UNIVERSITY OF CALIFORNIA, SAN DIEGO

Representations of Hierarchical Structure in Visual Memory

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Psychology

by

Timothy Franklin Lew

Committee in charge:

Professor Edward Vul, Chair
Professor Timothy Brady
Professor Jonathan Cohen
Professor Harold Pashler
Professor Zhuowen Tu

2017

The Dissertation of Timothy Franklin Lew is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2017

DEDICATION

To my family, friends, colleagues and all the people who have supported me over the past five years

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

There are so many people I'd like to thank who have helped support me while working on my research and this dissertation. But of course, I have to start with Ed who has just been a fantastic adviser and friend. I couldn't have asked for a wittier, more caring (and mischievous) adviser. I would like to request here that for my Finnish post-dissertation sword, I would prefer a lightsaber.

I also want to thank my committee, Hal Pashler, Tim Brady (the "original" vision Tim), Zhuowen Tu and Jonathan Cohen for their great feedback on my dissertation.

I am also grateful to all of my lab mates who have made my time at UCSD a blast—Drew Walker, Kevin Smith, Kristin Donnelly and Rob St. Louis—with inanity ranging from testing Banjo's cognitive reasoning skills to finding new places to hide terrifying stuffed animals.

My family and friends have been a huge part of my journey in graduate school. Mom, Dad, Lauren, Scrappy and Bailey, video chatting while you were back east always brightened the harsh San Diegan winters. And to my friends, thank you for pretending to understand the words that come out of my mouth.

Chapter 1, in part, is currently being prepared for submission for publication of the material. Lew, Timothy; Vul, Edward. The dissertation author was the principal researcher and author of this material.

Chapter 2, in full, is a reprint of the material as it appears in *Journal of Vision* 2016. Lew, Timothy; Vul, Edward. The dissertation author was the principal researcher and author of this material.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Lew, Timothy; Vul, Edward. The dissertation author was the principal researcher and author of this material.

VITA

2017	Ph.D. in Psychology, University of California, San Diego
2016	Data Science Fellow, Insight Data Science
2012-2016	Graduate teaching assistant, University of California, San Diego
2013	M.A. in Psychology, University of California, San Diego
2012	B.A. in *Cognitive Science* & *Philosophy,* magna cum laude, University of Pennsylvania

PUBLICATIONS

Timothy F Lew, Edward Vul, "Knowledge and use of price distributions by populations and individuals", *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 2016.

Timothy F Lew, Edward Vul, "Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions", *Journal of Vision*, 2015.

Timothy F Lew, Harold E Pashler, Edward Vul, "Fragile associations coexist with robust memories for precise details in long-term memory", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 2015.

Timothy F Lew, Edward Vul, "Structured priors in visual working memory revealed through iterated learning", *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 2015.

Jeremy R Manning, Timothy F Lew, Peter Li, Michael J Kahana, "MAGELLAN: A cognitive map-based model of human wayfinding", *Journal of Experimental Psychology: General*, 2014.

ABSTRACT OF THE DISSERTATION

Representations of Hierarchical Structure in Visual Memory

by

Timothy Franklin Lew

Doctor of Philosophy in Psychology

University of California, San Diego, 2017

Professor Edward Vul, Chair

Visual working memory possesses a limited capacity for information but people can use objects' statistical structure to help remember their features. If you know that your papers are scattered around your desk, for example, this constrains their possible locations (e.g. it is unlikely they are in the bathroom) and can help you remember specifically where each paper is on your desk. However, it is often uncertain what information visual working memory should summarize to aid recall later on. Is it sufficient to remember that the papers were near the desk? Or will you need to know

where they were relative to each other? My dissertation investigates what statistical structure visual working memory seeks to encode by (Chapter 1) revealing what visuospatial groupings people expect and tend to use, (Chapter 2) examining how people use those expectations to form structured memories of objects' groupings and (Chapter 3) evaluating the cost of using this grouping structure—what information is lost by encoding objects as components of groups. Overall, my dissertation reveals reveals that while exploiting the statistics of scenes introduces structured biases into memories, doing so enables visual memory to build accurate, multi-level representations of scenes.

Introduction

People's ability to successfully explore and interact with the world stands in stark contrast to visual working memory's limited capacity for information (Cowan, 2001). Visual working memory can only remember a small number of objects with limited precision (Bays & Husain, 2008; Zhang & Luck, 2008) and often has difficulty recognizing even large changes in scenes (Pashler, 1988; Rensink, 2002). People may be able to comprehend the world despite the limits of visual working memory by exploiting recurring statistical structure in the world (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan, et al., 2014). For example, an observer trying to remember the locations of people in a crowd might infer that individuals are organized into groups. Later on, the observer might have forgotten people's exact locations and compensate by remembering individuals' locations around their group centers. Thus, in lieu of forming an exact representation of the world, observers can encode a gist that captures the important features of a scene.

What patterns then does visual working memory aim to encode when remembering scenes? There are often many structured representations that can summarize objects' structure—Are these people huddled in a cluster or do they form a line? How much do I need to remember about each person's location? Additionally, how visual working memory encodes that structure will influence how accurately different patterns of stimuli are remembered and the types of errors memories accumulate over time. Here I investigate how observers choose what statistical structures to encode and how the format of the resulting structured representations influence forgetting.

Priors in visual memory

Observers can resolve uncertainty about what patterns to represent by relying on prior expectations from the real world. Our everyday experiences give us sophisticated knowledge about objects' colors (Bae, et al., 2014; Persaud & Hemmer, 2014), sizes (Hemmer & Steyvers, 2009) and visuospatial arrangement (Orhan, et al., 2014). For example, based on my experiences with contours I expect that nearby line segments should form one continuous line (Orhan, et al., 2014). In perception, people's priors appear to follow Gestalt principles—people group objects that are near each other (the principle of proximity), form continuous lines (the principle of continuity) or are similar (the principle of similarity) (Wertheimer, 1923; Froyen, Feldman & Singh, 2015). Knowing what groupings are frequent in the world can help observers choose between different forms of structured memories that are capable of representing the same stimuli.

When the arrangement of objects matches people's expectations, people can also remember scenes with greater fidelity. As objects are organized more consistently with an observer's expectations, an observer's structured representation of the objects will accurately capture redundant details and allow the observer to focus on encoding unique deviations (Brady, Konkle, & Alvarez, 2009; Sims, et al., 2012; Orhan, et al., 2014). If visual memory has a prior that objects are arranged as horizontal lines and encodes the objects as such, for instance, it can ignore the objects' y-positions and focus on encoding their x-positions. Conversely, in situations where the arrangement of objects does not match people's priors, such as laboratory experiments in which objects' features come from uniform distributions, imposing expectations can distort memories and impair the

fidelity of recall (Orhan, et al., 2014). What visuospatial patterns then do people expect and consequently remember with ease?

The structure of memory

Encoding objects as members of the same groups can efficiently compress information in memory—rather than remembering the location, color, size, etc. of each object, people can just remember a few values and assume the objects are similar. However, there are many ways observers might represent objects' grouping structures. Different encoding schemes may allow observers to better retain information about individual objects and/or their grouping structures and yield different patterns of degradation over time. How then do observers represent the grouping structure of objects?

People appear to encode objects as components of a hierarchical generative model (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013). In this scheme, people infer the ensemble statistics of objects (like the average location of objects) and use these ensemble statistics when they are uncertain about individual objects. When remembering the locations of people in a crowd, an observer might compensate for forgetting individuals' locations by recalling them shifted towards their group centers (Brady & Alvarez, 2011). Whereas encoding objects independently would result in the locations of objects drifting independently as they are forgotten, encoding objects in a hierarchical model should result in objects becoming increasingly drawn towards their groups over time.

Additionally, studies of spatial memory have suggested that people encode the relative positions of objects within groups. People may infer the hierarchical structure underlying objects and encode the positions of the objects relative to their ensembles (Gershman, Tenenbaum & Jäkel, 2016; Mutluturk & Boduroglu, 2014). Rather than remember the absolute position of a person, you may remember their position relative to their group (e.g.: "the guy in the red shirt is one foot northwest of the group's center"). In this scheme, people would encode objects' and ensembles' spatial locations relative to their parents in a tree-like structure rather than in absolute coordinates. Encoding objects in a relative position tree should also introduce distinct patterns of correlated errors over time: Because child objects are defined relative to their parents, as the parents decay in memory their children will inherit their errors.

My dissertation explores what statistical structures visual memory seeks to encode when representing a scene. In Chapter 1, I use an iterated learning task to examine how people prefer to group object in memory. Previous studies have sought to test specific hypotheses about grouping cues derived from classic Gestalt principles, such as proximity, continuity and similarity (Wertheimer, 1923; Froyen, Feldman & Singh, 2015). However, people may use grouping principles that researchers have not been able to explicitly account for due to their subtlety or complexity. Thus, I used an iterated learning task similar to a game of Telephone to allow people to naturally express what groupings they expect. Participants successively remembering and passing on the locations of objects recalled objects in more compact groups, in more linear arrangements, and they recalled lines with more similar angles and lengths. Furthermore, I found that only a model that represented not only groups of objects, but also groups of

linear groups, was able to capture these patterns. In all, our results demonstrate a new method for uncovering the full scope of visual memory's grouping principles and suggest that classical Gestalt principles may arise from the format of structured memories.

Given that visual working memory organizes objects into groups, in Chapter 2 I determine how people represented objects within their groups to best retain memories of objects and their structure. To precisely examine how visual working memory encodes different spatial patterns, I asked participants to recall the locations of objects arranged in several predetermined spatial clustering structures. Consistent with objects being encoded as components of a hierarchical generative model, participants remembered objects shifted towards their cluster centers, enhancing accuracy at the cost of introducing bias. Participants also had more difficulty recalling larger relative distances, suggesting that they encoded objects in a relative position tree—objects relative to clusters—and recalled relative positions with Weber noise. In this scheme, clustering reduced the magnitudes of relative distances. Both of these encoding schema enable visual working memory to exploit objects' clustering structure to improve the overall fidelity of memory.

In Chapter 3, I examine whether relying heavily on objects' hierarchical structure can also impairing memories of objects' idiosyncratic details. To examine the ramifications of encoding a hierarchical generative model, we asked participants to remember the locations of objects in clusters with varying encoding and delay times to test whether longer delay times (and greater uncertainty) increased bias towards cluster centers. To examine the effects of encoding a relative position tree, we asked participants to remember the locations of objects in non-circular groups and tested whether visual memory retained objects' relative but not absolute positions. Encoding objects in a

hierarchical generative model impaired participants' ability to distinguish previously studied scenes from scenes in which objects were shifted towards their clusters. Encoding the relative positions of objects introduced correlated rotational errors, even for objects in separate clusters. Although relying on objects' hierarchical structure can improve memory, doing so may distort memories of individual objects and facilitate errors consistent with the objects' overarching structure.

