Encoding higher-order structure in visual working memory: A probabilistic model

Timothy F. Brady (tfbrady@mit.edu), Joshua B. Tenenbaum (jbt@mit.edu)
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

Abstract

When encoding a scene into memory, people store both the overall gist of the scene and detailed information about a few specific objects. Moreover, they use the gist to guide their choice of which specific objects to remember. However, formal models of change detection, like those used to estimate visual working memory capacity, generally assume people represent no higher-order structure about the display and choose which items to encode at random. We present a probabilistic model of change detection that attempts to bridge this gap by formalizing the encoding of both specific items and higher-order information about simple working memory displays. We show that this model successfully predicts change detection performance for individual displays of patterned dots. More generally, we show that it is necessary for the model to encode higher-order structure in order to accurately predict human performance in the change detection task. This work thus confirms and formalizes the role of higher-order structure in visual working memory.

Keywords: change detection; visual short-term memory; working memory; probabilistic model

Introduction

Working memory capacity constrains cognitive abilities in a wide variety of domains (Baddeley, 2000), including general intelligence and reading comprehension (Daneman & Carpenter, 1980). 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 (Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Cowan, 2001; Luck & Vogel, 1997; Wilken & Ma, 2004; Zhang & Luck, 2008). However, nearly all of these models focus on memory for extremely simple displays of presegmented objects. Furthermore, these models address only average performance across displays and do not make predictions about the difficulty of particular displays.

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, 2004). Moreover, they use the gist to guide their choice of which specific objects to remember (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 (Lampinen, Copeland, & Neuschatz, 2001). Existing models of the architecture of working memory do not address any of these hierarchical encoding or perceptual grouping factors. For this reason, they are unsatisfying as explanations of what observers will remember about more complex displays in which objects are not randomly chosen, but instead make up a coherent scene.

In this paper we reformulate change detection as rational probabilistic inference in a generative model (similar in spirit to Huber, Shiffrin, Lyle, and Ruys (2001) and Hemmer and Steyvers (2009b)). Rather than modeling the memory process per se, we model how observers encode a scene, and treat change detection as a probabilistic inference that attempts to invert this encoding model. We show that earlier models of visual working memory capacity are special cases of this framework, and show how our model can be extended to include the encoding of gist or higher-order structure. We thus take the first steps toward formalizing working memory capacity for displays in which the items are not all treated independently.

Visual working memory

One of the most popular ways to examine visual working memory capacity has been with a *change detection* task (Luck & Vogel, 1997). In this task, observers are presented with a small number of different colored squares (2, 4, 8, or 16) and told to remember which color appeared in which location. The squares then disappear for a brief period, and when they reappear they either are all the same colors as before, or contain one square which has changed color. Observers must report whether the display is the same or whether one of the squares changed color.

It is generally found that observers accurately detect changes when there are fewer than 3-4 simple colored squares, and as the number of squares increases above 4 observers accuracy steadily decreases (Luck & Vogel, 1997). In order to quantify this decrease and derive a capacity measure for the contents of visual working memory, change detection tasks have been modeled and formalized (Rouder et al., 2008; Wilken & Ma, 2004). For example, in the standard "slot" model of visual working memory (Cowan, 2001; Luck & Vogel, 1997; Rouder et al., 2008), it is assumed that on a display with N items observers perfectly recall the color of K items and completely forget the other N-K items on the display. Using this model, it is possible to convert change detection performance into an estimate of K, and these capacity estimates, termed Cowan's K, are widely reported in the literature on visual working memory (Alvarez & Cavanagh, 2004; Brady, Konkle, & Alvarez, 2009; Cowan, 2001; Luck & Vogel, 1997).

