Title

A Nonparametric Bayesian Model of Visual Short-Term Memory

Authors

Orhan, Emin
Jacobs, Robert

Peer reviewed

A Nonparametric Bayesian Model of Visual Short-Term Memory

A. Emin Orhan (eorhan@bcs.rochester.edu)   **Robert A. Jacobs (robert.jacobs@rochester.edu)**

Department of Brain & Cognitive Sciences, University of Rochester, Rochester NY 14627, USA

Abstract

We present a nonparametric Bayesian model of the organization of visual short-term memory based on the Dirichlet process mixture model. Our model implements the idea that items in visual short-term memory can be encoded at multiple levels of abstraction, where the appropriate levels of abstraction and how much weight should be given to each level can be automatically determined. A capacity limit is implemented in this model by favoring small numbers of clusters of items. We show that various biases and distortions reported in visual short-term recall and recognition memory literatures can be quite naturally and elegantly explained by the model.

Keywords: Visual short-term memory; chunking; Dirichlet process mixture model; nonparametric Bayesian model.

Introduction

In a standard visual short-term memory (VSTM) experiment, subjects view a display containing multiple items with simple features (e.g. colored squares or oriented Gabor gratings) for a brief period of time and, after a delay period, are asked to report about one of those items. This procedure allows researchers to address a number of important questions: What is the exact content of their visual memory for the display at the end of the delay period and how is this content organized in VSTM? Are the items encoded independently in VSTM, or do memories for different items influence one another? If so, how?

There is substantial evidence suggesting that items in VSTM are not encoded independently of one another (Brady & Alvarez, 2010; Huang & Sekuler, 2010; Jiang, Olson, & Chun, 2000). As a specific form of such dependence between the representations of different items in VSTM, it has recently been argued that VSTM is organized hierarchically with each item being represented at two different scales, a fine scale (i.e. individually) and a coarse scale through the ensemble statistics of all items in the display (Brady & Alvarez, 2010; Brady & Tenenbaum, 2010; although see our discussion of Brady and Alvarez (in press) below). In this paper, we present a generalization of this idea based on the Dirichlet process mixture model (DPMM) (Neal, 2000). DPMM is a popular nonparametric model that can describe a dataset in terms of a probability distribution over its different possible partitions. Through the use of multiple partitions, our model can represent an item in VSTM not just at two levels of abstraction as proposed by Brady and Alvarez (2010) and Brady and Tenenbaum (2010), but at multiple levels of abstraction, including intermediate levels of abstraction. For instance, in one partition, an item might form its own group (i.e., a fine-scale representation of the item). In another partition, this item might be grouped with one other item that is highly similar to it (a moderate-scale representation). And in a third partition, the item might be grouped with all other items (a coarse-scale

representation). The advantage of DPMM is that it can automatically determine the appropriate partitions for the particular dataset at hand and the weights that should be alloted to each partition in the posterior distribution.

Details of the Model

Consider a single trial of a hypothetical VSTM experiment in which an observer needs to remember the feature values (e.g. position of a square or orientation of a Gabor grating) of N items in a display. We denote the actual feature value of item i by θ_i. One of the items, called the target item, is then cued and the observer is asked to report its feature value. The index of the target item is denoted by t and its feature value by θ_t. We model this single trial using a non-conjugate DPMM that assumes the following generative process (Neal, 2000):

$$\theta_i | \mu_i, \tau_i \quad \sim \quad \mathcal{N}(\theta_i; \mu_i, \tau_i) \qquad (1)$$
$$\mu_i, \tau_i | G \quad \sim \quad G \qquad (2)$$
$$G \quad \sim \quad DP(G_0, \alpha) \qquad (3)$$
$$G_0(\mu, \tau) \quad = \quad \mathcal{U}(\mu; a, b) \cdot \mathcal{G}(\tau; \alpha_\tau, \beta_\tau) \qquad (4)$$

