Structured priors in visual working memory revealed through iterated learning

Timothy F. Lew (tflew@ucsd.edu) and Edward Vul (evul@ucsd.edu)
University of California, San Diego
Department of Psychology, 9500 Gilman Dr.
La Jolla, CA 92093 USA

Abstract

What hierarchical structures do people use to encode visual displays? We examined visual working memory's priors for locations by asking participants to recall the locations of objects in an iterated learning task. We designed a non-parametric clustering algorithm that infers the clustering structure of objects and encodes individual items within this structure. Over many iterations, participants recalled objects with more similar displacement errors, especially for objects our clustering algorithm grouped together, suggesting that subjects grouped objects in memory. Additionally, participants increasingly remembered objects as lines with similar orientations and lengths, consistent with the Gestalt grouping principles of continuity and similarity. Furthermore, the increasing tendency of participants to remember objects as components of hierarchically organized lines rather than individual objects or clusters suggests that these priors aid the perception of higher-level structures from ensemble statistics.

Keywords: Visual working memory; Markov chain Monte Carlo with people; non-parametric Dirichlet process

Introduction

Visual working memory stores object features (e.g. locations) according to their statistical structure (Alvarez & Oliva, 2009). If I see a crowd, for instance, I might organize them into groups and remember the locations of both the individuals and their higher-order groups. Although any stimulus has its own ensemble statistics, people also have expectations from the real world about how objects are organized. Here, we try to characterize Gestalt priors about the spatial arrangement of objects in an iterative visual working memory paradigm.

Visual working memory can use statistical structure to compensate for uncertainty about individual objects (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013). Although relying on objects' statistical structure biases memories of those objects, it can improve the overall fidelity of recall. Furthermore, encoding objects according to their statistical structure constrains the possible properties of those objects, allowing observers to remember the objects' exact features more precisely (Sims, Jacobs & Knill, 2012; Orhan, et al., 2014). For example, inferring that a set of objects generally fall on a horizontal line constrains their y-coordinates. This allows the observer to focus on encoding their x-coordinates with greater precision.

The effectiveness of an encoding scheme depends on how well it matches the statistics of a stimulus (Orhan & Jacobs, 2014a). Consequently, when people's priors about statistical structures fail to match what they observe, the fidelity of visual working memory will suffer. Orhan & Jacobs (2014b), for example, found that in typical studies of capacity, priors that stimuli are similar or form continuous lines conflict with stimuli that have uniformly distributed features. This mismatch can detrimentally bias memory and potentially explain a significant portion of performance limitations. In short: how people use the structure of displays to help encode visual information depends on what priors they have about the structure of objects in displays.

In the current study, we examined people's visual working memory priors using a "Markov chain Monte Carlo with people" task (Sanborn & Griffiths, 2007). In our task, one participant studies the positions of many dots on a screen, then reports those positions, and then the next participant studies the previous participant's responses, and so on. A long sequence of individuals encoding and reproducing the responses of previous participants yields a Markov chain that will emphasize the priors that people use to encode object locations. Kempe, Gauvrit & Forsythe (2015) previously used such an iterated learning task to examine visual working memory for binary grids and compared the complexity of transmitted information between children and adults. Their study, however, remained agnostic as to the actual structures that made up complex displays. A simpler display, for instance, could have reflected elements organized into less dispersed clusters, or more linear arrangements. Consequently, this study could not characterize the display structure priors that participants bring to bear to encode displays.

We used a non-parametric clustering algorithm (a Dirichlet process mixture model; Ferguson, 1983; Orhan & Jacobs, 2013; Austerweil, 2014) to predict what kinds of groupings subjects would infer from the displays. Participants used the groupings predicted by the clustering algorithm and grew more likely to group objects together over time. Our clustering model revealed that participants increasingly organized the objects into straight lines and in turn remembered the orientations and lengths of those lines using their ensemble statistics. These results suggest that people possess priors that objects are arranged linearly and those lines possess similar features. In this way, our study allowed us to recover the Gestalt grouping principles of continuity and similarity.

