Structured Representations in Visual Working Memory

Timothy F. Brady

Massachusetts Institute of Technology

Introduction

Working memory allows us to actively work with and manipulate information. Moreover, individual differences in working memory capacity predict general intelligence, reading comprehension and ability to learn (Baddeley, 1986; Baddeley, 2000). However, there seem to be striking limits on working memory capacity: in visual working memory tasks, observers have difficulty remembering the color of five circles for a second or two (Luck & Vogel, 1997), in spite of the fact that they can easily perceive all of the colors when the display is present. Many theories have been proposed to explain these severe capacity limits: for example, it has been proposed that our capacity is slot-based and limited to representing only 3 or 4 integrated objects (Luck & Vogel, 1997), or that we have separate limits of 3 or 4 slots for each visual feature, such as colors and orientations (Wheeler & Treisman, 2002; Xu, 2002).

My dissertation proposes a new model of visual working memory based on the interaction of structured, hierarchical representations rather than independent representations of individual items, chunks or slots. Using psychophysical data from human observers combined with computational modeling, I provide evidence that observers represent each visual display with at least three different levels of abstraction: independent visual features, integrated objects, and ensemble or summary statistics across sets of items (Chapter 1). Furthermore, I show that the levels are not independent, but instead interact with each other to form a combined memory representation (Chapter 2), and I provide a computational model that formalizes how individual item representations might interact with the representation of summary statistics (Chapter 3). In addition, I demonstrate that statistical learning can influence our representations of individual items in working memory displays (Chapter 4), resulting in another form of structured memory representations. I formalize such learning using information theory, showing that working memory limits are best characterized as information limits, not limits on numbers of objects or features. Taken together, my thesis demonstrates that working memory capacity cannot be understood in terms of the number of items, chunks or objects remembered. In order to understand the capacity limits of working memory, we need to take into account the underlying hierarchical structure of our memory representations.

Throughout graduate school my research has had an interdisciplinary focus, involving the intersection of vision science, computer vision and human learning and memory. I have used methods ranging from psychophysics (Brady & Oliva, submitted; Brady & Alvarez, 2011), to cognitive paradigms designed to tap episodic memory (Brady et al. 2008; Konkle, Brady et al.

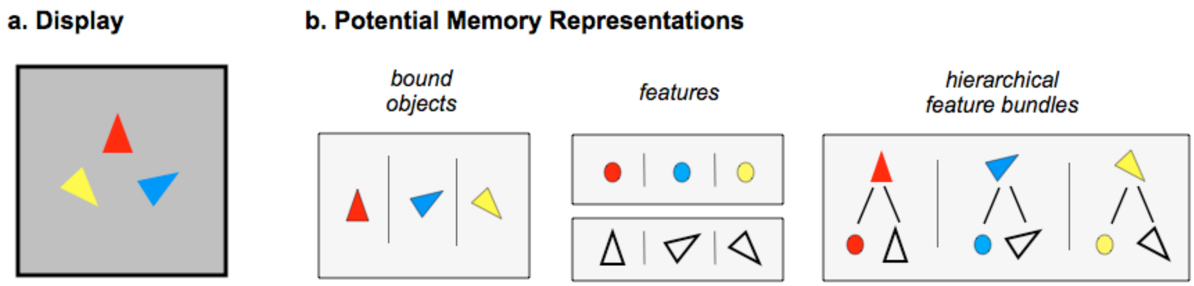

Figure 1. Possible memory representations for a visual working memory display. (a) A display of oriented and colored items to remember. (b) Potential memory representations for the display in (a). The units of memory do not appear to be integrated bound objects, or completely independent feature representations. Instead, we propose that they be characterized as hierarchical feature-bundles, which have both object-level and feature-level properties.

2010a, 2010b), to functional neuroimaging of visual scene and object recognition (Park, Brady et al. 2011). I have applied computational techniques ranging from connectionist modeling (Brady & Chun, 2007) to hierarchical Bayes (Brady & Alvarez, 2011; Brady & Tenenbaum, 2010) to information theory (Brady et al. 2009). In addition, I've sought a range of interdisciplinary collaborations, for example, collaborating in applying machine learning techniques to research at the intersection of animal cognition and music cognition (Schachner, Brady et al. 2009).