Here, μ_i and τ_i are the mean and precision of the Gaussian component (or cluster) that item i is assigned to and they are identical for different i if the corresponding items are assigned to the same component. $\mathcal{N}(\theta; \mu, \tau)$ is a normal distribution with mean μ and precision τ, $DP(G_0, \alpha)$ is a Dirichlet process with base distribution G_0 and concentration parameter α. Roughly, α acts a capacity parameter in our model: for small values of this parameter, the model favors a small number of clusters or groups of similar items ("chunks"), whereas for large values, it tends to assign each item to its own cluster. We place a $\mathcal{G}(\alpha_c, 1)$ prior on the concentration parameter α, treating α_c as a free parameter. $\mathcal{U}(\mu; a, b)$ is a uniform distribution within the range (a, b) and $\mathcal{G}(\tau; \alpha_\tau, \beta_\tau)$ is a gamma distribution with parameters α_τ and β_τ. We set the range of the uniform distribution to a large interval including minimum and maximum possible values for the relevant variable in each experiment below. For the parameters of the gamma distribution α_τ, β_τ, we put a $\mathcal{G}(1, 1)$ prior on β_τ and treat α_τ as a free parameter. This reduces the total number of free parameters to just 2, namely α_c for the concentration parameter, and α_τ for τ. We use the same parameter values α_τ, α_c for all trials of an experiment. Inference is performed via a Markov chain Monte Carlo (MCMC) sampling algorithm with auxiliary variables (Algorithm 8 in Neal (2000)).

Figure 1 schematically illustrates, in a hypothetical one-dimensional example, the idea of representing an item at multiple levels of abstraction with a DPMM. In this figure, feature values of three different items, θ_i, shown in a single trial of a hypothetical VSTM experiment, are represented by
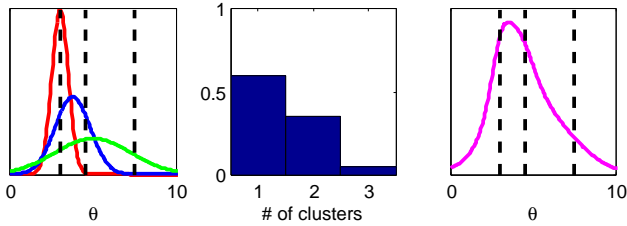
Figure 1: (Left) $p(\tilde{\theta}_t|\mu_t,\tau_t)$ at different iterations of the MCMC algorithm; (Middle) posterior distribution over the number of clusters; (Right) the proposed memory representation of target t: $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N}) = \int p(\tilde{\theta}_t|\mu_t,\tau_t)p(\mu_t,\tau_t|\{\theta_i\}_{i=1,...,N})d\mu_t d\tau_t$.

the dashed lines. At different iterations of the MCMC sampling algorithm, the target item (here the leftmost one) gets assigned to different Gaussian components. For instance, at one iteration, the target is assigned to a Gaussian component of its own (represented by the red line in the leftmost graph of the figure). In this case, the item is said to be encoded individually. At another iteration of the sampling algorithm, a different partition is drawn from the posterior distribution; the target is now grouped with the nearest item and the two items together get assigned to a common Gaussian component (represented by the blue line). In this case, the two items are said to be encoded together. At yet another iteration, the partition sampled from the posterior distribution assigns all three items into a single group and hence they are represented by a single Gaussian component (shown by the green line). This last representation is similar to the coarse representation of items by their ensemble statistics as in Brady and Alvarez (2010).

Formally, the three distributions shown in the leftmost plot represent $p(\tilde{\theta}_t|\mu_t,\tau_t)$, with $\tilde{\theta}_t$ representing an estimate of the feature value of the target item, for three different samples μ_t, τ_t from the posterior distribution over the Gaussian components that the target is assigned to, $p(\mu_t,\tau_t|\{\theta_i\}_{i=1,...,N})$. The middle plot in Figure 1 shows the posterior distribution over the number of groups or clusters. This plot shows that for this particular example, most of the time, all items get assigned to a single cluster as most of the posterior weight is on this case and items are rarely individually encoded. Finally the rightmost plot shows the posterior predictive distribution for the target item: $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N}) = \int p(\tilde{\theta}_t|\mu_t,\tau_t)p(\mu_t,\tau_t|\{\theta_i\}_{i=1,...,N})d\mu_t d\tau_t$, i.e. an infinite mixture of Gaussian components each weighted by its posterior weight.[1] This distribution is also computed via sampling. We use this last posterior predictive distribution as the memory representation of the target item t and refer to it as such.