Experiment

Participants studied and then recalled the locations of spots on a computer. Critically, we showed the locations one subject reported as the stimulus to the next subject, thus producing an "iterated learning" chain. Based on the logic of "Markov chain Monte Carlo with people" (Sanborn &

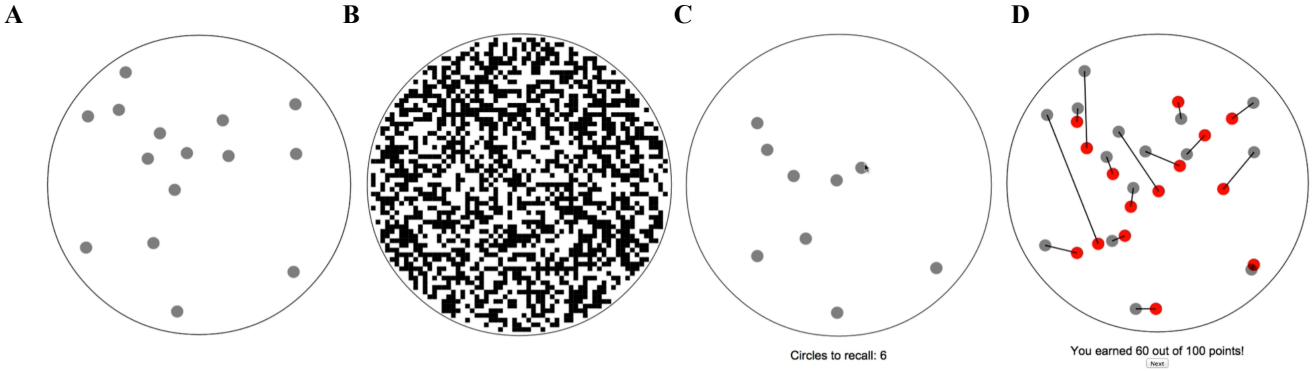

Figure 1: Example trial. (A) Participants saw 15 grey circles for 10 seconds followed by (B) a 1 second mask. (C) Participants then recalled the locations of all the circles and were told how many circles they had to recall. Participants could move around the circles until they were satisfied. (D) Participants then saw the correct object locations (grey) and their guesses (red) and the mapping between the targets and their guesses (black lines). Their score out of 100 was shown on the bottom.