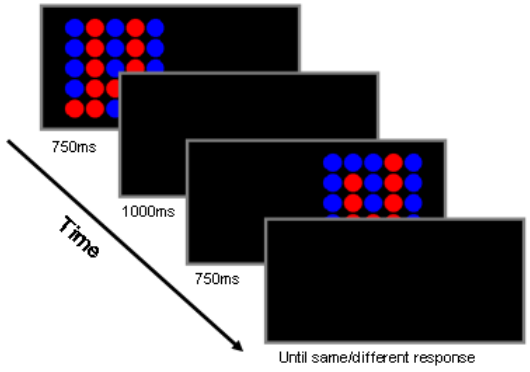

Figure 1: Methods of the change detection task modeled and used in Experiments 1 and 2. Observers are first briefly presented with a display, and then after a 1 sec blank are presented with another display where either the items are exactly the same or one item has changed color. They must say whether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

Aside from Cowan's K, there are other models used to quantify working memory capacity (Bays et al., 2009; Wilken & Ma, 2004; Zhang & Luck, 2008). However, the displays used always consist of simple stimuli like colored dots that are sampled uniformly, limiting any overarching structure or gist. All existing models of change detection thus ignore the presence of higher-order structure and prior knowledge that characterize change detection in real-world scenes (Simons & Rensink, 2005).

A probabilistic model of change detection

We present a probabilistic model of change detection that attempts to bridge the gap between the simple models used to formalize working memory capacity and the complicated phenomena that characterize memory for real-world scenes. Thus, we sought to model change detection in cases where the displays to be remembered were not just random colored dots but also exhibited some higher-order structure. As stimuli we created 5x5 patterns in which each space was filled in by a red or blue circle (or black and white square, see Figure 3). The items could form patterns that were anything from completely random to completely one color or vertical or horizontal lines. Our displays were thus simple relative to real scenes but were complex enough that we expected existing models, which encode dots at random, would fail to predict what people remember about these displays.

Our modeling preceded in two stages, mirroring the two stages of a standard change detection task: view and encode display one, then view display two and decide if a change occurred (See Figure 1).

While the observer is encoding the first display, they have access to the color of all the dots present in the first display. We propose that observers use this information to do two things: first, they infer what "gist" may have given rise to this display; then, using this gist, they select the subset of the dots least well captured by the gist and encode these items specifically into an item memory. The specific dots to encode are selected based on how unlikely they are under the gist. Those that are the biggest outliers (e.g., least well captured by the gist) are encoded into an item memory that specifically encodes their colors.

After a short viewing, the first display disappears and the observer is left with only what they encoded about it in memory. Then, some time later, a 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. Thus, at the time of the second display (detection), the observer has access to the new display and the information 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 each possible first display, and, using these probabilities, to calculate the likelihood of that the display changed.

Importantly, when the model encodes no higher-order structure it recovers the standard slot-based model of change detection. However, when the displays do have higher-order regularities or 'gist', our model uses this information to both select appropriate individual items to remember and to infer properties of the display that are not specifically encoded.

Encoding

The graphical model representation of the encoding model (shown in Figure 2) specifies how the stimuli are initially encoded into memory. We observe the first image (D^1), and we use this to both infer the higher-order structure that may have generated this image (G) and to choose the specific set of K items to remember from this image (S).

In the model, any given "gist" must specify which displays are probable and which are improbable under that gist. 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 gists by hand and specifying the likelihood each one gives to each of the 2^{25} displays infeasible. Thus, as a simplifiying asumption we chose to define gists using Markov Random Fields, which allow us to specify a probability distribution over all images by simply defining a small number of parameters about how nodes tend to differ from their immediate neighbors; such models have been used extensively in computer vision (Geman & Geman, 1984). We use only two gist parameters, which specify how often dots are the same or different color than their horizontal neighbors (G_h) and how often dots are the same or different color than their vertical neighbors (G_v). Thus, one particular gist ($G_h = 1, G_v = -1$) 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 gist would give high likelihood to displays that have many similar horizontal neighbors and few similar vertical neighbors.