Next, we show that our model can both qualitatively and quantitatively explain certain biases and distortions previously reported in visual short-term recall and recognition

---

[1] Note that this integral only involves the components μ_t, τ_t that the target is assigned to, and is thus different from the overall posterior predictive distribution, $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$.

memory literatures. Specifically, we focus here on results from two short-term recall tasks (Brady & Alvarez, 2010; Wilken & Ma, 2004) and from a short-term recognition task (Viswanathan, Perl, Visscher, Kahana, & Sekuler, 2009).

Explaining Biases in Short-Term Recall

Brady and Alvarez (2010): Brady and Alvarez (2010) demonstrated that ensemble statistics of items in a display influence memory for individual items. In their experiment, subjects were presented with three red, three blue and three green circles with different sizes for a brief duration and they were asked to remember the sizes of red and blue circles only and to ignore the green circles. After a delay interval, a black circle appeared at the location of a red or a blue circle in the original display. Subjects had to indicate the size of the original (red or blue) circle that was at that location by using the mouse to resize the black circle. Brady and Alvarez (2010) showed that the reported size of the target circle was biased toward the average size of the circles having the same color as the target. It is important to note that they used pairs of identical configurations in their design such that the only difference between trials in a matched pair was that the target circle and a randomly chosen circle of the opposite color swapped their colors. They measured subjects' bias by first finding, for each matched pair of trials, the colors associated with the smaller and larger average circle sizes. They, then, measured the ratio between the mean reported size when the target circle appeared in the same color as the larger average size circles and the mean reported size when it appeared in the same color as the smaller average size circles. The value of this ratio indicated whether subjects had a bias toward the average size of the circles appearing in the same color as the target. A value of 1 would indicate that subjects did not have any bias toward the average size, as they would be equally likely to report smaller or larger sizes than the actual size of the target; whereas a value significantly greater than 1 would indicate that subjects were more likely to report larger sizes when the target appeared in the same color as the larger circles and smaller sizes when it appeared in the same color as the smaller circles, even though the actual size of the target was the same between the trials of a matched pair.

Brady and Alvarez (2010) reported an average bias of 1.2 (SEM: 0.05, 6 subjects and 400 trials each), suggesting that subjects had a bias toward the average size of the circles appearing in the same color as the target. We simulated 3000 trials of the Brady and Alvarez (2010) experiment to test if our nonparametric Bayesian model would be able to account for the experimentally observed bias. Sizes of the blue and red circles in each trial were generated in accordance with their procedure: first, the mean sizes of the blue and red circles for a single trial were randomly chosen from the interval [0.625°, 3.125°]. The actual sizes of the three red and three blue circles were then drawn from normal distributions with corresponding means and standard deviations that were equal to one-eighth of the corresponding means. After generating stimuli for 1500 such trials, stimuli for the remaining 1500
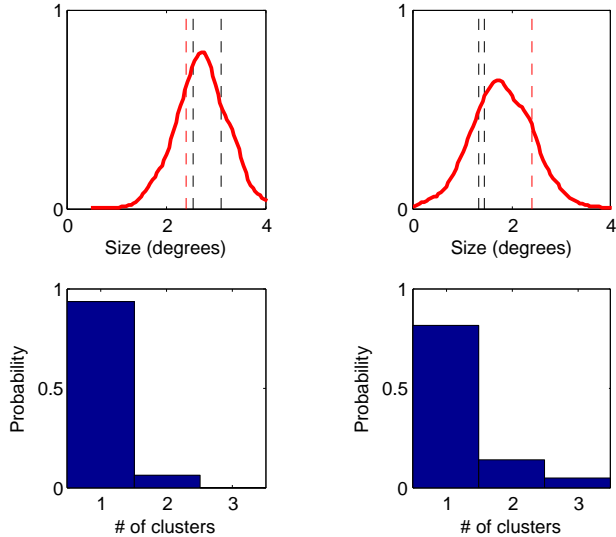
Figure 2: (Top row) feature values of the items (dashed lines, feature value of the target is represented by the red dashed line) and the memory representation of the target; (bottom row) posterior distribution over the number of clusters.