Altogether, my dissertation demonstrates how people's prior expectations about objects' statistical structural encourages different forms of structured representations. Expanding our understanding of people's priors can reveal not only what patterns are easy for visual memory to learn and remember but also when people will likely impose an incorrect or unintended patterns, introducing structured biases.

References

Bae, G.-Y., Olkkonen, M., Allred, S., Wilson, C., & Flombaum, J. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision , 14* (4).

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science , 321*, 851-854.

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

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

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 , 24* (1), 87-114.

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.

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian Hierarchical Grouping: Perceptual Grouping as Mixture Estimation. *Psychological Review , 122* (4), 575-597.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research , 126*, 232-241.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science , 1* (1), 189-202.

Mutluturk, A., & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, & Psychophysics , 76* (8), 2276-2285.

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., 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.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics , 44* (4), 369-378.

Persaud, K., & Hemmer, P. (2014) Interactions between categorical knowledge and episodic memory across domains. *Frontiers in Psychology* 5, 584.

Rensink, R. (2002). Change detection. *Annual review of psychology , 53* (1), 245-277.

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

Wertheimer, M. (1923). Laws of organization in perceptual forms. *A Source Book of Gestalt Psychology* .

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

Chapter 1 **Structured priors in visual memory revealed through iterated learning**

Timothy Lew and Edward Vul

Abstract

What hierarchical structures do people use to encode visual scenes? 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 nonparametric 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, 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 encoding of higher-level structures from ensemble statistics.

Introduction

Although visual working memory possesses a limited capacity for information, it can exploit statistical structure in the world to aid recall (Brady & Alvarez, 2011; Orhan, et al., 2014). For example, an observer trying to remember the locations of people in a crowd might infer that individuals are organized into groups. Later on, the observer might have forgotten people's exact locations and compensate by remembering individuals' locations biased towards their group centers (Lew & Vul, 2015). However, even though people spatially group objects, it is not always clear what structures people should encode: Are these people in the same group? Are these people huddled in a cluster or do

they form a line? What statistical structures should people encode to accurately remember a scene?

Typically, researchers examine what spatial groupings visual memory encodes by designing stimuli that test whether people use specific grouping strategies. In perception, this approach has allowed psychologists to identify a host of Gestalt grouping principles (Wertheimer, 1923; Froyen, Feldman & Singh, 2015). For example, people group elements that are near each other (the principle of proximity), form continuous lines (the principle of continuity) or are similar to each other (the principle of similarity). This research strategy has confirmed that visual memory also relies on these Gestalt principles. For instance, observers tend to recall objects as closer together (Orhan & Jacobs, 2013; Im & Chong, 2014; Lew & Vul, 2015) and more similar (Brady & Alvarez, 2011; Orhan, et al., 2014) than they originally were.

Building upon these findings, rather than test whether people possess *particular* priors, in the current study we adopted a data-driven design to discover the grouping structures people expect by virtue of the memory biases that arise in an iterated learning paradigm. We had participants reveal their grouping expectations by performing a task similar to a game of Telephone: each participant studied and recalled the locations of objects and then the next participant studied and recalled the previous participant's responses, and so on. In this kind of iterated learning task, participants successively filtering stimuli will yield responses increasingly resembling their prior expectations (Bartlett, 1932; Kirby, 1999; Sanborn & Griffiths, 2007; Kempe, Gauvrit & Forsythe, 2015)—in our case, the spatial groupings that participants expected.

Participants initially retained information about the displays but gradually introduced biases, resulting in the locations of objects drifting systematically over time. We assessed what kinds of structures were evident in subjects' reports by constructing a grouping algorithm (similar to Orhan & Jacobs (2013) and Froyen, Feldman & Singh (2015)) that infers whether participants organized objects into clusters and/or lines. Consistent with classical perceptual Gestalt principles (Wertheimer, 1923), we found that participants recalled objects in more compact groups (following the principle of proximity), in more linear arrangements (following the principle of continuity) and recalled lines with more similar angles (following the principle of similarity).

To identify what structural priors could explain the biases that emerge through iterated learning, we designed a suite of four hierarchical memory models and used them to simulate new iterated learning chains. We found that only a model that represents objects as parts of clusters and lines, uses multiple levels of representation—at the levels of individual objects, clusters and lines and groups of lines—and applies distinct priors to different levels was able to capture these patterns. Altogether, human prior expectations about visual structure encourages the formation of sophisticated, hierarchical representations that in turn introduce biases into visual memory.

Experiment

Participants briefly saw a set of circles on a computer screen and after a short delay clicked on the screen to recall where the circles had been. Critically, we showed the locations one participant reported as the stimulus to the next participant, thus producing an iterated learning chain. Based on the logic of iterated learning (Bartlett, 1932; Kirby,

1999; Sanborn & Griffiths, 2007), such a process will tend to converge to people's prior expectations, in our case yielding samples of the spatial structure people expect in images.

Participants

We gathered participants from the Amazon Mechanical Turk marketplace and rewarded participants with a base payment and a performance-based bonus. We allowed participants to perform multiple trials of our experiment for different initial displays, resulting in 1614 unique participants performing a total of 2000 experiment runs. Participants were not told that the stimuli they studied were another participant's responses.

Stimuli

Each display had a radius of 275 pixels and contained 15 identical grey circles, each with a radius of 10 px. In the first iteration of each chain, the locations of the circles were generated from a uniform distribution. The circles, however, could not overlap.

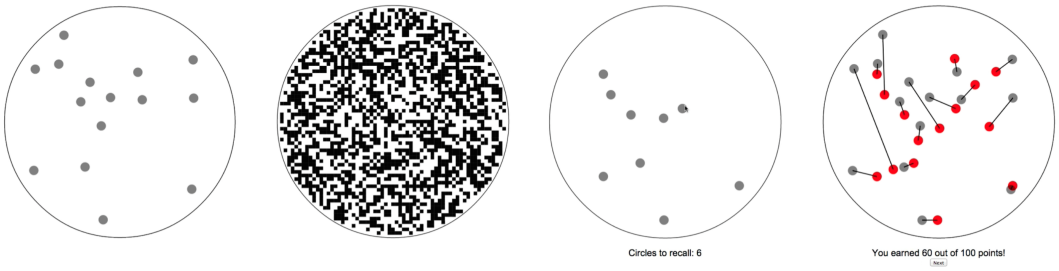

Figure 1.1 Example trial. (A) Participants saw 15 grey circles for 10 seconds followed by (B) a 1 second mask. (C) Participants then recalled the locations of all the circles and were told how many circles they had to recall. Participants could move around the circles until they were satisfied. (D) Participants 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.

Procedure

In each trial, participants observed the locations of the circles for 10 seconds (Figure 1.1A), followed by a 1 second mask (Figure 1.1B). Participants then recalled the locations of the circles by clicking the mouse (Figure 1.1C). 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 1.1D). 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.

Design

We generated 10 unique initial seed displays, each containing 15 circles with uniformly distributed locations. We set up 10 chains for each seed display and then ran each chain for 20 iterations. Thus, for each seed display there were 10 separate chains that began with the objects in the exact same locations and then diverged after the first iteration.

In each experimental run, participants first performed a randomly generated practice trial to familiarize themselves with the task. The second trial was our main test in which participants saw locations from the iterated learning chain (either the seed display of the chain for the first iteration, or the locations reported by the previous participant in

the chain in subsequent iterations). In the third trial, participants studied the first display of the chain, giving a measure of baseline performance (so participants who performed the first iteration of a chain would see the same seed display twice). The fourth trial was a randomly generated performance check: if a participant's 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 1.2 shows several example chains from our study (movies of all the chains are located on our website at www.evullab.org/dots.php).

(Kuhn, 1955)

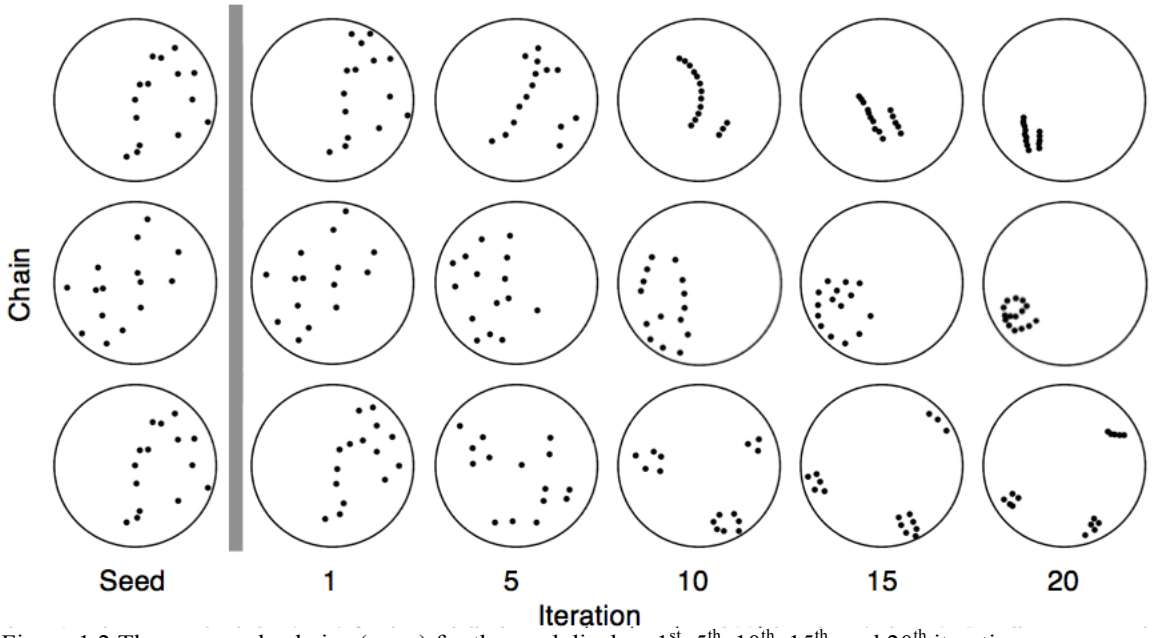

Figure 1.2 Three example chains (rows) for the seed display, 1st, 5th, 10th, 15th, and 20th iterations (columns). Grey lines separate the seed displays from the iterated trials. Circles are black in this figure for clarity (participants actually saw grey circles as in Figure 1A). Note that the 1st and 3rd chains begin from the same initial display and then diverge. Despite objects being initially uniformly distributed in the displays, participants gradually organized them into complex structures.