We treat each dot in these change detection displays as a random variable D_i^1, where the set of possible values of each

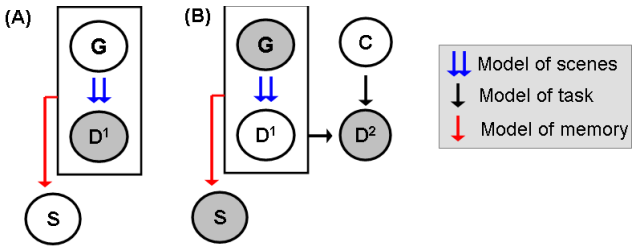

Figure 2: The model expressed in graphical model notation for (A) encoding and (B) detection. Filled circles indicate a node is observed (the model has access to it). Unfilled circles indicate the model must infer the value of the node. The arrows are colored based on what kind of process they represent. D1=the first display, D2=the second display, G=the gist S=specific items, C=the presence of a change.

D_i^1 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:

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

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

where the partition function:

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

normalizes the distribution. $\psi(D_i^1, D_j^1)$ is 1 if $D_i^1 = D_j^1$ 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 "gist" 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 $p(display|gist)$. To complete the encoding model we also need to define $p(items|display, gist)$ ($p(S|D^1, G)$). To do so, we define a probability distribution that preferentially encodes outlier objects (objects that do not fit well with the gist).

We choose whether to remember each object from the display by looking independently at the conditional probability of that object under the gist, assuming all of its neighbors are fixed $p(D_i^1|G, D_{/i}^1)$. 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 gist. Thus, the probability of encoding a set of objects (S) is:

$$p(S|G, D^1) = \prod_{j:s_j\in S} [1 - p(D_j^1|G, D_{/j}^1)] \prod_{j:s_j\notin S} p(D_j^1|G, D_{/j}^1) \quad (2)$$

This defines $p(S|D^1, G)$, which provides the probability of encoding a particular set of specific items in a given display, $p(items|display, gist)$, 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 gist 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 gist (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 gists 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 gists (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, $P(D^1|G)$ is given by Equation 1, and $P(S|G, D^1)$ is given by Equation 2.

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 simple 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.

Experiment 1 and 2

To examine human memory performance, we collected data using Amazon Mechanical Turk, where we had observers perform a change detection task for each of 24 different displays. We then compared this performance to our model.

The model makes predictions about how hard it is to detect changes in particular displays of colored dots (i.e., some changes will be more difficult to detect than others). In addition, it makes predictions about overall accuracy for a particular set of displays. We can thus examine how well the model fits with human memory performance in two distinct ways: (1) How many particular items (K) the model needs to recall to match human performance overall, and (2) how well the model's predictions about the difficulty of particular displays correlate with human memory performance.

Method

We sampled a set of 16 displays from the Markov Random Field model we use to define our gist (using Gibbs Sampling). Four of these displays were sampled from each of $G_h = \pm 1$, $G_v = \pm 1$. In addition, we generated 8 displays randomly. In Experiment 1, these 24 displays consisted of red and blue dots. In Experiment 2 they were exactly the same displays, but composed of black and white squares instead.

The displays were presented to 65 participants in Exp. 1 and a separate set of 65 participants in Exp. 2 using Amazon Mechanical Turk. The first display was flashed up for 750 ms (timing was controlled using Javascript), followed by a 750ms blank period; then the second display was flashed up for 750ms in a different screen location. Observers' task was simply to say whether the two flashed displays were the same or different (See Figure 1). Each display was presented to each observer in both a "same" and "different" trial, so observers completed 48 trials each, with the entire experiment lasting approximately 4 minutes. The order of the 48 trials was randomly shuffled for each subject. Observers were paid 30 cents for their time.

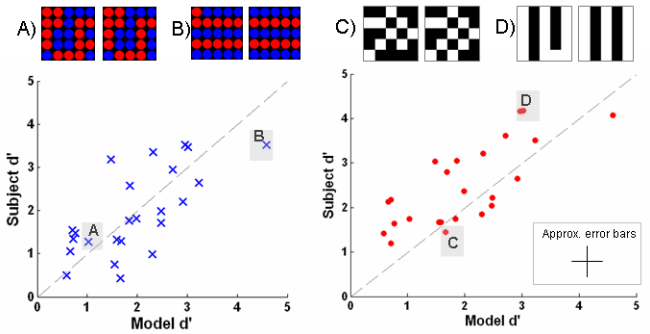

Figure 3: The fit of our probabilistic model to the observers' data with K=4 in the model for Experiments 1 and 2. Each point is the d' for a pair of displays. Approximate error bars are shown for both the subjects and model, calculated by bootstrapping standard errors. Example of both a hard and easy pairs of displays is shown for each experiment.