trials were generated by simply swapping, for each trial, the color of the target circle with that of a randomly chosen circle of the opposite color. We assumed that subjects would group the target with other circles of the same color and therefore considered only those circles while running the model. The free parameters of the model (α_c and α_τ) were optimized via a simple grid search method to fit the experimentally observed bias as closely as possible. Once the proposed memory representation for the target, $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N})$, is computed, there are different ways to generate a response. One possibility is to take the mean of this distribution as the model's response. Instead, we draw a single value from this distribution and consider this random value, $\tilde{\theta}_t^m$, to be the model's response. Although both of these response generation methods yielded similar biases on average, the second method provided a better fit to standard deviations of subjects' average absolute error reported in Brady and Alvarez (2010) (0.5°on average, SEM: 0.06°). Due to the stochasticity of the responses, the sampling procedure was repeated 10 times for each trial, each time drawing different samples from the distribution representing the memory for the target item. Here we report the mean biases of these 10 repetitions. The mean bias of the fitted model was 1.2030 (SEM: 0.0033) and its mean absolute error was 0.4890°(SEM: 0.0018°), both very close to the experimentally observed values (mean bias: 1.2, SEM: 0.05; mean absolute error: 0.5, SEM: 0.06).

Figure 2 shows the model results for a typical matched pair of trials. In this figure, the top row shows the feature values of the items (dashed lines, feature value of the target is represented by the red dashed line) together with the memory representation of the target, $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N})$, for the two matched trials. Note how the memory representation of the

target is shifted toward the mean feature value of the items in both cases. The bottom row shows the posterior distribution over the number of clusters for the two trials. These posterior distributions (and those from other trials not shown here) are dominated by 1-cluster partitions, suggesting that subjects' performance can be explained as relying overwhelmingly on a 1-cluster, global, or ensemble representation of the items.

Wilken and Ma (2004): Using a recall paradigm, Wilken and Ma (2004) showed that subjects displayed systematic biases in their judgments in a short-term memory experiment that used spatial frequency as the relevant feature. In their Experiment 9, subjects are first briefly shown a number of Gabor stimuli with different spatial frequencies randomly drawn from 16 spatial frequency values linearly spaced between 4 cycles/degree and 8 cycles/degree. Different set sizes used in the experiment were $N = 2, 4, 6, 8$. After a delay interval, one of the N Gabors is cued and the subjects' task is to adjust the spatial frequency of a probe Gabor stimulus using the arrow keys so that it matches the frequency of the cued Gabor. Wilken and Ma (2004) showed that subjects tended to overestimate the spatial frequencies of low-frequency Gabors, but tended to underestimate the spatial frequencies of high-frequency Gabors, i.e. subjects showed a bias toward the mean spatial frequencies in their judgments. Wilken and Ma (2004) also observed that the magnitude of this bias depended on the set size with smaller set sizes leading to smaller biases (see Figure 8 in Wilken and Ma (2004); also reproduced in Figure 3 here).

We sought to determine whether our nonparametric Bayesian model would be able to explain, both qualitatively and quantitatively, these observed biases. We first generated a dataset according to the procedure described above. For each set size, 3000 trials were simulated. We then ran the DPMM model on these data. In fitting the model to observed biases, α_τ was allowed to vary across different set sizes, but α_c was fixed across different set sizes; and both parameters were optimized to fit the experimentally observed biases as close as possible.