My thesis represents this interdisciplinary focus well. The three main components of my thesis involve collaborations not only with my advisor, Dr. Aude Oliva, but also Dr. Josh Tenenbaum at MIT, in work formalizing a structured model of visual working memory; and Dr. George Alvarez at Harvard, in work using psychophysics to examine the content of working memory representations. The contributions of my thesis are also interdisciplinary: It uses tools from machine learning, information theory and Bayesian statistics to describe phenomena in human cognition and memory, and it proposes a theoretical picture of working memory capacity and architecture that is both formally specified as well as strongly related to the known neuroscientific basis of visual representation.

Chapter 1: Thesis Background / Structured Representations in Visual Working Memory
(*published as Brady, Konkle & Alvarez, 2011)*

Estimates of memory capacity must be expressed with some unit; however, the appropriate unit for memory capacity 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). In addition to providing an overview of the background literature on visual working memory, Chapter 1 of my thesis proposes a novel theory for how individual items are represented in working memory: the hierarchically structured feature-bundle.

This model resolves an apparent contradiction in the existing literature, which had suggested limits at the level of both integrated objects and at the level of separate features. For example, Luck and Vogel (1997) 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), suggesting that the limitation on memory was how many objects can be stored, rather than how many features composed those objects. By contrast, when Fougnie, Asplund and Marois (2010) examined observers' representations of multi-feature objects (oriented triangles of different colors), they found that even within a single simple object, remembering more features results in significant costs in the fidelity of each feature representation. In other words, remembering the color and orientation of each triangle resulted in a less precise representation of both the color and orientation than remembering only a single dimension. These data provide strong evidence against slot-based models, since more information about an already-represented object cannot always be encoded without cost (e.g., Luck and Vogel, 1997); but at the same time argue against the idea of entirely separate memory capacities for each feature dimension.

So what *is* the basic unit of representation in visual working memory – visual features or integrated objects? No existing model captures all of the relevant data on the storage of objects and features in working memory. We propose 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 1). This "unit" is composed of an integrated object representation at the top level, and low-level feature representations at the bottom level, with these representations hierarchically 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. This structured representation accounts for the previous apparent contradiction: 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 would 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 from the visual working memory literature (see Chapter 1) and 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).

It is important to note that our proposed 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 (such as features). Consequently, chunk-based 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). Any limit in terms of numbers of items or chunks would not capture the hierarchically structured content of the representations maintained in memory.

Chapter 2: Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items *(published as Brady & Alvarez, 2011, Psychological Science)*

In addition to storing information about individual items and features, observers can quickly and accurately compute *ensemble* statistics about a display, like the mean size of the items (Ariely, 2001; Chong & Treisman, 2003), mean facial expression (Haberman & Whitney, 2009), mean orientation (Parkes et al. 2001), mean location (Alvarez & Oliva, 2008), and even higher-level spatial layout statistics (Alvarez & Oliva, 2009). However, little work has explored how these statistics are used, 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; Zhang & Luck, 2008).

Contrary to the assumptions of previous models of visual working memory, we propose that items are not represented independently, even in working memory displays: Instead, ensemble statistics allow observers to encode 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 size). Such hierarchical encoding would allow observers to represent information about every item in the display simultaneously,

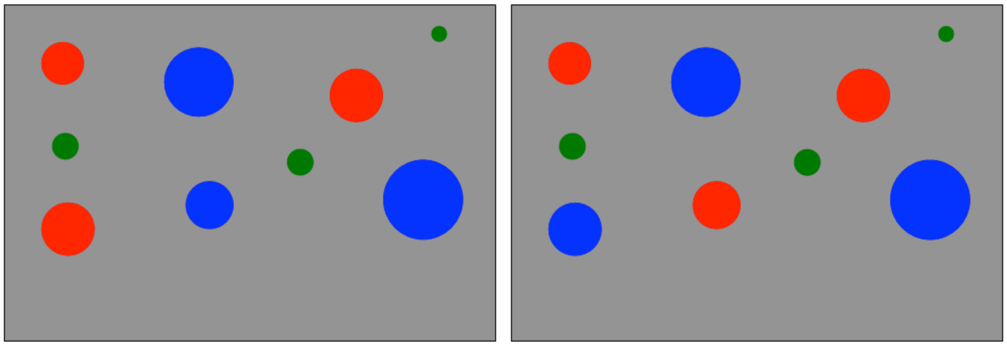

Figure 2. An example pair of matched displays from Chapter 2. Observers had to remember the size of the red and blue dots and ignore the green dots. After each trial they were tested on the size of a single dot using the method of adjustment (a continuous report procedure).

significantly improving the fidelity of their memory representations compared to encoding only 3-4 individual items.

