visual short-term memory: The roles of location and connectedness. Perception & Psychophysics, 68, 815-828.

Xu, Y. & Chun, M.M. (2007). Visual grouping in human parietal cortex. Proceedings of the National Academy Of Sciences, 104(47), 18766.

Xu, Y. (2002a). Encoding color and shape from different parts of an object in visual short-term memory. Perception & Psychophysics, 64, 1260-1280.

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

Yuille, A. & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis?. Trends in Cognitive Sciences, 10(7), 301-308.

Zhang, W. & Luck, S.J. (2008). Discrete fixed-resolution representations in visual working memory. Nature, 452, 233-235.

Zosh, J. M., & Feigenson, L. (2009). Beyond 'what' and 'how many': Capacity, complexity, and resolution of infants' object representations. In B. Hood & L. Santos (Eds.), The Origins of Object Knowledge (pp. 25- 51). New York: Oxford University Press.