Biases obtained from the best-fitting DPMM model are shown in Figure 3 together with the experimentally observed biases. The model was able to capture the two main patterns in the observed biases: the linear relationship between the bias and target frequency and the increase in the magnitude of bias with set size, as well as providing a good quantitative fit: root mean squared error, RMSE: 0.2437. Figure 3 also shows the posterior distributions over the number of clusters averaged over all trials for different set sizes. These posterior distributions are dominated by partitions with small numbers of clusters, suggesting that subjects' performances can be explained as relying overwhelmingly on coarse representations of the items.

Explaining the Inter-Item Similarity Effect

Kahana and Sekuler (2002) showed that inter-item similarity between stimuli influences subjects' performances in a standard old/new recognition task. In their Experiment 1, on each
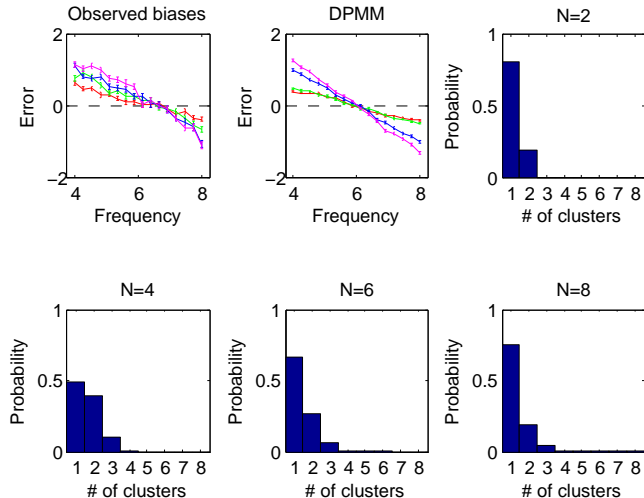
Figure 3: The first two plots show the observed biases and biases predicted by the DPMM for different set sizes (magenta $N = 8$, blue $N = 6$, green $N = 4$, red $N = 2$); error bars represent ∓ 1 SEM across trials. The remaining four plots show the posterior distributions over the number of clusters averaged over all trials for different set sizes.

trial, subjects were first successively shown sinusoidal gratings with different spatial frequencies. After a delay interval, they were shown a test grating (or probe) which, on half of the trials, had the same spatial frequency as one of the gratings presented on the corresponding trial (OLD) and on the other half of the trials, had a different spatial frequency than all the gratings presented on the corresponding trial (NEW). The task was to decide if the spatial frequency of the test grating was OLD or NEW. Kahana and Sekuler (2002) then fit the human data collected from this experiment by a simple 'noisy exemplar' (NEMO) model that included separate terms for the effects of the similarity between the test probe and each of the study items and the inter-item similarity between the study items. They found that the inter-item similarity term had a significant effect on OLD/NEW decisions over and above the effect of individual probe-item similarities. In particular, when the probe-item similarities were fixed, larger inter-item similarities increased the likelihood of a NEW response, making the subject more likely to judge the probe to be a novel item. Note that a model that represented the study items independently would not be able to show any inter-item similarity effects, as the inter-item similarity depends on all the study items simultaneously.

Kahana and colleagues replicated the main inter-item similarity effect in later works. Here, we specifically consider an article by Viswanathan et al. (2009) and focus on one of their experiments demonstrating that homogeneity or inter-item similarity influences subjects' responses. The design of their experiment was, for our purposes, identical to the design of the experiments in Kahana and Sekuler (2002) described in the previous paragraph. There were two conditions

in the Viswanathan et al. (2009) study: a medium homogeneity condition and a high homogeneity condition. These two conditions are schematically described in the left and middle plots in Figure 4. In this figure, the spatial frequencies (in just noticeable difference or JND units) of study gratings are represented by black dashed lines (at 1, 4 and 8 JND in the medium homogeneity condition, and at 3, 4 and 8 JND in the high homogeneity condition) and the spatial frequency of the probe grating is represented by the green dashed line (at 2 JND in both conditions). Note that the probe is a NEW item (or a lure) in both conditions. Also note that the individual probe-study item similarities are identical in the two conditions so that the only difference between these conditions is the inter-item similarity of the study items, with the high homogeneity condition having a higher inter-item similarity than the medium homogeneity condition. The inter-item similarity effect refers to the finding that subjects had a significantly higher probability of responding OLD (or YES) in the medium homogeneity condition than in the high homogeneity condition (mean $P(YES) = 0.69$ vs. mean $P(YES) = 0.57$).