Results

Did participants' responses drift across iterations?

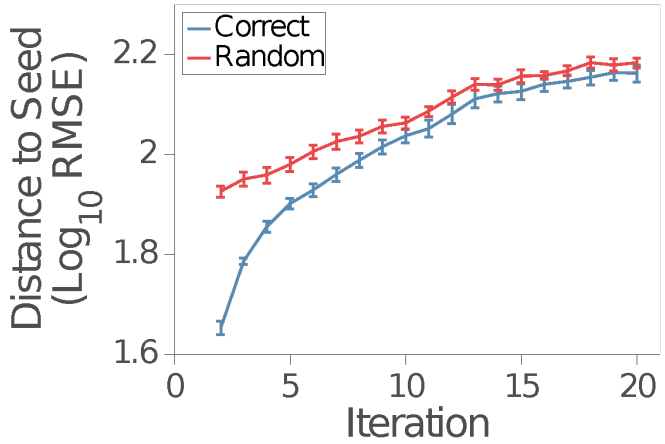

Figure 1.3 The distance between objects in each iteration and their locations in the initial seed displays. The lines indicate (blue) the distance between objects in a given iteration and the original seed and (red) the distance between objects in a given iteration and a random seed. Participants' responses initially resembled the seed displays, but became increasingly dissimilar over time. Error bars indicate SEM.

In our iterated learning task, the first participants remembered the one of ten randomly generated seed displays and later participants remembered the locations recalled by previous participants. To determine whether participants retained information about the initial displays, we measured the distance between objects in each iteration and their original seed displays. We used the Hungarian algorithm (Kuhn, 1955) to match objects from each iteration to either their original seed or a random seed and calculated the $\log_{10}$ root mean square error (RMSE) between objects (Figure 3). The $\log_{10}$ RMSE between objects and their seed locations increased over time (*Correct responses linear model slope: .049, 95% CI= .017—.081*), indicating that participants gradually lost information about the seed. Nevertheless, participants retained *some* information about the initial seed; even in the second half of iterations, responses were consistently more similar to the original seed than to random seeds (*paired t-test: $t(9)=19.7$, $p<.001$*).

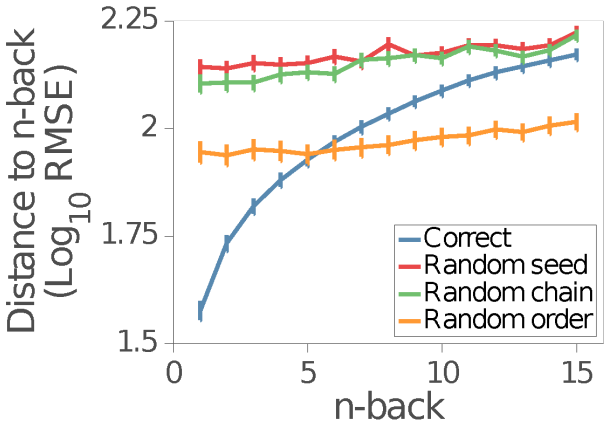

Figure 1.4 The distance between objects in a given iteration and the n-back iteration. The different lines indicate the distance between trials and the n-back (blue) displays in the same chain, (green) displays from the same seed but from different chains, (orange) the same chain with the order of iterations shuffled and (red) different seeds. Displays from the same chain in the correct order (blue) showed a clear drift with iteration distance as compared to all the shuffled controls. Error bars indicate SEM.

Information about the seed may have deteriorated due to each participant remembering objects somewhat inaccurately, resulting in locations gradually drifting over time. Similar to our previous analysis, we evaluated whether participants' responses drifted by measuring the distance ($\log_{10}$ RMSE) between objects from different iterations. Rather than compare each iteration to the original seed displays, however, we compared the locations of objects in each iteration to the n^{th} previous display (Figure 4). This comparison allowed us to measure how the locations of objects changed trial-to-trial.

The locations of objects grew more dissimilar as the number of iterations between two trials increased (*Correct responses linear model slope: .036, 95% CI= .029—.044*), demonstrating that participants' responses drifted over time. The displays diverged less when we shuffled the iterations (*Random order linear model slope: .0054, 95% CI= .0044—.0064*), providing further evidence that responses drifted sequentially from iteration to iteration. These patterns were also chain specific—when we compared displays to preceding displays from other chains and seeds, the responses were much

further apart (*Correct vs. Random chain paired t-test: t(14)=4.49, p<.001*, *Correct vs. Random seed paired t-test: t(14)=4.68, p<.001*). Altogether, participants appeared to retain some information about the initial seed displays but gradually introduced small errors, resulting in the locations of objects drifting over time.

Were participants biased towards grouping objects?

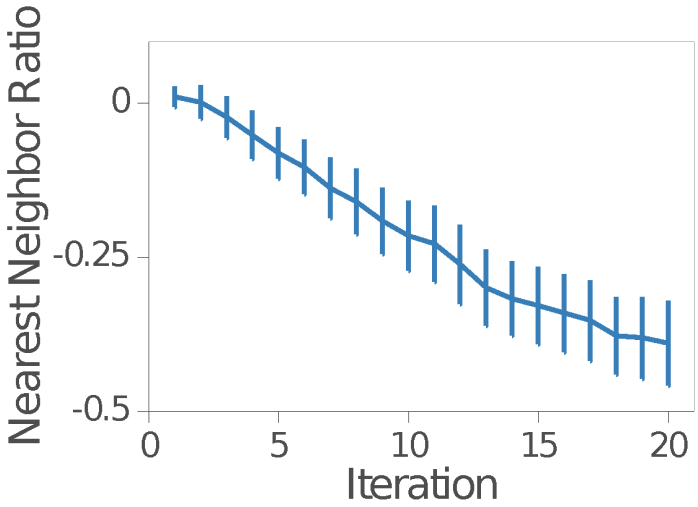

Figure 1.5 The log-ratio of nearest neighbor distance between objects recalled by participants vs. the locations of objects expected by independent drift over time. Positive and negative values respectively indicate that objects were more and less spread out than expected from a random distribution. Over iterations, objects became more closely clustered than expected by chance, suggesting that participants were biased to group objects together. Error bars indicate SEM.

Why did the locations of objects drift over time? One possibility is that participants remembered the locations of objects somewhat imprecisely and added independent noise each time they recalled the objects. Alternatively, participants may have grouped objects in memory and recalled the grouped objects closer together.

To evaluate these sources of errors, we compared the mean nearest neighbor distance between each object for the actual locations participants recalled to the expected distance if locations were recalled with independent noise. We compared participants'

responses to independent drift by calculating the log-ratio of participants' nearest neighbor distance to the nearest neighbor distance expected if objects drift independently following a homogeneous Poisson process (Figure 5). If participants recalled the objects independently, the log-ratio should stay at 0. Positive and negative log-ratios respectively indicate the objects are further apart and closer together than expected from a uniform random distribution.

The log-ratio decreased over iterations, indicating that participants recalled objects closer together than expected by independent drift (*nearest neighbor ratio linear model slope: -.022, 95% CI=-.024— -.021*). These patterns suggest that drift was the result of participants grouping objects and imposing compressive biases and not just independent noise.

What grouping structures did participants use?

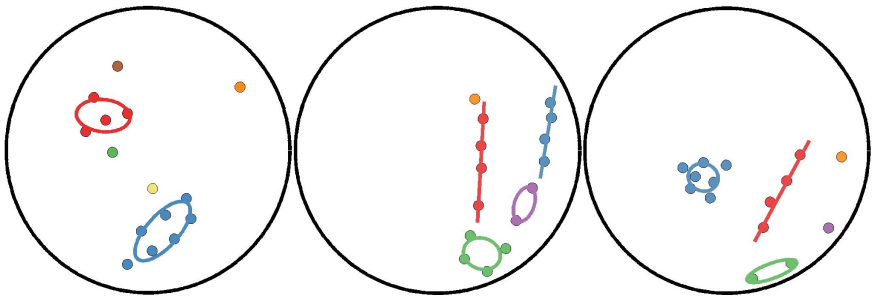

Figure 1.6 Examples of the Dirichlet grouping algorithm's inferred grouping for three trials. The grouping algorithm estimates the assignment of objects to groups (objects color-coded by group membership) and the parameters of the group structure: either a Gaussian cluster (represented by a covariance ellipse) or a line.

Non-parametric Dirichlet grouping algorithm. Thus far, we have demonstrated that the locations of objects drifted over iterations and that this drift was the result of participants remembering grouped objects close together. But what kinds of groups did visual

memory impose? To discover some of the grouping biases participants used, we designed a Dirichlet-process grouping algorithm (similar to Orhan & Jacobs (2013) and Froyen, Feldman, & Singh (2015)) that infers how participants grouped the objects in each iteration. 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 median location, length and orientation (Figure 6). We fit the grouping algorithm using a Gibbs sampler (Geman & Geman, 1984) which we ran for 800 iterations. Our analyses primarily use the maximum likelihood groupings inferred by the grouping algorithm. Further details about the grouping algorithm are located in Appendix 1.A.

In the following sections, we use our grouping algorithm to uncover a portion of the structures that participants converged towards. First, we confirm that participants used the groupings recovered by the grouping algorithm. We then demonstrate that the structured biases introduced by visual memory reflect the classical Gestalt grouping principles of Proximity, Continuity and Similarity.

Did participants group objects as predicted by the grouping algorithm? If participants grouped objects together per our grouping algorithm, then objects in the same group should have correlated errors (i.e., would tend to be misreported shifted 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 ($\mathbf{q}$) as:

$$q_{ij} = \frac{\boldsymbol{x_i x_j}^T}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}$$

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.

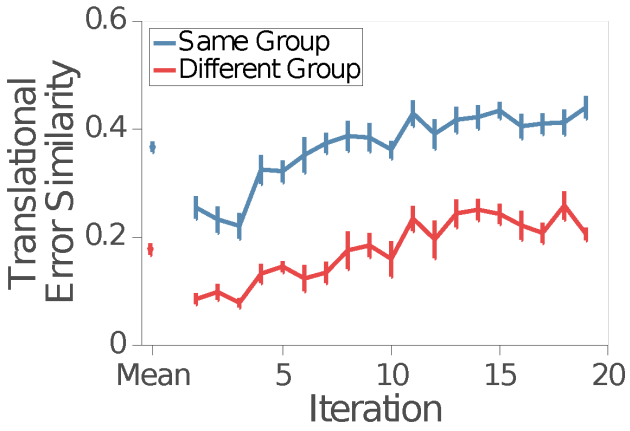

Figure 1.7 Translational error similarity for participants' responses. The points (Mean) indicate the error correlations averaged over iterations. The continuous lines indicate error correlations over iterations. Different Group (red) represents the error correlation for objects that the grouping algorithm inferred were in different groups, Same Group (blue) represents the error correlation for objects that the grouping algorithm inferred were in the same group. Errors were more similar for objects grouped together by the grouping algorithm. Error bars indicate SEM.