To test this hypothesis, we gave observers a memory task in which they had to remember the size of multiple red and blue dots (Figure 2). After each trial, observers were tested on the size of a single dot using the method of adjustment (a continuous report procedure). We asked whether the ensemble statistics of a display would bias memory for individual items in this task. We hypothesized that on displays with 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 size bias would suggest that observers had taken into account the mean size of the set of same-colored items in addition to the size of the individual item when reporting its size.

To allow a direct test for this mean set bias, we generated displays in matched pairs. Thus, we were able to compare reported size for an identical item when it was one color with the reported size when it was another color (see Figure 3).

We found that 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 ($t(20)=4.17$; $p=0.0004$). Furthermore, a series of control experiments revealed that: this bias was present on each individual trial; there was a bias not only towards the mean size of the same colored items but also towards the mean size of all of the items on the display; and the data was compatible with a simple hierarchical Bayesian model where information is stored at multiple levels of abstraction and integrated to form a final memory representation.

Thus, we find that observers are biased by the ensemble statistics of the display when representing items in visual working memory. 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

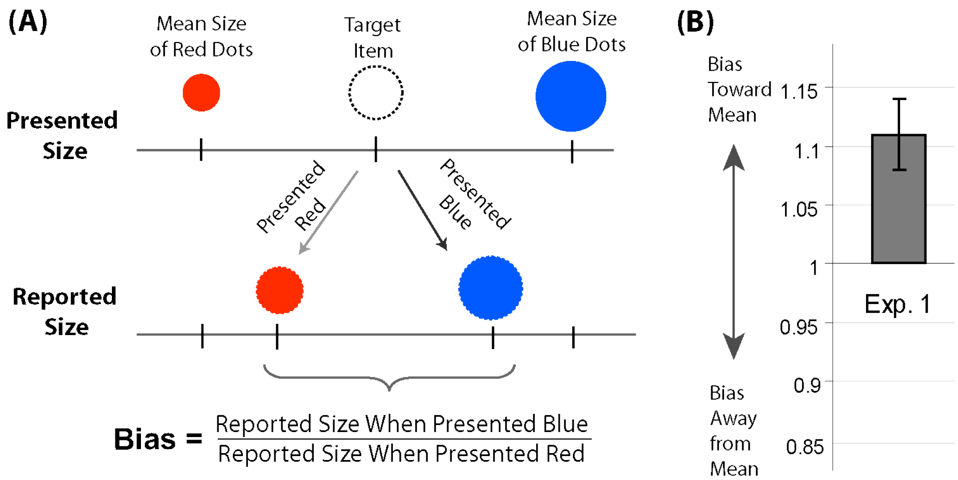

Figure 3. (A) Schematic illustration of calculating bias from a pair of matched displays. In this example, the blue dots were larger than the red dots on average. We then measured whether observers reported different sizes for the tested dot when it was red versus when it was blue (the dot was in fact the same size in both presentations). (B) The bias observed in Experiment 1, indicative of observers using ensemble statistics.

(Bartlett, 1932; Lampinen, Copeland, Neuschatz, 2001); and working memory representations must be thought of as based on a structured representation consisting of both ensemble and individual item information.

Chapter 3: A probabilistic model of visual working memory: Incorporating higher-order regularities into working memory capacity estimates *(parts published as Brady & Tenenbaum, 2010, CogSci proceedings)*

The standard way of reporting performance in visual working memory tasks is to report the number of colors remembered ("Cowan's K"). These values are calculated using a particular model of change detection (a 'slot model'), which assumes that the decline in performance when more colors 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.

Nearly all visual working memory papers report such values, often without considering whether the model that underlies them is an accurate description of observers' representations 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 are not encoding items independently.