Results

For each display we computed a d', measuring how difficult it was to detect the change in that display (averaged across observers). Performance in Experiments 1 and 2 was highly similar, as the correlation in the display-by-display d' was r=0.91 between the two experiments. Thus performance was collapsed across both experiments for the remaining analyses.

On average, human observers d' was 2.18 (S.E. 0.06) suggesting they were quite good at detecting changes on these displays. Since the displays contain 25 dots, this d' corresponds to a Cowan's K of nearly 16.1 dots if the items are assumed to be represented independently and with no summary information encoded (Cowan, 2001). This is nearly 5 times the number usually found in simpler displays and thus represents a challenge to standard models of change detection and visual working memory capacity.

Importantly, our claim is not that observers remember 16 individual dots. Instead, our model provides an alternative explanation. The model achieves the same performance as people (d'=2.18) with a K value of only 4, thus encoding only four specific dots in addition to the display's gist (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. The model thus aligns nicely 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 which suggests a hierarchical representation with both gist and item information (e.g. Lampinen et al., 2001).

In addition to describing overall memory capacity, we can also examine the difficulty of particular displays. Previous models of change detection treat all displays as interchangeable, since they choose which objects to encode at random and do not represent any summary information about the display. They thus make no predictions about which particular changes will be hard or easy to detect. However, observers reliably find it more difficult to detect change in some displays than others, as measured both by averaging 200 split-half correlations on d-prime (r=0.75) and by bootstrapping standard errors on observers' d-prime (see Figure 3).

Our model does not treat each item independently, and chooses which items to encode by making strategic decisions based on the display's gist. Thus, our model does make predictions about the difficulty of detecting particular changes. In fact, the correlation between the model's difficulty with individual displays and the human performance on these displays was quite high (overall: r=0.71, $p<0.0001$; Exp.1: r=0.65, Exp.2: r=0.73; See Figure 3). Thus, the model's simple gist representation captures which changes people are likely to detect and which they are likely to miss.

Discussion

We here present a formal model of change detection which relies upon probabilistic inference to make predictions about visual working memory. The model takes into account the hierarchical nature of memory typically found in real-world scenes. It successfully predicts the display-by-display difficulty of visual working memory displays, indicating which changes observers will find easy to detect and which they will find difficult. The model also converges with the standard visual working memory literature on an estimate of 3-4 individual objects remembered, even in more complex patterned displays.

Importantly, the model recovers previous models of visual working memory capacity as a special case, and thus captures the properties of those models in displays with no higher-order information. However, by formulating change detection in terms of probabilistic inference, we can make much richer models of working memory than those typically used to calculate capacity in visual working memory experiments.

Non-independence in Visual Working Memory

While almost all experiments on visual working memory treat the items to be remembered as indepedent units, there are several exceptions (e.g., Jiang, Olson, & Chun, 2000; Sanocki & Sulman, 2008; Jiang, Chun, & Olson, 2004; Vidal, Gauchou, Tallon-Baudry, & Oregan, 2005). The most prominant exception to this assumption of independence is the work of Jiang et al. (2000), who suggested that the spatial context of other items is important to simple change detection tasks. On displays where the item that changed is presented in the context of the other items present at encoding, observers perform better at change detection (Jiang et al., 2000). This suggests the items are not represented independently of their spatial context. This is compatible with the encoding of both summary information and specific items used in our probabilistic model.

In addition, previous work by Brady et al. (2009); Brady and Alvarez (2010) demonstrates that observers can be induced to encode displays with colored dots using statistical regularities present between the dots, rather than treating each dot separately. Observers not only use information about co-occurence between items to form more compressed representations of these displays (Brady et al., 2009) but also encode the displays at multiple levels of abstraction, combining both an overall summary of the display and information about particular dots (Brady & Alvarez, 2010).