The error similarity of objects that our grouping algorithm grouped together was significantly greater than the similarity of objects in different groups (*Same vs. Different paired t-test: t(9)=35.21, p<.001*; Figure 1.7, Mean points), indicating that the grouping algorithm predicted the structure of errors in participants' responses, and therefore the display structure that participants inferred.

Did participants increasingly use the groupings inferred by the grouping algorithm? If participants converged towards grouping structures that were consistent with their priors, over multiple iterations grouping strength should increase, and the translational errors of items in the same group should grow more similar. As a result, over time the translational error correlation of objects in the same group should have increased. To test this prediction, we measured the translational error correlation for objects that the grouping algorithm inferred were in different groups and the same group at each iteration (Figure 1.7). The translational error correlation of objects that the grouping algorithm predicted would be in the same group increased over iterations (*Same group linear model slope: .010, 95% CI= .0074—.014*), demonstrating that participants became more likely to remember objects in coherent groups and relied on priors that encouraged the grouping of objects.

Prior for proximity: Did participants remember objects in more compact groups?

We first examined whether visual working memory possesses a prior analogous to the principle of proximity observed in Gestalt studies of perception (Wertheimer, 1923). The principle of proximity states that observers tend to group objects that are near each other. Similarly, visual working memory appears to expect grouped objects to be close together; for example, people often remember the locations of objects biased towards their center (Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013; Im & Chong, 2014; Lew & Vul, 2015). If this bias in visual working memory arises from a prior for proximity

similar to that observed in perception, we expected that participants would arrange objects in increasingly compact groups over time.

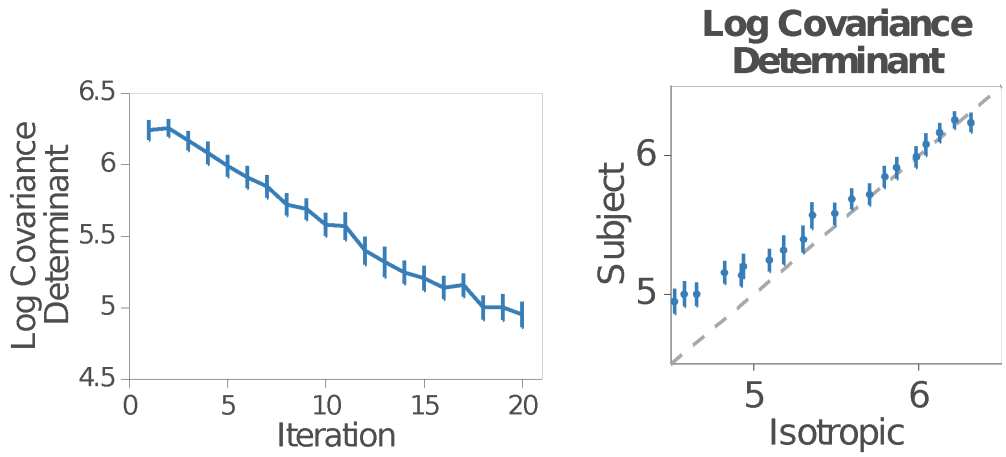

Figure 1.8 A) The log of the determinant of the group covariance matrices for participants. Larger log determinants indicate groups were more dispersed. Participants recalled locations increasingly close together. B) The dispersion of groups recalled by the isotropic clustering model vs. participants' recalled groups. The dashed grey line indicates equality. The isotropic clustering model predicted participants' convergence towards more compact groups. Error bars indicate SEM.

Behavioral. Following the Gestalt principle of proximity, participants should have recalled objects more compactly over iterations. For each iteration, we calculated the dispersion of objects within groups by finding the log of the determinant of the locations' covariance matrices. The determinant measures the magnitude of groups' dispersion such that smaller determinants indicate more compactly spaced objects within groups. The within-cluster spread of objects decreased over iterations (*Log covariance determinant linear model slope: -.074, 95% CI= -.070— -.78*) (Figure 8A). Thus, participants recalled locations increasingly compactly within groups, suggesting that visual memory encodes objects using a prior that resembles the principle of proximity.

Isotropic clustering model. Did encoding objects according to their structure cause participants to remember objects in increasingly compact groups? Participants may have encoded objects as parts of clusters and compensated for uncertainty about their individual locations by recalling them biased towards their cluster centers (Lew & Vul, 2015). To test this possibility, we designed a model that would emulate human behavior: inferring a grouping structure for a particular display, then noisily recalling those displays. Thus, by providing the output of this model to itself, we can simulate a whole chain of participants playing the iterated memory game. The objective of designing such a model is to ascertain whether the key features of human behavior are reproduced by iterated learning via a particular model of human learning—this logic has been used to study features of language evolution (Kirby, 1999; Griffiths & Kalish, 2007).

Our first human-learner model is rather simplistic and assumes that objects are arranged only in isotropic Gaussian clusters and then infers this grouping structure and recalls objects biased towards their cluster centers. The isotropic clustering model recalls objects biased towards their clusters, inversely weighted by the covariance of the clusters and the noise with which the objects are encoded (σ_o) (Brady & Alvarez, 2011; Lew & Vul, 2015). Intuitively, the more uncertain participants are about the locations of objects, the more they will rely on their memories of the objects' clusters and vice versa. When the determinant of Σ is large (indicating that clusters are very dispersed), the model will rely on its memories of the objects and exhibit little bias. When σ_o is large (indicating that memories of objects are noisy), the model will rely on the clusters' statistics and exhibit a strong bias. We fit two parameters for this model: α and σ_o (further model fitting details are contained in Appendix 1.B).

Over iterations, the isotropic clustering model remembered objects in more compact groups (*Log covariance determinant linear model slope: -.096, 95% CI= -.094— -.99*) and accurately predicted the dispersion of groups that participants used (*r=.99, p<.001; Participant vs. Isotropic model linear model slope: 1.29, 95% CI= 1.22—1.37*) (Figure 8B). This pattern suggests that participants encoded objects as members of clusters and recalled objects biased towards their clusters as they forgot the individual objects' locations. In this way, encoding and relying on the grouping structure of objects is sufficient to explain visual memory's prior for proximity.

Prior for continuity: Did participants remember objects arranged in lines?

We next examined whether visual memory possesses a prior analogous to the Gestalt principle of continuity (Wertheimer, 1923), such that observers remembered objects arranged in continuous lines. Constraining objects to be continuous may help compress stimuli in dimensions with statistical regularities (Orhan, et al., 2014). If objects are arranged in a vertical line, for instance, an observer does not have to remember their x-coordinates and can focus on encoding their y-coordinates more precisely. If visual memory relies on a prior for continuity, participants should go from recalling objects as parts of amorphous clusters to recalling objects arranged in highly constrained lines.

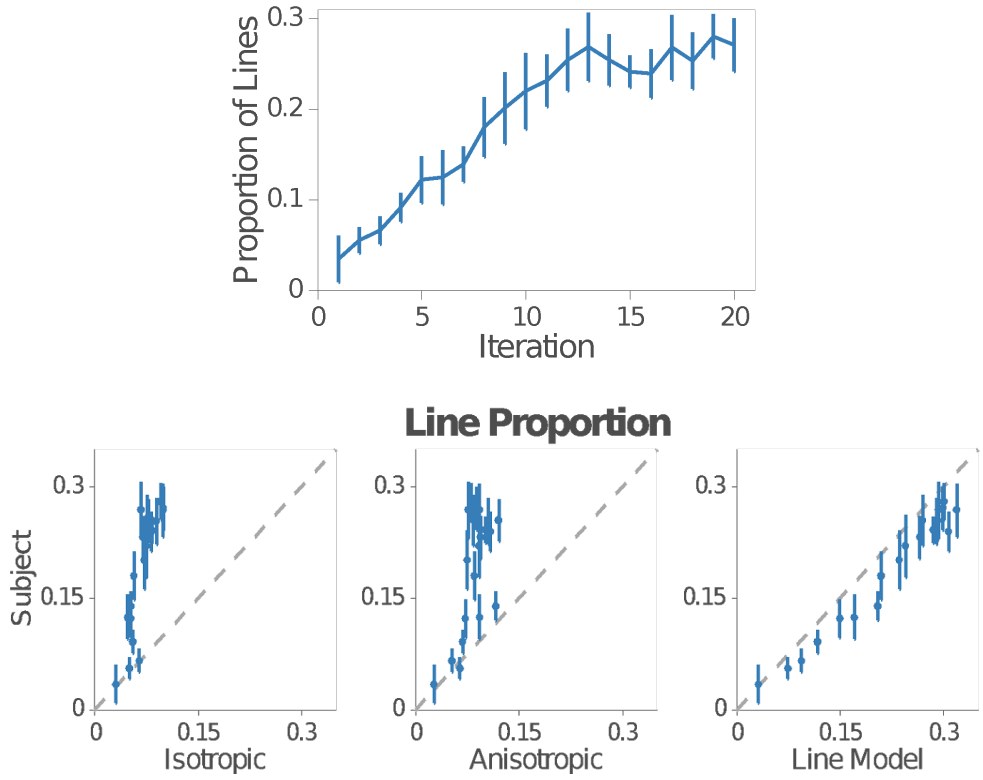

Figure 1.9 A) The proportions of groups participants recalled that were straight lines rather than Gaussian clusters. Participants organized more objects into lines over time. B-D) The proportion of lines participants (corresponding to blue in (A)) formed as a function of the proportion of lines formed by (B) the isotropic clustering model, (C) the anisotropic clustering model and (D) the line model. The dashed grey line indicates equality. Each point represents the proportion of lines at each iteration. The line model best predicted the proportion of groups arranged into lines. Error bars indicate SEM.

Behavioral. If participants relied on a prior resembling the Gestalt principle of continuity, they should have remembered an increasing proportion of groups as lines. We used the grouping algorithm to calculate the proportion of groups that were lines (Figure 1.9A). Participants increasingly grouped objects into lines. The proportion of lines had a linear regression slope of .013 (*95% CI= .011–15*), supporting convergence towards linear groups. The increasing proportion of linear groups suggests that visual working memory relies on an expectation that objects are arranged linearly, similar to the Gestalt principle of continuity.