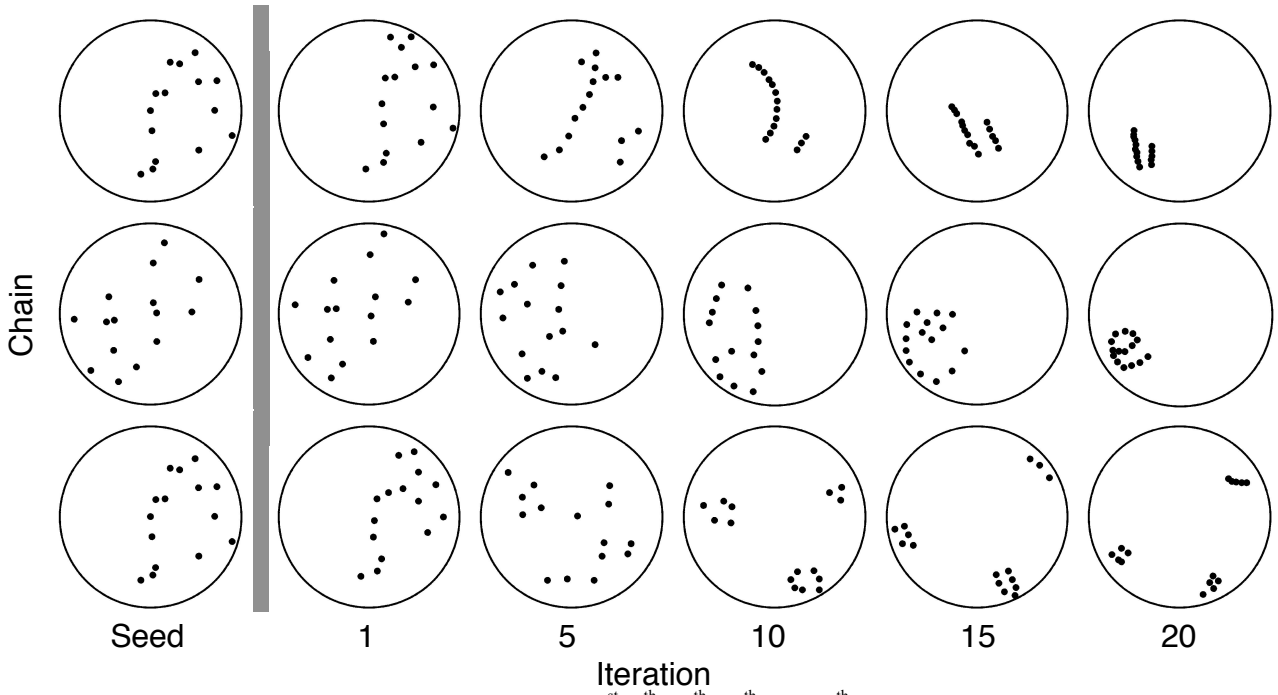

Figure 2. Three example chains (rows) for the seed display, 1st, 5th, 10th, 15th, and 20th iterations (columns). Grey lines separate the seed displays from the iterated trials. Circles are black in this figure for clarity (participants actually saw grey circles as in Figure 1A). Despite objects being initially uniformly distributed in the displays, participants gradually organized them into complex structures.

Griffiths, 2007), such a process will tend to converge to the prior, in our case yielding samples of the sorts of location structure people expect in images.

Methods

We generated 10 initial seed displays, each containing 15 circles with uniformly distributed locations. For each seed display, we ran 10 iterated learning chains for 20 iterations each. We allowed participants from the Amazon Mechanical Turk Marketplace (who performed our study for payment and a performance-based bonus) to repeat the experiment for different seed displays, resulting in a total of 1581 unique subjects yielding 2000 experimental runs.

In each trial, participants observed the locations of 15 circles for 10 seconds (Figure 1A), followed by a 1 second mask (Figure 1B). Participants then recalled the locations of the circles by clicking the mouse (Figure 1C). Participants had unlimited time to recall the locations of the circles and could move them (by dragging) as much as they wanted. Once participants indicated that they were done reporting the locations (by pressing enter), we gave them feedback by showing the correct and recalled locations along with lines indicating how far off they were (Figure 1D). We determined the mapping between guesses and targets using a greedy search that minimized root mean square error (RMSE). Participants also received a score between 0 and 100 based on the average distance between guesses and

targets normalized by the standard deviation of object locations. Participants were instructed that their final bonus would reflect their scores.

In each experimental run, a participant first performed a randomly generated practice trial to familiarize themselves with the task. The second trial was our main test in which they saw locations from the iterated learning chain (either the seed display for the first iteration, or the locations reported by the previous participant in the chain). In the third trial, participants observed the seed display, giving a measure of baseline performance (so subjects who were exposed to the first iteration of a chain would see the same display twice). The fourth trial was a randomly generated performance check: if their score was below criterion on this test, their responses were not included in the iterated learning chain to prevent a single inattentive subject from ruining an entire chain. Figure 2 shows several example chains from our study (movies of all the seeds and chains are located on our website at www.evullab.org/dots.php).

Non-parametric Dirichlet clustering algorithm

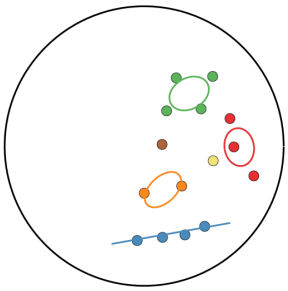

Figure 3: Example of the Dirichlet clustering algorithm's inferred grouping for a single trial. The clustering algorithm estimates the assignment of objects to groups (objects color-coded by group membership) and the parameters of the group structure: either a Gaussian cluster (represented by a covariance ellipse) or a line.

We designed a Dirichlet-process clustering algorithm similar to Orhan & Jacobs (2013) to estimate the grouping structure that subjects might infer. Critically, this grouping model allows the number of groups to vary and each group to be either a Gaussian cluster with a mean location and a spatial covariance matrix or a line segment with a particular location, length, and orientation (Figure 3). To minimize false positive identifications of lines, we set the standard deviation of objects around lines to be very small (ensuring that lines were thin) and required that lines contain at least four objects (to ignore coincidentally collinear objects). We held the two free parameters constant throughout all analyses (concentration=.25 – a prior on the number of groups; and line noise=2.5 pixels – the standard deviation of reported locations around a line. For reference, each circle had a radius of 10 pixels). We used a Gibbs sampler (Geman & Geman, 1984) to fit the model to each trial. In our analyses, unless otherwise stated, we use the maximum likelihood (MLE) groupings.