We applied the DPMM model to the two conditions in Viswanathan et al. (2009) and optimized the parameters to fit the experimentally observed probabilities of YES responses. Since this task is an old/new recognition task, we cannot use the memory representation of the target $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N})$ introduced previously, as unlike in other examples considered in this paper which are all recall tasks, there is no single target item in an old/new recognition task. Instead, we use the overall predictive distribution, $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$, for this recognition task, which intuitively quantifies the likelihood that $\tilde{\theta}$ was the value of any one of the study items; whereas the distribution $p(\tilde{\theta}_t|\{\theta_i\}_{i=1,...,N})$ that is used elsewhere in this paper, quantifies the likelihood that $\tilde{\theta}_t$ was the value of the target item t. In the computation of the predictive distribution for the target item, only the components that the target is assigned to (μ_t and τ_t) are considered; whereas in the computation of the overall predictive distribution, $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$, all components in the posterior distribution are taken into account, regardless of whether the target is assigned to them or not. The overall predictive distribution, $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$, can be thought of as a nonparametric density estimator of the given dataset and it can be easily computed by sampling (see section 4 of Görür and Rasmussen (2010) for details).

Since $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$ is a probability density, we need to transform it into a probability of responding YES (or OLD). We do this by normalizing $p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})$ between 0.2 and 0.8: $P(YES) = 0.6 * (p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N})/max(p(\tilde{\theta}|\{\theta_i\}_{i=1,...,N}))) + 0.2$ so that the maximum probability of responding YES is 0.8 and the minimum probability of responding YES is 0.2. We did not normalize probabilities between 0 and 1 because in these experiments subjects tend to have a relatively high probability of responding YES even for very dissimilar lures, and a relatively low probability of responding YES even for perfect matches or targets. The specific values of 0.2 and 0.8
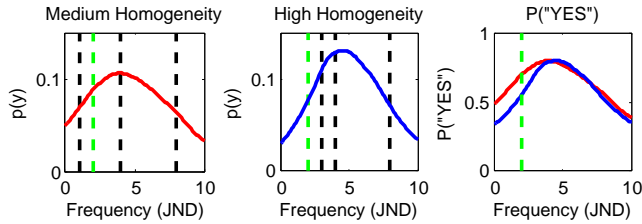
Figure 4: (Left) Overall predictive distributions in the medium and high homogeneity conditions. Spatial frequencies of the study items are shown with black lines and the probe frequency is represented by the green line. (Right) Probability of responding YES in the two conditions.

were chosen based on a similar previous study by Kahana, Zhou, Geller, and Sekuler (2007).

Results are shown in Figure 4. The DPMM was able to reproduce the inter-item similarity effect: the probability of responding YES was 0.70 for the medium homogeneity condition and 0.55 for the high homogeneity condition. Also note that the model predicts that the lure probe (represented by the green line) will elicit a higher percentage of YES responses than the rightmost study item in both conditions: for the rightmost study item the model predicts $P(YES) = 0.56$ and $P(YES) = 0.53$ in the medium and high homogeneity conditions, respectively. This prediction is confirmed in the Viswanathan et al. (2009) study, with observed mean $P(YES) = 0.57$ and $P(YES) = 0.52$ for the rightmost study item in the respective conditions. The model was not explicitly fit to these last two observed probabilities.

Intuitively, the reason that the DPMM can successfully account for the inter-item similarity effect is that the posterior distribution was dominated by 1-cluster or 1-component partitions. Since the items in the high homogeneity condition have a lower variance, this one component, in turn, had lower variance in the high homogeneity condition. Even though the probe has roughly the same Euclidean distance to the mean in both conditions, it has a higher Mahalanobis (or effective) distance to the mean in the high homogeneity condition, making the probe less similar to the study items in that condition.