Yet, there are different forms of prior expectations that could have yielded linear groupings. Simply grouping objects close together, like the isotropic clustering model, may have allowed nearby objects to coincidentally form lines. Alternatively, lines may have arisen from participants encoding anisotropic clusters that grew more eccentric over time. Finally, visual memory may have actually encoded linear structures in addition to Gaussian clusters. We next tested which of these structured representations best captured participants' tendency to organize objects into lines.

Isotropic clustering model. We first tested whether merely grouping objects into isotropic clusters and recalling objects biased towards their cluster centers could have resulted in the formation of lines. The isotropic clustering model grouped an increasing proportion of objects into lines over iterations (*line proportion linear model slope: .0030, 95% CI= .0024—.0036*) and was correlated with participants' behavior ($r=.86$, $p<.001$) (Figure 1.9B). However, the model systematically underestimated the proportion of lines and provided a poor regression fit (*Participant vs. Isotropic model linear model slope: .20, 95% CI= .14—.26*) suggesting that encoding objects as members of isotropic clusters was not the main cause participants forming lines.

Anisotropic clustering model. We next examined whether a model that encodes objects as components of anisotropic clusters could capture the increasing proportion of lines. Like the isotropic cluster model, we set Λ to be 0. Unlike the isotropic clustering model, we allowed the anisotropic clustering model to have clusters with covariance matrices, Σ, that were asymmetrical, such that the dispersion of objects varied along the axes of the

clusters. As a result, when the anisotropic clustering model recalls objects biased towards their group center, the extent to which objects are drawn towards the center will differ along the axes of the cluster. Objects will be drawn more strongly towards the minor axis of the cluster (that is, the axis with lower variance) than towards the major axis. We predicted that over multiple iterations this regularization towards the minor axis would yield lines (for in the limit, a cluster with an eccentricity of 0 would be indistinguishable from a line). We fit two parameters for this model: α and σ_o (further model fitting details are contained in Appendix 1.B).

Contrary to our expectations, the proportion of groups that the anisotropic clustering model inferred were lines increased only slightly over iterations (*line proportion linear model slope: .0023, 95% CI= .0008—.0037*) and still provided a lackluster regression fit (Participant vs. Anisotropic model linear model slope: .17, 95% CI= .072—.28) (Figure 1.9C). The lack of lines may have arisen from objects being weakly biased towards the center of the cluster along the major axis. Although for a given cluster the model recalls objects more strongly biased along the less variable minor axis, it appears that the weak bias along the major axis is sufficient to disrupt the organization of objects into lines.

Line model. Finally, we assessed whether representing objects as parts of clusters and lines could capture participants' behavior. The line model allows groups to be either anisotropic Gaussian clusters or lines (Λ can equal 0 or 1), using the same parameterization as our grouping algorithm. If an object is part of a Gaussian cluster, the model recalls it biased towards the cluster's center, just like the anisotropic clustering

model. If an object is part of a line, the model first noisily recalls the length of the line using a log-normal distribution with standard deviation σ_λ and the angle of the line using a von Mises distribution with standard deviation σ_θ. The model then recalls the object biased towards its position on the line, adjusted for the scaling and rotational transformations introduced by recall of the line.

Like the previous models, the extent to which the line model recalls an object biased towards its position on the line is determined by how precisely the object is encoded (σ_o) and the variance of objects around the line (like the grouping algorithm, we set σ_Λ=2.5). Unlike the previous models, because the line model recalls objects biased towards their positions on the lines rather than the center of the cluster, line-like groupings should not readily collapse into clusters. We fit four parameters for this model: α, σ_λ, σ_θ, and σ_o (further model fitting details are contained in Appendix 1.B).

The line model increasingly organized objects into lines (*line proportion linear model slope: .013, 95% CI= .011—.017*) and accurately predicted participants' tendency to recall objects arranged as lines (*Participant vs. Line model linear model slope: 1.06, 95% CI= .96—1.16*) (Figure 1.9D). The line model's success provides further evidence that participants possess a prior for continuity and, in addition to encoding objects as parts of clusters, utilized lines as a qualitatively distinct form of representation.

Prior for similarity: Did participants remember lines with more similar orientations?

Thus far, we have investigated whether visual working memory relies on priors for proximity and continuity when remembering the locations of individual objects. However, participants' prior expectations may have motivated not only the organization

of objects into compact clusters and lines, but also the organization of lines into groups of lines.

In particular, we expected that participants may have relied upon a prior comparable to the Gestalt principle of similarity to group lines (Wertheimer, 1923). The Gestalt principle of similarity states that observers tend to organize objects with similar features into groups. Consistent with this principle, visual memory appears to rely on the ensemble statistics of groups of lines when remembering the orientations of individual lines, recalling lines biased towards their average feature values (Sims, Jacobs & Knill, 2012). If visual memory also relied on a prior for similarity at the level of groups of lines, we expected that lines' angles would become more similar each iteration. Furthermore, this pattern would demonstrate that the priors of visual memory drive the formation of sophisticated, multiple-level representations of displays.

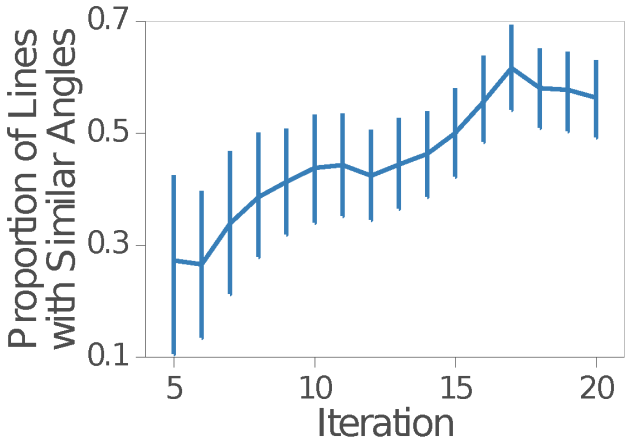

Figure 1.10 The proportion of line pairs recalled with angle differences less than the overall median angle difference over iterations. Due to the small number of lines in early blocks, we smoothed the proportions for each iteration using a sliding window of 5 iterations. Participants became more likely to recall lines with similar orientations. Error bars reflect 95% bootstrapped confidence intervals.

Behavioral. We tested whether participants remembered lines according to their hierarchical structure by examining whether they recalled lines in the same display with

increasingly similar orientations. For each trial containing more than one line, we calculated the differences in orientations for each pair of lines. For each iteration, we then aggregated all the orientation differences across displays and chains, performed a median split and found the proportion of differences in the lower half as function of iteration[1]. If participants remembered lines biased towards their ensemble statistics such that lines' orientations became more similar, then the proportion of small angular differences (smaller than the median) should have increased over iterations.

Participants remembered lines with increasingly similar orientations (Figure 1.10). Participants became more likely to recall lines with angular differences below the median, which was confirmed by the positive slope of a linear regression *(.021, 95% CI= .017–.025*). The increasing similarity of lines suggests that visual working memory relies on ensemble statistics applied not only at the level of individual elements, but also at the level of linear groups.

Line model. Although the line model does not integrate information across different line groups, lines may have nonetheless become more similar through other means (such as parallel lines being less likely to intersect and thus interfere with each other). However, the line model did not recall lines with more similar angles (*similar angle linear model slope: .0024, 95% CI= -.0013—.0062*) over iterations, and was a poor fit to participants' performance (*r=.25, p=.24; Participant vs. Line model linear model slope: .078, 95%*

[1] Because the number of groups arranged in lines increased over time, later iterations reflect more differences between lines. A linear regression on the number of line pairs had an intercept of 85.45 (*95% CI= 17.04–153.86*) and a slope of 30.87 (*95% CI= 23.80–37.95*).

CI= -.093—.25, Figure 1.11A) suggesting that the increasing similarity of features arose from another source.

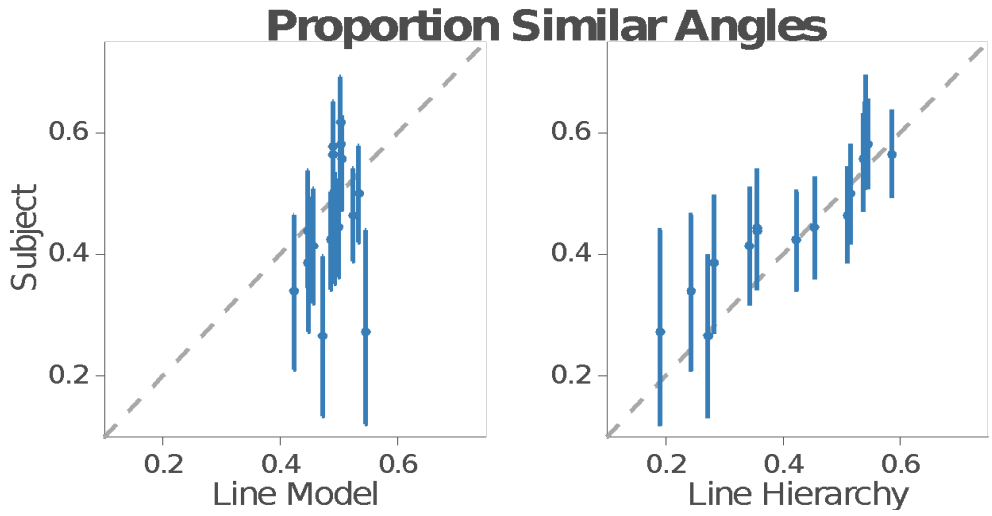

Figure 1.11 (A) The line model and (B) the hierarchical line model's ability to predict the similarity of angles recalled by participants. "Similar" indicates the the difference between two lines' orientations fell below the median angular difference. Dashed grey lines indicate equality. The hierarchical line model, but not the line model, predicted that participants would recall lines with more similar angles and lengths. Error bars reflect 95% bootstrapped confidence intervals.

Hierarchical line model. Based on the line model's limitations, we designed a model that relies on the ensemble statistics of lines during recall. The hierarchical line model is identical to the line model except when recalling objects grouped into multiple lines. When there are two or more lines, the hierarchical line model calculates the mean angle of the lines. The model then recalls the lines' features biased towards their ensemble statistics, inversely weighted by the noise with which the lines' features were encoded (σ_{λ} and σ_{0}) and the variance of the lines' features (Sims, Jacobs & Knill, 2012). The hierarchical line model then recalls the positions of the individual objects just like the line model. By recalling the features of lines biased towards their ensemble statistics, the hierarchical line model should recall lines with more similar orientations over iterations.