Results

Did participants group objects?

If participants grouped objects together per our clustering algorithm, then objects in the same group should have correlated errors (i.e. would tend to be misreported in the same direction). We matched participants' responses to objects' correct locations using the Hungarian algorithm (Kuhn, 1955) to minimize total root mean square error, thus finding the translational error $\boldsymbol{x}_i$ for each object i. For each pair of objects, we define the similarity of their displacement errors (**q**) as:

$$\boldsymbol{q}_{ij} = \frac{\boldsymbol{x}_i \boldsymbol{x}_j^T}{\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|}$$

Where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are vectors containing the translational errors of the reported locations. This error-similarity metric will be **q**=1 if the recalled locations of two objects were shifted in the exact same direction, and **q**=-1 if they were shifted in the exact opposite direction. If participants recalled objects independently, then the expected value of **q** would be 0.

The error similarity of objects that our clustering algorithm grouped together was significantly greater than the similarity of objects in different groups ($t(9)=13.71$, $p<.001$), indicating that the clustering model predicted the structure of errors in participants' responses, and therefore the display structure that participants inferred.

What priors did participants converge towards?

In an iterated learning chain, participants' responses should converge towards their priors (Sanborn & Griffiths, 2007). Therefore we can assess the structured priors that people use by estimating the properties of the structures to which participants' reported locations converge over the iterated learning chain.

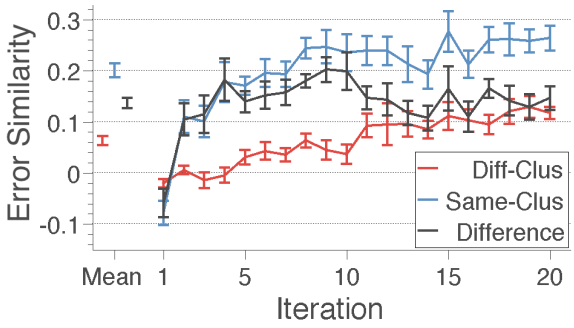

Figure 4: Translational error correlation. The continuous lines indicate error correlations over iterations. The points (Mean) indicate the error correlations averaged over iterations. Diff-Clus (red) represents the error correlation for objects in different groups as estimated by the clustering algorithm, Same-Clus (blue) represents the error correlation for objects in the same cluster according to the clustering algorithm and Difference (grey) represents the difference between the different and same cluster error correlations. Errors became more correlated over iterations and were more similar for objects grouped together by the clustering algorithm.

Translational error correlation over time. If the iterated learning task yielded more grouping structure, and thus more reliance on the grouping, over time, then the translational error correlation should increase over iterations. To test this prediction, we measured the translational error correlation for objects the clustering model inferred were in different groups and the same group at each step of the iteration (Figure 4). Not only are errors of objects that the clustering algorithm predicted would be in the same group more similar than translational errors for objects in different groups, but this diagnostic translational error correlation increased over iterations for both measures. The increasing similarity of errors for objects in the same cluster demonstrates that participants became more likely to remember objects in coherent groups.

Structured memory model convergence predictions. Insofar as our clustering model captures the priors participants used to encode objects, we should expect both the participants and the clustering model to converge towards the same structures. Intuitively, the model compensates for uncertainty about individual objects by recalling objects biased towards their structures; as such, we expect that over multiple simulated iterations of learning and recalling displays by this model, the reported displays will converge towards the structured prior. To generate such simulated "model chains" from the model, we constrained one free parameter: the noise with which it encodes the objects' locations (we set this to 90 pixels). Larger encoding noise indicates more uncertainty about the objects' locations and results in the objects being recalled with greater bias.

The "model chains" produced by this structured memory model converged towards remembering objects in tighter groups, with fewer, and more defined groups on the whole. Additionally, in the model chains, objects were increasingly organized into lines. In the subsequent analyses, we compare these model chain predictions to participants' actual performance.