Accounting for Multiple Levels of Abstraction

More recently, in an experiment similar to the recall experiment of Brady and Alvarez (2010) discussed above, Brady and Alvarez (in press) showed that items can be represented at multiple levels of abstractions in VSTM (including intermediate levels), not just at two fixed-resolution levels. Experiment 1 of Brady and Alvarez (in press) replicated the main result of Brady and Alvarez (2010). They hypothesized that the fact that color was a task relevant feature in this experiment (subjects had to remember only the red circles and the blue circles and ignore the green circles) might have increased the salience of this feature, thereby inducing subjects to use a color-based encoding for the circles. This implies that in a similar task where color is not task-relevant, the observed bi-

ases toward the mean size of the same colored circles will disappear. In their Experiment 2, Brady and Alvarez (in press) tested this prediction by removing the green circles from the display and presenting only the red circles and the blue circles in each trial. Subjects were told to remember the sizes of all circles in the display. Therefore, color was no longer a task-relevant feature. Consistent with their hypothesis, Brady and Alvarez (in press) found that subjects did not show any bias toward the mean size of the same colored circles. However, they observed a bias toward the mean size of all circles in the display, suggesting that in different contexts, stimuli might be represented at different levels of abstraction in VSTM.

To explain this distinctive pattern of biases when color was a salient feature versus when it was not, Brady and Alvarez (in press) used a two-level hierarchical model for Experiment 2 and a three-level hierarchical model for Experiment 1. In the two-level hierarchical model, the two levels were the level of individual circles (individual encoding) and the overall mean of the sizes of all circles in the display. Since the color-based grouping of circles is not taken into account in this two-level hierarchical model, in estimating the sizes of individual circles only a bias toward the overall mean size is predicted by the model, in accordance with the results of Experiment 2. On the other hand, in the three-level hierarchical model, the three levels were the level of individual circles (individual encoding), the group level means of the sizes of red and blue circles and the overall mean of the sizes of all circles in the display. Since the color-based grouping of circles is explicitly incorporated into the model as a separate level, both a bias toward group level means and a bias toward the overall mean are predicted by this three-level hierarchical model, largely consistent with the results of Experiment 1 (it is not clear if the results from their Experiment 1 support a bias toward the overall mean in addition to the observed bias toward the mean size of the same-colored circles).

Although their three-level hierarchical model allows for the representation of items at multiple levels of abstraction (including intermediate levels), there are two problems with (Brady & Alvarez, in press)'s hierarchical modeling approach. First, using hierarchical models with different numbers of levels in order to account for the pattern of results observed in different experiments is *ad hoc*. It is not clear what determines the number of levels of hierarchy that should be used for a given experiment and the appropriate 'grain' of those levels. Second, related to the first problem, note that in their three-level hierarchical model, in the group level representation of the circles, the number of groups (i.e. the number of colors) and the assignment of each circle to a single group had to be explicitly specified beforehand. Although it is rather easy to do this in their experiment (as the number of colors and the color of each circle are clear), it is not clear how to define groups or how to assign each item to a group in more ambiguous cases. It would be much better if the model could determine these automatically.

The DPMM can deal with each of these problems and pro-

vides a much more parsimonious and elegant explanation of Brady and Alvarez (in press)'s results. We first represent the colored circles used in Brady and Alvarez (in press) as two-dimensional points in the feature space of color and size (for now using arbitrary numerical values for the blue and red colors). Following Brady and Alvarez (in press), we assume that the manipulation in Experiment 2 (removing the green circles) reduces the salience or weight of the color dimension and consequently the distances along the color dimension shrink in Experiment 2, effectively making the stimuli one-dimensional. This is illustrated in Figure 5 in the context of a hypothetical example. A straightforward two-dimensional extension of the DPMM presented in this paper, where one uses bivariate normal component distributions instead of univariate ones and appropriate base distributions, produces exactly the observed results. In Experiment 1, since color is a salient dimension, color-based partitions of the stimuli, where the red circles are grouped into a single cluster and the blue circles into another, are prominent, producing a bias toward the mean size of the same-colored circles. In Experiment 2, on the other hand, the stimuli effectively collapse into a single dimension (size) and color-based partitions of the stimuli are no longer prominent or likely under the model. Indeed, the most likely partition in this case is the partition where all stimuli are grouped into a single cluster, producing a bias toward the overall mean size of all the circles. Note that the DPMM automatically determines the appropriate levels of abstraction to represent the stimuli and the weights that should be alloted to each of them without any supervision.