In this chapter 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 in the test display and the information in memory to a decision about the initial display, and 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), and make predictions about memory capacity when items are non-independent or where summary statistics are encoded in addition to specific items.

Across a series of experiments, we find that observers are much more successful at detecting changes to displays containing spatial patterns than would be expected if they were remembering only 3-4 individual items from each display. Furthermore, we find that observers are highly reliable in which particular changes they find easiest and hardest to detect. Using our Bayesian model, we formalize a model of observers' memory representations: a summary-based representation, where observers encode both a spatial summary of the display (e.g. "items tend to be the same color as their horizontal neighbors") plus specific outlier items, combined with a simple perceptual grouping model. Using this change detection model, we can explain nearly all of the variance in change detection performance in patterned displays as well as in simple displays of colored dots.

Thus, modeling more structured memory representations allows us to successfully understand what information observers represent in visual working memory, and predict for the first time what particular working memory displays will be easy or difficult for observers to remember.

Chapter 4: Compression in visual working memory: Using statistical regularities to form more efficient memory representations *(published as Brady, Konkle, & Alvarez, 2009, Journal of Experimental Psychology: General)*

Observers are highly sensitive to statistical regularities in the world, and this sensitivity has been used to explain effects from speech segmentation to the emergence of visual objects (Saffran, Aslin & Newport, 1996; Turk-Browne, Isola, Scholl, & Treat, 2008). However, such regularities also provide an opportunity for memory systems to form more efficient representations by eliminating redundancies. According to information theory, 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 (Cover & Thomas, 1991).

We asked whether human observers use such regularities in working memory in a way that is compatible with an information theoretic compression analysis. To do so, we combined a visual statistical learning paradigm with a working memory paradigm. On each trial, 8 colors were presented, grouped in pairs (see Figure 4). After a delay, one location was cued and observers reported the color that had been at the cued location. For one group of observers (the "uniform