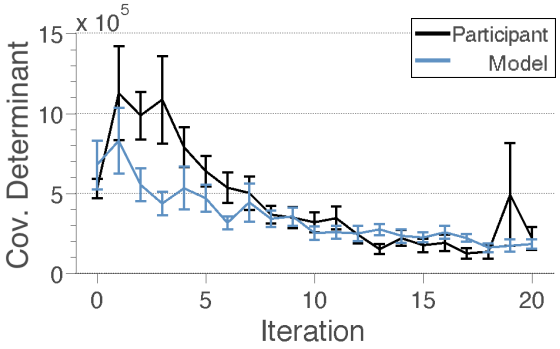

Figure 5. The determinant of the group covariance matrices. The black line indicates participants' responses and the blue line indicates model chains. Larger determinants indicate larger location dispersion. Locations were recalled increasingly close together.

Variance of groups. Objects were initially uniformly distributed in the display; did participants expect objects to be arranged more closely together? For each iteration, we measured the dispersion of objects within groups by finding the determinant of the locations' covariance matrix, where larger determinants indicate greater spread of objects within clusters. The chains of responses produced by humans showed the same decreasing within-cluster spread of objects (Figure 5) as we saw in the model chains (r=.80, 95% CI: .56–.91), indicating that participants recalled locations increasingly compactly within groups.

Number of groups. Given working memory's limited capacity, how many groups did participants remember? We estimated the number of groups that were evident at each step of the chain of human responses (Figure 6A). People reported objects in fewer groups in later iterations, consistent with the simulated model chains, (r=.89, 95% CI: .74–.95). Additionally, participants asymptoted around approximately five groups, slightly higher than, but comparable to previous studies of working memory capacity (Cowan, 2001).

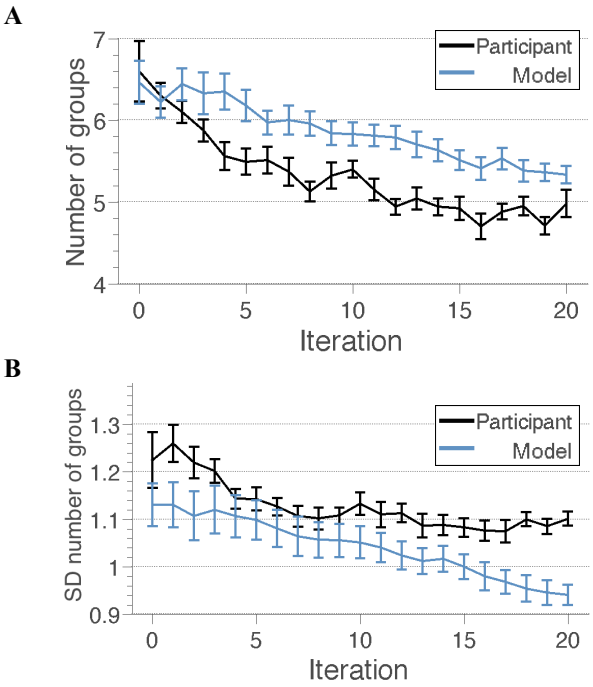

Figure 6. (A) The number of groups inferred by the clustering algorithm. (B) The posterior standard deviation of the number of groups. The black line indicates participants' responses and the blue line indicates model chains. The number of groups and the posterior standard deviation of the number of groups both decreased over time.

Additionally, as the within-group spread of objects, and the number of groups decreased, uncertainty about the number of groups in a given display decreased over simulated iterations with the structured memory model. We measured confidence in grouping structure as the standard deviation of the posterior distribution of the number of

groups in a display. The posterior standard deviation of the number of groups decreased over iterations of human chains (Figure 6B), similar to simulated iterations from our clustered memory model (r=.81, 95% CI: .58–.92). This suggests that, in addition to the number of groups decreasing, the distinction between different groups of objects became more pronounced.

How did participants encode lines?

Participants recalled objects in increasingly coherent and compact groups. Our Dirichlet clustering algorithm allowed us to infer whether these groups were Gaussian clusters or lines with orientations and lengths. Here, we use our clustering algorithm to characterize participants' prior expectations of linear groupings.