We fit four parameters for this model: α, σ_{λ}, σ_{θ}, and $\sigma_{\circ}$ (further model fitting details are contained in Appendix 1.B).

The hierarchical line model recalled lines with increasingly similar angles (*similar angle linear model slope: .026, 95% CI= .023—.030*) and this increase was strongly correlated with participants' responses (*r=.93, p<.001; Participant vs. Hierarchical line model linear model slope: 1.10, 95% CI= .85—1.36*, Figure 1.11B). These patterns suggest that participants' prior for similarity arose from their reliance on ensemble representations of line features. In addition, the hierarchical line model was able to capture how participants remembered objects in increasingly compact groups and lines (Table 1.1). Together, the hierarchical line model's ability to replicate these phenomena demonstrates that people represent scenes using multiple hierarchical levels of representation and apply distinct priors at each level.

Table 1.1. Correlations between participants and the models for the proximity, continuity and angle similarity analyses. Due to the small number of lines, we did not perform the angle similarity analysis for the Isotropic and Anisotropic models. * indicates p<.05, ** indicates p<.001. The hierarchical line model accurately predicted participants' performance in each analysis.

	Isotropic	**Anisotropic**	**Line**	**Hierarchical line**
Proximity	.99**	.93**	.91**	.87**
Continuity	.86**	.64*	.98**	.97**
Angle	--	--	.25	.93**

Discussion

Sequences of humans grouping and recalling the locations of objects resulted in structured patterns of drift over time. Participants increasingly remembered objects in

compact groups and encoded many of these groups as dense clusters and lines with similar lengths and orientations. Only a model that encodes objects as components of anisotropic clusters and lines and utilizes the ensemble statistics of lines was capable of replicating participants' behavior. These results suggest that people remember scenes with complex grouping biases that arise from encoding sophisticated, hierarchical representations.

Priors for the structure of visual memory

People used priors for proximity, continuity and similarity to help encode the locations of individual objects, resulting in objects being recalled biased towards their grouping structures. In this way, over many iterations reliance on objects' grouping structure transformed the arrangement of objects from uniformly distributed elements to arrays of sophisticated structures.

Reliance on structured priors distorted people's memories of objects, such that the final iterations of each chain were unrecognizable from their initial seed. In the real world, however, using such priors may improve the fidelity of visual memories. In our task the initial uniform distribution of object locations conflicted with expectations that objects would follow principles of proximity, continuity and similarity. In contrast, when objects are arranged consistently with people's priors, these priors can help people compensate for uncertainty about individual objects (Brady & Alvarez, 2011; Orhan, et al., 2014).

Furthermore, relying on structured priors may have aided performance by compressing information in visual working memory (Brady, Konkle, & Alvarez, 2009;

Sims, et al., 2012; Orhan, et al., 2014). While participants converged towards Gestalt priors when remembering the locations of objects, participants failed to organize objects into structured patterns in a separate perceptuomotor control experiment (Appendix 1.C). Visual working memory may have relied on these prior particularly strongly in order to compensate for its limited capacity. Whereas remembering a set of objects as falling along a horizontal line can help visual working memory focus on encoding their x-positions, perceptual or motor systems can simply refer back to the original stimulus.

From ensembles to objects

Instead of encoding lines independently, visual memory extracted the orientations of lines and recalled them biased towards their ensemble statistics. Only the hierarchical line model, which integrates feature information across lines, recalled lines with more similar angles over iterations. These findings suggest that rather than solely remember objects as members of groups, participants encoded objects as components of higher-order, object-like structures.

Representing objects as parts of structured memories allowed visual memory to exploit information from different hierarchical levels (Orhan & Jacobs, 2014). In the real world, inferring that a scene is a forest causes one to expect the presence of trees, seeing trees causes one to expect leaves, etc. (Orhan & Jacobs, 2014). Similarly, in our task the ensemble statistics of lines constrained the features of the lines and in turn the statistical structure of lines constrained the position of objects. Lines and groups of lines may have also been constrained by even higher-order statistical structure that we could not detect using our analyses, such as contours or everyday shapes (see Limitations below).

However, it is not always obvious what the units of storage and ensemble processing at a given level are. The patterns of convergence in our study demonstrate that observers' prior expectations about the world influence visual working memory's units of representation. One possibility is that people's priors interact with their observations yielding distributions of possible groupings (Gershman, Vul, & Tenenbaum, 2012; Froyen, Feldman & Singh, 2015). These groupings in turn may represent stimuli at the levels of parts, whole objects, groups of objects, etc. Future work may further examine how prior expectations determine how observers build from basic representational units like individual elements and ensembles to more sophisticated structures like parts of objects, whole objects and scenes (Palmer, 1977; Biederman, 1987; Orhan & Jacobs, 2014).

Limitations

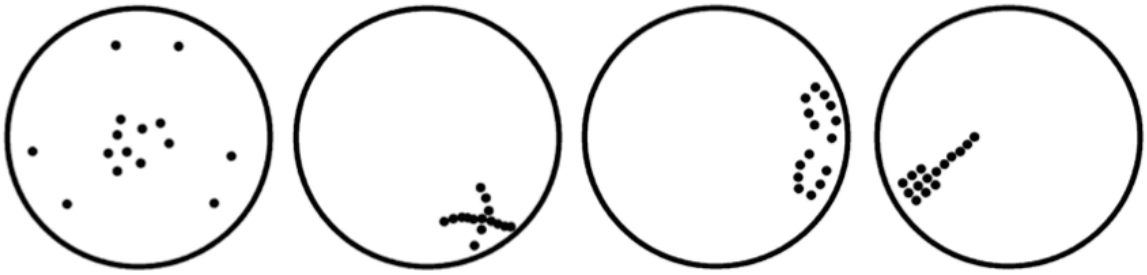

Figure 1.12 Examples of sophisticated structures that we were unable to account for using our model.

Our initial analyses revealed that locations drifted over time due to participants remembering objects close together in groups and our grouping algorithm subsequently suggested that many of those groups were clusters and straight lines. However, 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 1.2) reveals perpendicular lines, winding contours and even structures like letters and shapes that suggest the use of long-term knowledge (Figure 1.12 displays several more particularly complex structures). Although we were able to capture much of how people grouped objects in visual memory using a simple model, there is much more structure that we plan on examining in the future. Grouping algorithms that account for contours and bound shapes, such as the model used in Froyen, Feldman & Singh (2015), may reveal further structure in visual memory.

Additionally, our task may have encouraged participants to rely particularly heavily on the statistical structure of objects. First, long encoding times in our study may have allowed participants to verbally encode the locations of objects. An exceptionally apropos example is the possibility of participants verbally encoding objects arranged in the shape of the letter "e" (Figure 1.2, row 2, iteration 20). Participants may have used verbal labels to help recall grouping structures, increasing apparent biases. Eliminating verbalization using a verbal interference task may consequently decrease the amount of structure in participants' responses.

Second, it is unclear whether biases arose due to purely biases in visual memory or motor planning interacting with memory. The patterns of convergence in the perceptuomotor control condition suggest that motor actions alone did not cause structural biases (Appendix 1.C). Faced with a difficult reconstruction task involving planning motor actions, participants may have chosen to encode objects according to their grouping structure. Alternatively, participants may have remembered objects

without bias but introduced structural biases by transforming those memories into motor actions.

Summary

Using an iterated learning task, we revealed rich, sophisticated structure in visual working memory's prior expectations about the spatial arrangement of objects. Locations drifted as they were transmitted from person to person, gradually growing closer together. These structured patterns of drift in part reflected classical Gestalt grouping principles: Participants organized objects into more compact groups that were increasingly arranged as lines with similar orientations and lengths. Model comparison suggested that these Gestalt priors are the result of visual memory encoding objects as components of sophisticated, hierarchical representations of the world. Future studies must account for the influence of these priors when examining the fidelity of visual working memory.

Chapter 1, is currently being prepared for submission for publication of the material. Lew, Timothy and Edward Vul. "Structured priors in visual memory revealed through iterated learning." The dissertation author was the principal researcher and author of this material.

References

Bartlett, F. (1932). Remembering: A study in experimental and social psychology. *British Journal of Educational Psychology , 3* (2), 187-192.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review , 94* (2).

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 , 24* (1), 87-114.

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.

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian Hierarchical Grouping: Perceptual Grouping as Mixture Estimation. *Psychological Review , 122* (4), 575-597.

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.

Gershman, S. J., Vul, E., & Tenenbaum, K.B.. (2012). Multistability and perceptual inference. *Neural Computation , 24*, 1-24.

Griffiths, T., & Kalish, M. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science , 31* (3), 441-480.

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. *136*, 247-254.

Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals.* OUP Oxford.

Kuhn, H. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly , 2.1*, 83-97.

Lew, T., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision , 15* (4).

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., 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.

Palmer, S. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology , 9* (4), 441-474.

Sanborn, A., & Griffiths, T. (2007). Markov chain monte carlo with people. *Advances in Neural Information Processing Seminar* , 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.

Wertheimer, M. (1923). Laws of organization in perceptual forms. *A Source Book of Gestalt Psychology* .

Appendix

Appendix 1.A: Grouping algorithm

The grouping algorithm assumes that objects' locations are generated from a mixture of groups, $g=\{g_1, g_2, ..., g_n\}$. The number of objects in a group determines that group's mixture weight, or the probability that an object came from the group. Crucially, the more objects that are in a group, the more likely new objects will be assigned to that group. The concentration parameter, α, determines the likelihood of an object coming from a new group. We set α to .081 (see Appendix 1.B for details). Let G be a set containing the number of objects in each of the current groups, $\{G_1, G_2, ... G_n\}$. The group assignment of X_i object is generated by:

$$g^* \sim DP(G, \alpha)$$

Where *DP* indicates a Dirichlet process and g* is the new group sampled by the Dirichlet process. Crucially, the Dirichlet process allows the number of groups to vary. If an object is assigned to α, then a new group containing that object is created. If a group contains no objects, it is removed from g.