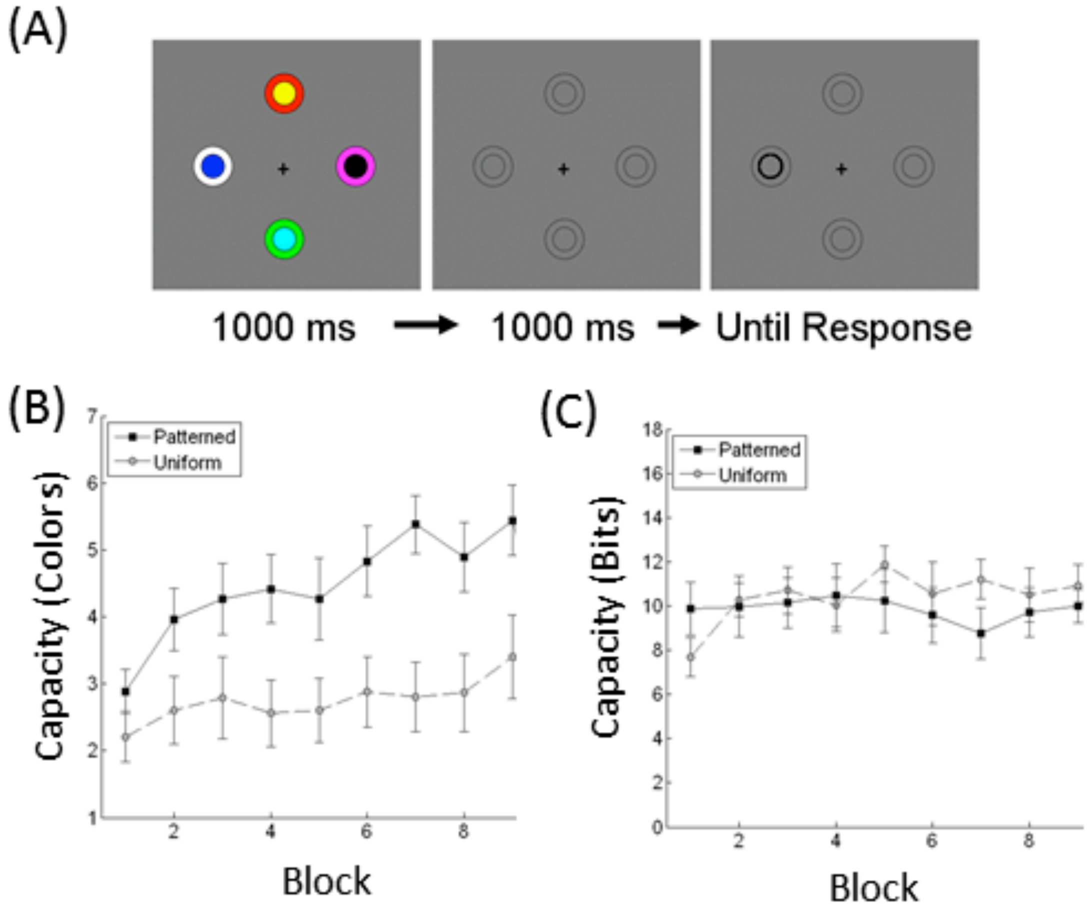

Figure 4. (*A*) A sample trial. (*B*) Results of Experiment 1. In the patterned condition some colors appeared on the same object more frequently than others. Observers' performance increased over time as they learned these regularities. Capacity (K) is the number of colors observers successfully recalled. (*C*) The size of memory estimated in bits, rather than number of objects. Note that the amount of information remembered is constant, even though the number of colors remembered increases.

group"), the colors were randomly paired from trial to trial. For a separate group of observers (the "patterned group"), the colors were paired such that 80% of the time certain colors co-occurred (e.g., red with yellow).

We found that, over the course of the experiment, the number of colors remembered by the patterned group doubled from 3 to 6, but remained constant for the uniform group at about 3 (see Figure 5b). Additional analyses ruled out the possibility that the improved performance for the patterned group was simply due to a guessing strategy. Instead, it appeared that observers in the patterned group learned the regularities and were able to use them to store more colors in working memory.

To formalize the hypothesis that learning visual statistical regularities enables efficient compression, we performed an information theoretic analysis of the data. First, we modeled the

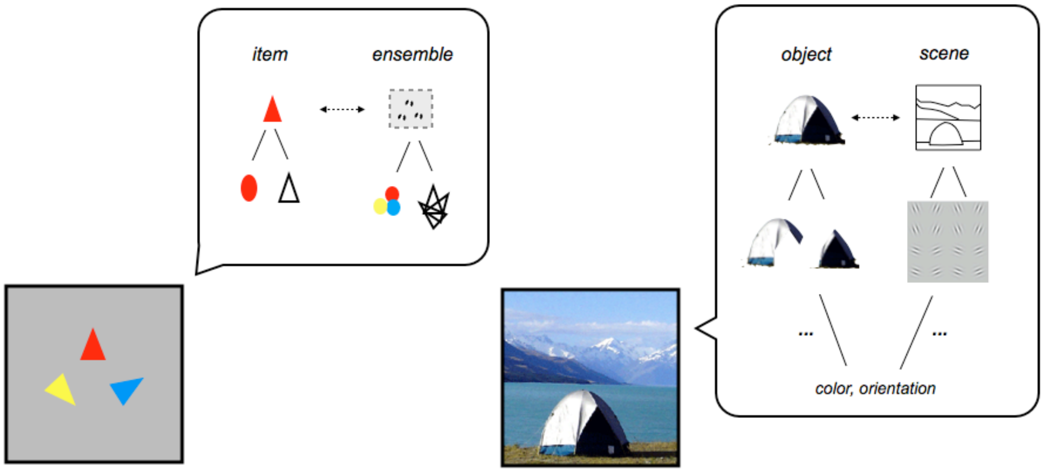

Figure 6. The proposed structure of memory representations in both simple and real-world displays. Information is represented both at the item level (as a hierarchical feature-bundle) and across individual items at the ensemble level, possibly in parallel but interacting processing streams.

learning of the color regularities based on the number of times observers saw each pair of colors using a Bayesian learning model (a Dirichlet-Multinomial model). Second, we assessed how these learned statistics translate into representations in bits, using Huffman coding (Huffman, 1952). The Huffman codes provide a measure of the average number of bits per color, and the memory performance gives a measure in number of colors remembered. Thus, if we multiply the average size of 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 4c). Both groups of observers in the *uniform* condition and the *patterned* condition show roughly the same total capacity in bits, despite the difference in the number of items remembered between the groups. This analysis suggests that observers in the patterned group were able to use an optimal encoding scheme, taking advantage of the regularities between items to compress information. Importantly, this modeling approach provides a formal framework that makes specific, testable predictions for how many items can be stored in working memory, given the particular set of regularities encountered, and provides a formal specification of when observers will and will not successfully form 'chunks' from a display.