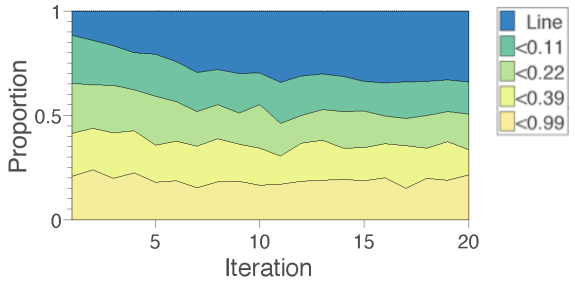

Figure 7. The proportions of groups recalled that were straight lines and Gaussian clusters. Clusters are divided into quartiles based on their eccentricity. Here, low eccentricities indicate less circular, more linear clusters. Participants organized more objects into lines over time.

Proportion of grouping types. What kinds of structures did participants encode over time? We used the Dirichlet clustering algorithm to calculate the proportions of the groups that were lines and clusters (Figure 7). We further subdivided the clusters into quartiles based on their eccentricity. Eccentricity measures how much the covariance of a cluster deviates from circularity, such that an eccentricity of 1 would indicate perfect circularity and an eccentricity of 0 would indicate clusters with zero width along the minor axis – in other words: lines. This allows us to measure how linear/circular groups were.

Participants increasingly grouped objects as lines consistent with the convergence produced in simulated model chains. For humans, the proportion of lines went from 13% to 34% and had a linear regression slope of .0087 (95% CI: .0076–.0098), consistent with the trend seen in the model chains (r=.91, 95% CI: .78–.96). This change seemed primarily to arise from regularizing greatly anisotropic clusters toward regular lines, suggesting that visual working memory relies on an expectation that objects are arranged linearly, consistent with the Gestalt principle of continuity.

Properties of lines. Both participants and the model simulations increasingly remembered objects as components of lines. Although the structured memory model inferred objects were organized into independent lines, it is possible that participants imposed further hierarchical structure on lines, grouping them together, and remembering their properties based on the ensemble statistics of *groups of lines*. We tested whether participants remembered lines according to their hierarchical structure by examining whether they recalled lines in the same trial with similar orientations (Figure 8A) and lengths (Figure 8B).

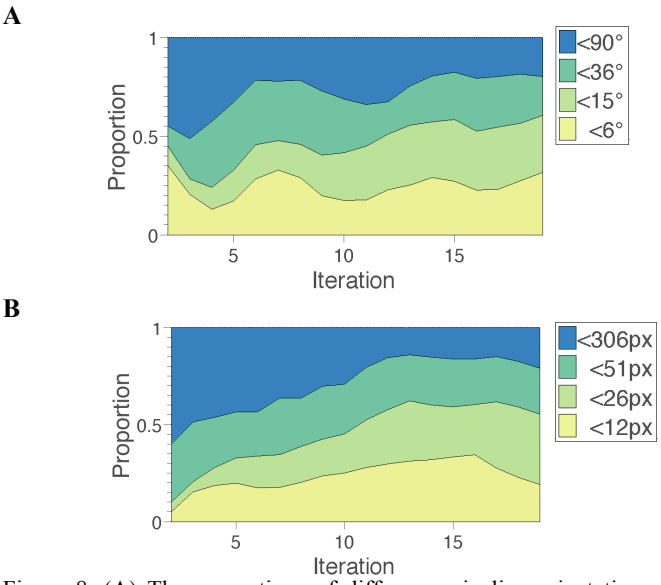

Figure 8. (A) The proportions of differences in line orientations and (B) the proportions of differences in line lengths. Due to the small number of lines in early blocks, in this figure we smoothed the proportions for each iteration by aggregating over the current, previous and next iteration. Line differences are organized into quartiles, such that bluer colors indicate larger differences. Participants became more likely to recall lines with similar orientations and lengths.

For each trial containing more than one line, we calculated the difference in feature values (orientation and length) for each pair of lines. For each iteration, we then aggregated all the feature differences across displays and chains, binned the differences into quartiles calculated from the entire study and found the proportion of differences in each quartile. Because the number of groups arranged in lines increased over time, later iterations reflect differences between more lines.