We allow groups to be either clusters or straight lines. Clusters are two-dimensional Gaussian distributions with anisotropic covariance matrices, defined by their centers (the means of their objects, μ) and covariances (Σ). A line is determined by its center (the median of the objects, *m*), angle (θ) and length (λ). To ensure that lines are thin, we set the noise of locations orthogonal to the line (σ_Λ) to 2.5. We treat a new group (when the Dirichlet process samples α) as a cluster with the mean and covariance equal to

the mean and covariance of all the objects. We also define Λ as an indicator variable that determines whether a group is a line or cluster. If objects x are assigned to group g_j, then whether the group is a cluster (Λ=0) or a line (Λ=1) is determined by the function:

$$\Lambda = \begin{cases} 1 & if\ g_i < 4, p = 0 \\ & if\ g_i \geq 4, p = \dfrac{\frac{1}{\lambda}N(d,0,\sigma)}{mvN(x,\mu_i,\sigma) + \frac{1}{\lambda}N(d,0,\sigma)} \\ 0 & if\ g_i < 4, p = 1 \\ & if\ g_i \geq 4, p = \dfrac{mvN(x,\mu_i,\sigma)}{mvN(x,\mu_i,\sigma) + \frac{1}{\lambda}N(d,0,\sigma)} \end{cases}$$

Where d is the distance between the objects in x and the nearest point on the line, N is the normal probability density function and mvN is the multivariate normal probability density function. This function states that: If there are less then 4 objects in the group, the group cannot be a line; this minimizes the possibility of the grouping algorithm spuriously inferring the existence of lines from randomly collinear objects. If there are at least 4 objects in the group, the probability of the group being a line is equal to the marginal probability of the objects coming from the line (vs. the cluster). Alternatively, the probability of the group being a cluster is the marginal probability of the objects coming from a cluster (vs. a line). Thus, the generative process for the location of an object, X_i, that has been assigned to group g_j is:

$$X_i \sim (1-\Lambda)\left(\mathrm{N}(\mu_i, \Sigma_i)\right) + \Lambda\left(\mathrm{m} + (\cos(\theta), \sin(\theta))\mathrm{U}(\frac{-\lambda}{2}, \frac{\lambda}{2}) + \left(\cos\left(\theta + \frac{\pi}{2}\right), \sin\left(\theta + \frac{\pi}{2}\right)\right)\mathrm{N}(0, \sigma_\Lambda)\right)$$

Where U indicates the uniform distribution. The first line indicates the location of an object that came from a cluster and the subsequent lines indicate (second) the location of the object along the line with (third) some orthogonal noise.

Appendix 1.B: Parameter fits

Table 1.B.1. Model parameter fits. α is the concentration parameter, σ_o is the encoding noise, and σ_λ and σ_θ are the length and angle encoding noise. Because the grouping algorithm does not recall the locations of objects, we did not fit any of the encoding noise parameters. Because the isotropic and anisotropic clustering algorithms do not possess lines, we did not fit the line encoding noise parameters σ_λ and σ_θ for these models.

	α	**σ_o**	**σ_λ**	**σ_θ**
Grouping algorithm	.081	---	---	---
Isotropic	.1	30	---	---
Anisotropic	.37	45	---	---
Line	.067	55	.15	.59
Hierarchical Line	.081	50	.14	.26

We fit several different parameters for our cognitive models and grouping algorithm. α is the concentration parameter, and influences how many groups the model infers. σ_o is the encoding noise and determines how noisily individual objects are recall

and how strongly they are biased toward their groups. σ_λ and σ_θ are the length and angle encoding noise and determine how noisily the lengths and orientations of lines are recalled and how strongly they are biased towards their ensemble statistics. Due to the difficulty of fitting the noise parameters for a large number of possible groupings, we chose to first fit the α parameter, use the fitted α to determine the maximum likelihood grouping for each display and then fit the noise parameters to the maximum likelihood grouping.

Fitting α. To fit the α concentration parameter, we selected a subset of displays from our main iterated learning experiment and asked a new set of participants to group the objects of these displays in a grouping experiment. For each model we then fit α to best predict participants' groupings.

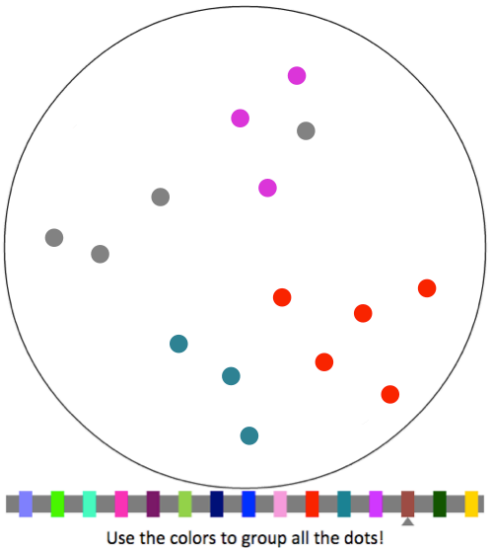

Figure 1.B.1 Example of the color grouping experiment. Participants initially saw a set of 15 grey dots. They then clicked colors in the palette to select them. Here, the participant has selected brown, indicated by the triangle below the brown patch. Once the participant selected a color, they clicked a circle to paint it.

We sought to maximize the heterogeneity of displays by selecting displays for which the grouping algorithm expressed varying levels of certainty. For each display, we measured the grouping algorithm's certainty in the grouping of objects by fitting the grouping algorithm[2] and calculating the standard deviation of the number of groups across the posterior distribution. Larger standard deviations indicate the grouping algorithm sampling a larger variety of possible groupings and being more uncertain. We then ranked the displays by the standard deviation of the number of groups and selected two displays for each 4th percentile. This yielded a set of 50 displays that we split into two sets of 25 displays, one display for each 4th percentile. Each participant saw only one of the two sets.

156 participants from the Amazon Mechanical Turk marketplace studied the first set of displays and 137 studied the second set; we rewarded participants with a base payment. The displays were identical to those from the main experiment, except for a color palette containing 15 colors (one for each object) below each display (Figure 1.B.1). We randomized the order of the displays for each participant and the order of the colors in the palette for each trial. We instructed participants to use the colors to group the objects. Participants clicked on a color to select it and then clicked on an object to paint it. Participants could distribute the colors however they wished (e.g., all the objects could be the same colors or all different colors) and could change the colors of the objects afterwards. However, we required that participants paint all the objects before moving on to the next display.

[2] As a first approximation, we set α=.1, finding that it resulted in seemingly reasonable fits

We varied the α parameter of each cognitive model to maximize how accurately the model predicted what groupings participants used. For each display, we compared the model's grouping to participants' grouping by calculating the probability that participants/the model reported each pair of objects as members of the same group and converting those probabilities into bits. We then found the absolute difference between the number of bits participants vs. the model needed to represent the grouping of objects. We used a grid search to find the value of α that minimized the absolute bit difference.

Fitting σ_o. σ_o indicates how noisily participants encoded the locations of individual objects and consequently how strongly objects were recalled biased towards their groups. We used the Hungarian algorithm (Kuhn, 1955) to match the locations of objects to participants' responses and for each model found the likelihood of the model recalling the object in that location given σ_o[3]. We used a grid search to find the value of σ_o that maximized the likelihood of the responses.

Fitting σ_λ and σ_θ. σ_λ and σ_θ indicate the noise with which the lengths and orientations were recalled, respectively. For each trial that the model inferred contained a line, we used the Hungarian algorithm to match objects that were part of the line in the studied display to the objects that were recalled. We then used principle components analysis (PCA) to fit lines to the studied and recalled objects and compared the lines' angles and lengths. Using separate grid searches, we then found the values of σ_λ and σ_θ that

[3] For simplicity, we assume that the grouping of objects in the studied display is preserved in the responses.

maximized the likelihood of the lengths and orientations of the PCA-fitted lines given the model inferred lines.

Appendix 1.C: Perceptuomotor experiment

Task. We ran a perceptuomotor experiment with 1399 unique participants to distinguish structure that arose from memory versus structure that arose solely from perception or motor planning. The structure of the perceptuomotor task was similar to the memory task—participants studied and reported the locations of objects and their responses were passed on to the next participant. However, instead of briefly studying and then recalling the objects, participants saw the display they were instructed to reconstruct the entire time.

Each trial, participants saw two environments side-by-side. The left environment contained the circles in the target locations (identical to Figure 1.1A) and remained onscreen for the entire trial. The right environment was empty and participants were instructed to copy the locations from the left environment onto the right environment (identical to Figure 1.1C). Once the participant finished, they received feedback in the right environment using the same criteria as in the memory task.

Error similarity. We used the grouping algorithm to infer what groupings participants used. Just like in our main memory experiment, we compared the translational error similarity of objects from the same vs. different groups (Figure 1.C.1). Objects from the same group were recalled with more similar translational errors than objects in different groups ($t(9)=19.64, p<.001$), suggesting that participants used the grouping structure of

objects in both the perceptuomotor and memory tasks. However, the translational error similarity of objects in the same group remained constant over time (*Same group error similarity linear model slope: -.004, 95% CI= -.0067— -.0027*), suggesting that participants did not increasingly infer groupings consistent with their prior expectations.

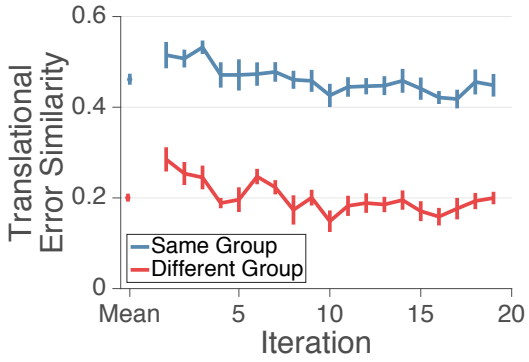

Figure 1.C.1 Translational error similarity for participants' responses in the perceptuomotor experiment (identical to Figure 1.7 for the primary memory task). Participants recalled objects in the same group with more similar errors, indicating that they relied on the groupings of objects. But, unlike in the memory task the translational error similarity of objects did not increase, suggesting that participants did not increasingly organize objects into groups resembling their prior expectations.

Compactness. We hypothesized that participants' prior for proximity arose from clustering in visual memory. However, perceptual or motor uncertainty may have also caused participants to rely on the grouping structure of objects. To determine whether the prior for proximity came from a source besides memory, we used groups inferred by the grouping algorithm and calculated the log of the groups' covariance determinants. Overall, the dispersion of groups in the perceptuomotor experiment did not predict the dispersion of groups in the memory task (*r=-.85, p<.001, Perceptual vs. Memory linear model slope: -6.15, 95% CI= -8.06— -4.26*) (Figure 1.C.2).

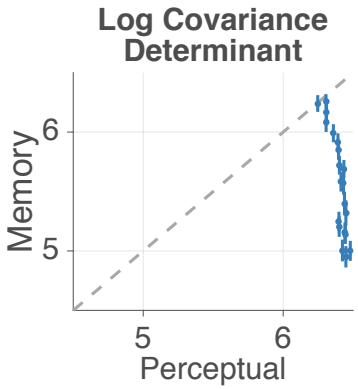

Figure 1.C.2 The log determinant of covariance of groups in the memory task as a function of the log covariance of groups in the perceptuomotor task. The dispersion of groups tended to remain constantly high in the perceptuomotor task compared to the memory task.