More broadly, the idea that memory encoding and retrieval are based on information represented at multiple levels of abstraction is common in the literature on reconstructive memory (Bartlett, 1932). Recent computational models similar in spirit to the one presented here have formalized this in both the domains of object size memory (Hemmer & Steyvers, 2009b) and more recently in the combination of gist and specific objects in real-world scenes (Hemmer & Steyvers, 2009a).

Chunking, Perceptual Grouping and Gist

One of the most popular explanations for observers' better-than-expected performance with more complex stimuli is chunking, or forming larger units out of smaller subsets of the stimuli (Miller, 1956; Cowan, 2001). In this framework, performance on our displays of patterned dots could be a result of observers' remembering only 3-4 independent items from the display and not encoding any overarching gist or structure. Instead, the items they remember would simply consist of multiple dots grouped into single items. This explanation has been proposed, for example, 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 & Jung, 1986).

This kind of chunking could potentially explain observers' performance on our displays. However, our preliminary work with a model that partitions the display into contiguous regions of the same color and remembers K of these regions suggests that such a model does not adequately explain performance in the current experiments. Instead, such a model either fails to capture the pattern of human errors or requires memory for an overly large number of regions ($K>5$) to achieve human levels of performance. However, future work is needed to examine models that perform such grouping or chunking and compare them with models, like ours, that represent the displays at multiple levels of abstraction.

Model of Gist

In the model the "gist" is encoded using Markov Random Fields, and thus the only information that can be represented are local spatial continuity properties of the colors in the display (similiarity between horizontal and vertical neighbors). Obviously, this is too impoverished to be a fully accurate model of human visual memory, even for such simple dot displays. For example, we could draw letters or shapes in the dot patterns, and people would recall those patterns well by summarizing them with a gist-like representation. Our model cannot capture such representations. However, we believe that our model nonetheless represents a step forward in understanding how people make use of such gist during change detection.

References

Alvarez, G., & 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.

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

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

Bays, P., Catalao, R., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7.

Brady, T. F., & Alvarez, G. A. (2010). Ensemble statistics of a display influence the representation of items in visual working memory. *Visual Cognition*, *18*(1), 114–118.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior. Vol*, *19*(4), 450–466.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Machine Intel.*, *6*, 721–741.

Hemmer, P., & Steyvers, M. (2009a). Integrating Episodic and Semantic Information in Memory for Natural Scenes. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society* (pp. 1557–1562). Austin, TX: Cognitive Science Society.

Hemmer, P., & Steyvers, M. (2009b). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, *16*(1), 80.

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*(3), 519–537.

Hollingworth, A., & Henderson, J. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, *7*(1), 213–235.

Hollingworth, A., Hyun, J., & Zhang, W. (2005). The role of visual short-term memory in empty cell localization. *Perception and Psychophysics*, *67*(8), 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.

Huber, D., Shiffrin, R., Lyle, K., & Ruys, K. (2001). Perception and preference in short-term word priming. *Psychological Review*, *108*(1), 149–182.

Jiang, Y., Chun, M., & Olson, I. (2004). Perceptual grouping in change detection. *Perception and Psychophysics*, *66*, 446–453.

Jiang, Y., Olson, I., & Chun, M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*(3), 683–702.

Lampinen, J., Copeland, S., & Neuschatz, J. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(5), 1211–1222.

Luck, S., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–280.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81–97.

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

Rouder, J., Morey, R., Cowan, N., Zwilling, C., Morey, C., & Pratte, M. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975.

Sanocki, T., & Sulman, N. (2008). Visual short term memory for location: Does objecthood matter? *Journal of Vision*, *8*(6), 203-203.

Simons, D., & Rensink, R. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Vidal, J., Gauchou, H., Tallon-Baudry, C., & Oregan, J. (2005). Relational information in visual short-term memory: The structural gist. *Journal of Vision*, *5*(3), 244–256.

Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135.

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