Participants remembered lines with increasingly similar orientations and lengths. Participants became more likely to recall lines with angular differences in the 1st and 2nd quartiles, which was confirmed by the positive slope of a linear regression (.038, 95% CI: .029–.047). Participants also increasingly recalled lines with length differences in the 1st and 2nd quartiles, as indicated by the positive linear regression slope (.038, 95% CI: .033–.043).

In contrast, the simulated model chains failed to predict participants remembering lines with similar orientations (r=-.30, 95% CI: -.65-.17) and lengths (r=.29, 95% CI: -.17–.65). This is not surprising, given that the structured memory model incorporates no method to integrate information across groups. Indeed, it provides evidence that,

unlike the model, participants remembered lines using ensemble statistics applied not only at the level of objects grouped into clusters, but also at the level of the groups.

Subjects also tended to remember more vertical lines. A v-test revealed that lines were on average 90° (vertical) and that this average did not reflect a uniform distribution of orientations ($v(1942)=129.65$, $p<.001$). This bias provides further evidence that subjects imposed higher-order structure upon groups. In this case, however, the structure came from the prevalence of vertical lines in natural scene statistics (Switkes, Mayer & Sloan, 1978).

These results indicate that participants encoded lines using their ensemble statistics, consistent with the Gestalt principle of similarity. The ensemble encoding of lines provides evidence that linear priors helped participants encode the basic stimuli as higher-level constructs.

Discussion

We used a Markov chain Monte Carlo with people visual working memory task to infer people's priors about the spatial arrangement of objects. Participants organized the objects into groups that were consistent with the predictions of a non-parametric Dirichlet process. Over iterations, objects became organized into more compact, stable groups, these groups became increasingly structured into lines, and these lines were grouped themselves to become more similar in orientation and length. The convergence towards organizing objects into lines and remembering those lines with similar orientations and lengths suggests that visual working memory's priors reflect classical Gestalt grouping principles such as continuity and similarity, respectively. Additionally, a model that used the hierarchical structure of objects grouped objects into lines similarly to participants, demonstrating that these priors facilitated the organization of objects into higher-level constructs. Notably, however, our cognitive model was unable to predict how people used the statistical structure of lines, raising questions about what are the units of ensemble encoding (Im & Chong, 2014).

In contrast to the clusters and lines observers encoded in our study, in the real-world observers frequently encode objects in complex shapes—we need only consider examples like stargazing for constellations or tealeaf reading. Likewise, a quick glance at responses in later iterations of our study (Figure 3) reveals perpendicular lines, winding contours and even structures like letters and shapes that suggest the use of long-term knowledge. Despite the heterogeneity and complexity of the patterns observers could have possibly used to remember objects, we found that a relatively simple model that encoded objects as components of clusters and lines was able to capture much of how people grouped objects in visual working memory.

Acknowledgements

This work was supported by NSF CPS grant 1239323.

References

Alvarez, G.A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*(18), 7345—7350.

Austerweil, J. (2014). Testing the psychological validity of cluster construction biases. In *Proceedings of the 36th annual meeting of the cognitive science society*. (pp. 122—127)

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., & Tenenbaum, J.B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, *120*(1), 85—109.

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.

Ferguson, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent advances in statistics*, *24*, 287—302.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6), 721—741.

Im, H.Y., & Chong, S.C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*(7), 663-676.

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247—254.

Orhan, A.E., & Jacobs, R.A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological review*, *120*(2), 297—328.

Orhan, A.E., & Jacobs, R.A. (2014a). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, & Psychophysics*, *76*(7), 2158—2170.

Orhan, A.E., & Jacobs, R.A. (2014b). Are performance limitations in visual short-term memory tasks due to capacity limitations or model mismatch. *Manuscript under review*.

Orhan, A.E., Sims, C.R., Jacobs, R.A., & Knill, D.C. (2014). The adaptive nature of visual working memory. *Current Directions in Psychological Science*, *23*(3), 164—170.

Sanborn, A., & Griffiths, T.L. (2007). Markov chain Monte Carlo with people. *Advances in neural information processing systems*, (pp. 1265—1272).

Sims, C.R., Jacobs, R.A., & Knill, D.C. (2012). An ideal observer analysis of visual working memory, *Psychological review*, *119*(4), 807—830.

Switkes, E., Mayer, M.J., & Sloan, J.A. (1978). Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. Vision Research, (13), 1393—1399.