Almost all real-world items are associated with other objects in the environment, and we frequently have long-term knowledge that allows us to constrain the information we need to remember. Thus, taking into account our prior knowledge about regularities between items is crucial to determining the structure of our memory representations.

Conclusion

Taken together, my thesis demonstrates that working memory capacity cannot be understood in terms of the number of items, chunks or objects that can be remembered. Instead, we need to study working memory capacity by taking into account the hierarchical structure of our memory representations. I have shown that observers simultaneously represent visual displays with at least three different levels of abstraction: independent visual features, integrated objects, and ensemble or summary statistics across sets of items, and that these levels of representation interact to form the memory representation of the entire display.

My thesis draws inspiration not only from computer vision and machine learning, but also from the neuroscience of the visual system, including the basic visual hierarchy (e.g., Riesenhuber & Poggio, 1999; Ullman, 2007) and the distinction in the behavioral and neuroimaging literatures between object processing and scene processing regions (e.g., Epstein & Kanwisher, 1998; Grill-Spector et al. 2001; Torralba et al. 2006; Oliva, 2005), including the known involvement of scene processing regions in representing summary statistics of sets of objects (e.g., Cant & Xu, 2010). Combined with the fact that improved understanding of working memory capacity is relevant to a broad range of research areas, I believe that my thesis aligns well with the principles of the Robert J. Glushko Dissertation Prize in Cognitive Science.

References

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

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

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.

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

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press

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

Barlett, F.C. (1932). *Remembering*. University Press.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567-575.

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. and Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cueing. *Journal of Experimental Psychology: Human Perception & Performance*, 33(4), 798-815.

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

Brady, T. F., Konkle, T., & 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., and Oliva, A. (submitted). Spatial frequency integration during active perception: Perceptual hysteresis when an object recedes.

Brady, T. F. & Tenenbaum, J.B. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In S. Ohlsson & R.Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 411-416). Austin, TX: Cognitive Science.

Cant, J.S. & Xu, Y. (2010). The flipside of object individuation: Neural representation for object ensembles. Journal of Vision, 10(7): 111.

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

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, 87–185.

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

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.

Cowan, N., & AuBuchon, A.M. (2008). Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review*, 15(1), 230-235.

Epstein, R. & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392. 598–601.

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.

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

Frank, M.C. & Barner, D. (2011). Representing Exact Number Visually Using Mental Abacus. *Journal of Experimental Psychology: General*.

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.

Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11), 1409-1422.

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.

Harrison, S.A. & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632-635.

Jeneson, A. & Squire, L.R. (2012). Working memory, long-term memory, and medial temporal lobe function. *Learning & Memory*, 19(1), 15-25.

Kibbe, M. (2011). Charting the boundaries of infant working memory. Doctoral thesis, Rutgers.

Konkle, T., Brady, T. F., Alvarez, G. A. and Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558-78.

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

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.

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

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.

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).

Park, S., Brady, T. F., Greene, M.R., and Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the PPA and LOC in representing real-world scenes. *Journal of Neuroscience*, 31(4), 1333-1340.

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

Pashler, H. (1988). Familiarity and the detection of change in visual displays. *Perception & Psychophysics*, 44, 369–378.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Schachner, A., Brady, T. F., Pepperberg, I. and Hauser, M. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, 19(10), 831-836.

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.

Torralba, A., Oliva, A., Castelhano, M.S. & Henderson, J.M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766.

Turk-Browne, N.B., Isola, P.J., Scholl, B.J. & Treat, T.A. (2008). Multidimensional visual statistical learning. *Journal Of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 399.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences,* 11(2), 58–64.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92–114.

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

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.

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

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