General Discussion

We introduced a computational model of the organization of VSTM based on the DPMM. The DPMM provides an elegant implementation of the idea that items in VSTM can be represented at multiple levels of abstraction. For all the experiments considered here, our model suggested that subjects' performances in these experiments could be explained as relying on relatively coarse representations of the items.

As illustrated in the previous section, our model can be easily extended to multi-dimensional features using multivariate component distributions instead of univariate ones. Future work might use combinations of feature dimensions considered in this paper, such as orientation or spatial frequency, with other features such as spatial or temporal locations of items.

Acknowledgments

References

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

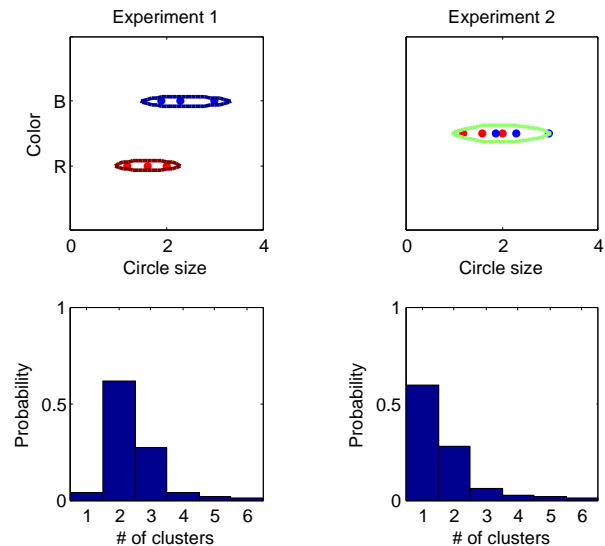Brady, T., & Alvarez, G. (in press). Hierarchical encoding in

Figure 5: Two hypothetical trials from Experiment 1 (left column) and Experiment 2 (right column) of Brady and Alvarez (in press). The top row shows the configurations of stimuli in the color-size space and the most likely partitions (represented by the contours) under a two-dimensional DPMM. The bottom row shows the posterior distributions over the number of clusters for the corresponding trials.

visual working memory: ensemble statistics bias memory for individual items. *Psychol Sci*.

Brady, T., & Tenenbaum, J. (2010). Encoding higher-order structure in visual working memory: A probabilistic model. In *Proceedings of the 32nd annual conference of the cognitive science society*.

Görür, D., & Rasmussen, C. (2010). Dirichlet process gaussian mixture models: Choice of the base distribution. *J Comput Sci Technol*, *25*, 653–664.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *J Vis*, *10*, 1–27.

Jiang, Y., Olson, I., & Chun, M. (2000). Organization of visual short-term memory. *J Exp Psychol Learn*, *26*, 683–702.

Kahana, M., & Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Res*, *42*, 2177–92.

Kahana, M., Zhou, F., Geller, A., & Sekuler, R. (2007). Lure-similarity affects visual episodic recognition: Detailed tests of a noisy examplar model. *Mem Cognition*, *35*, 1220–32.

Neal, R. (2000). Markov chain sampling methods for dirichlet process mixture models. *J Comput Graph Stat*, *9*, 249–265.

Viswanathan, S., Perl, D., Visscher, K., Kahana, M., & Sekuler, R. (2009). Homogeneity computation: How interitem similarity in visual short-term memory alters recognition. *Psychon B Rev*, *17*, 59–65.

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