Proportion of lines. The perceptuomotor condition may have encouraged the formation of lines by allowing participants to form more complex patterns like triangles of squares or by virtue of lines arising from smooth motor planning. To examine whether participants in the perceptuomotor condition organized objects into lines we calculated the proportion of groups organized into lines vs. Gaussian clusters for each iteration. Participants tended to organize objects into fewer lines compared to participants in the memory task (*paired t-test: $t(19)=9.16$, $p<.001$;* Figure 1.C.3), suggesting that arranging objects into lines arose primarily from memory constraints. It is worth noting, however, that because our grouping algorithm only accounts for straight lines, it is possible the perceptuomotor condition allowed participants to use more sophisticated structure like contours (as in Froyen, Feldman & Singh (2015)).

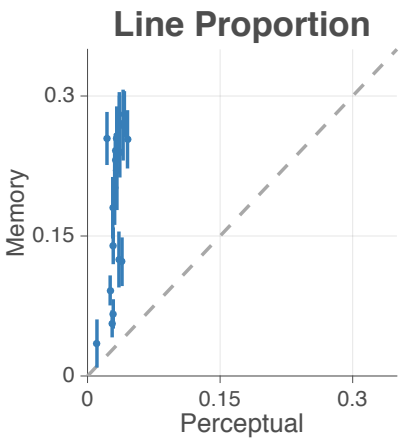

Figure 1.C.3 The proportion of groups organized into lines for participants in the perceptuomotor task vs. participants in the memory task. Participants recalled very few groups as lines in the perceptuomotor task.

Chapter 2 **Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions**

Timothy Lew and Edward Vul

Abstract

People seem to compute the ensemble statistics of objects and use this information to support the recall of individual objects in visual working memory. However, there are many different ways that hierarchical structure might be encoded. We examined the format of structured memories by asking subjects to recall the locations of objects arranged in different spatial clustering structures. Consistent with previous investigations of structured visual memory, subjects recalled objects biased towards the center of their clusters. Subjects also recalled locations more accurately when they were arranged in fewer clusters containing more objects, suggesting that subjects used the clustering structure of objects to aid recall. Furthermore, subjects had more difficulty recalling larger relative distances, consistent with subjects encoding the positions of objects relative to clusters and recalling them with magnitude-proportional (Weber) noise. Our results suggest that clustering improved the fidelity of recall by biasing the recall of locations towards cluster centers to compensate for uncertainty and by reducing the magnitude of encoded relative distances.

Introduction

Our visual working memory is limited in its ability to remember objects. In addition to remembering the individual elements of scenes, people may also extract the higher order structure of an image, such as elements average size (e.g., Ariely, 2001) or average location (e.g., Alvarez & Oliva, 2009). People can then use that statistical structure to help remember objects (Brady, et al., 2009; Brady & Alvarez, 2011; Sims, et

al., 2012). Knowing that your papers are scattered in a pile around your desk, for example, constrains their possible locations (e.g., it is unlikely they are in the bathroom) and can help you remember where individual papers are. Given that people appear to encode and utilize not only individual objects but also the higher order structure of objects, what is the format of structured memories?

In contrast to the traditional assumption that objects in visual working memory are encoded independently (Bays & Husain, 2008; Zhang & Luck, 2008; Anderson, Vogel & Awh 2011; for review see Ma, Husain & Bays, 2014), recent studies have demonstrated that memory exploits the statistical structure of scenes. Specifically, people infer the ensemble statistics of objects (like the average location of objects; Ariely, 2001; Alvarez & Oliva, 2009) and combine these ensemble statistics with uncertain estimates of individual object properties (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013). This encoding strategy can be described as reliance on a hierarchical generative model: people infer that object features are drawn from a distribution of features, and make uncertain inferences accordingly. In our desk example, this would imply that if you do not know exactly where a paper was, you may recall it as closer to the center of the pile to compensate for your uncertainty; although this strategy will yield some bias in in your estimate of the location, it will decrease variance, and thus improve overall memory fidelity.

The structure of multiple objects may also constrain the individual constituent objects more rigidly into multi-object chunks (Miller, 1956; Cowan, 2001; Brady & Tenenbaum, 2013). Chunking accounts tacitly assume that an inferred chunk completely constrains its subparts (e.g., encoding FBI fully determines its constituent letters). Thus,

chunking is classically considered to be a fixed memory structure (what we might call hard chunking) such that people remember only the chunk, and nothing about its constituent elements. However, if this encoding strategy is softened to allow some information to be preserved about the constituent elements of a chunk (soft chunking), such an account is consistent with encoding a hierarchical generative model that probabilistically constrains individual elements.

Additionally, studies of spatial memory suggest that people encode the relative positions of objects: Rather than remember the absolute position of a paper, you may remember its position relative to your desk (e.g., the paper is one foot northwest of your desk; Huttenlocher, Hedges & Duncan, 1991; Hollingworth, 2007). This relative encoding may be adapted to accommodate hierarchical structures via an assumption that people encode the relative discrepancy between features of individual objects and the average features of the ensemble. This relative encoding view is consistent with vector-summation models of multi-object motion parsing (Johansson, 1973; Gershman, JŁkel & Tenenbaum, 2013), and spatial positions (Mutluturk & Boduroglu, 2014). Intuitively, instead of remembering the locations of your papers relative to your desk, you may remember the locations of individual papers relative to the centroid of all the papers.

Thus, the space of possible structures that people might use to encode objects can be considered along several dimensions: (a) do people encode individual items with no information about their structure (independent encoding)? Or do they only encode the structure, losing all information about constituent elements (hard chunking)? Or something in between such that the overarching structure informs individual object features (hierarchical generative model or soft chunking encoding)? (b) Insofar as people

encode both higher order structure and individual element features, are these both encoded in absolute terms and inform one another probabilistically (absolute encoding), or are objects in the hierarchy encoded relative to their parent (objects relative to their ensembles and ensembles relative to cluster groups), such that object properties are ascertained by accumulating relative offsets in the hierarchy (relative encoding)?

Here we evaluate these dimensions of visual memory structure by asking people to remember and report the locations of objects arranged in different spatial clustering structures. Subjects recalled objects more accurately when they were arranged in fewer clusters that each contained more objects separated by smaller relative distances. To directly evaluate the format of subjects structured memories, we compared human behavior to that of three cognitive models-a hard chunking model, a hierarchical generative model and a relative position model. The relative position model best accounted for human performance, followed closely by the hierarchical generative model, with the hard chunking model missing key aspects of human behavior. Our results demonstrate two compatible ways in which hierarchical encoding improves the fidelity of visual working memory. First, objects are biased towards their ensemble statistics to compensate for uncertainty about individual object properties. Second, objects are encoded relative to their parents in the hierarchy, and relative positions are corrupted by Weber noise[4], such that larger relative distances yield greater errors.

Experiment

To distinguish different hierarchical encoding strategies that people may use, we

[4] In this study, Weber noise refers to errors that are normally distributed in log space.

asked subjects to report the positions of objects arranged in different clustering structures. Different encoding strategies will yield distinct patterns of errors across scenes that varied in the number of objects and the number of clusters in which they were arranged. Thus, we then examined if subjects responses across different types of environments were consistent with different forms of structured encoding.

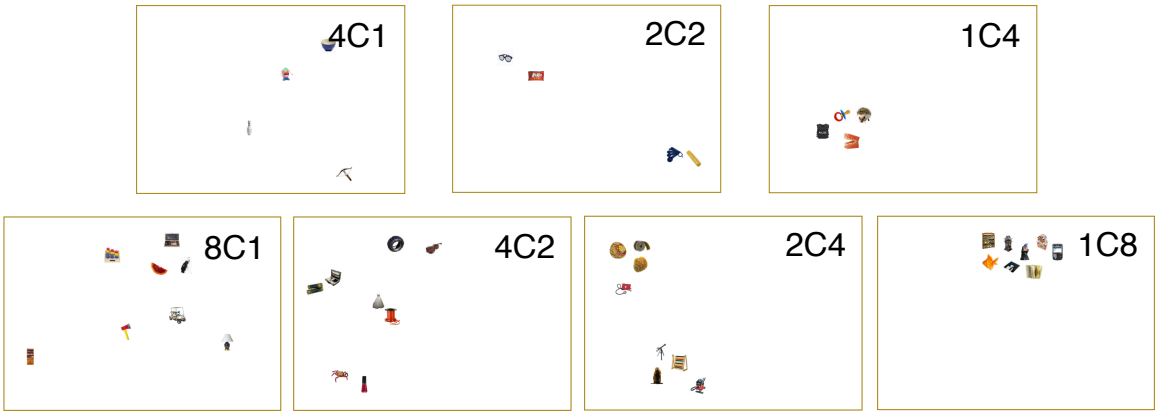

Figure 2.1 Examples of environments from each of the clustering structures. From left to right, each row is arranged in order of increasing clustering (clusters contain more objects). For this figure, a label indicating each environment's clustering structure is superimposed. Labels are read 4C2=4 clusters each containing 2 objects. Images of objects from Brady, et al. (2008).

Methods

Participants. Thirty-five students from the University of California, San Diego participated for course credit.

Stimuli. We generated 70 environments, each containing objects arranged into different clustering structures. We selected 440 images from Brady, at al. (2008) for the objects.

Although we did not control how much objects varied perceptually and semantically, we made sure each object type was unique (e.g., there was only one bicycle, clock, etc.). The dimensions of the environments were 700 x 1000 pixels. Each subject saw the same environments, but in a random order.

Each environment had one of seven clustering structures: 4 clusters each containing 1 object (4C1), 2 clusters containing 2 objects (2C2), 1C4, 8C1, 4C2, 2C4, 1C8 (Figure 2.1). We generated the locations of the clusters and objects by selecting cluster centers from a uniform distribution across the entire environment and then sampling object locations from each center using a two-dimensional isotropic normal distribution (SD=45) with the restriction that objects could not overlap. There were ten unique environments for each clustering structure, for a total of 70 environments.

Procedure. Subjects studied the 4-object environments (4C1, 2C2 and 1C4) for 4 seconds and the 8-object environments (8C1, 4C2, 2C4 and 1C8) for 8 seconds. After a 1 second pause, subjects saw an empty environment with the objects located at the bottom of the screen and had unlimited time to place the objects in their correct locations by clicking and dragging with the mouse. Our analyses focus on the reported spatial locations of all the objects in a display.

Results

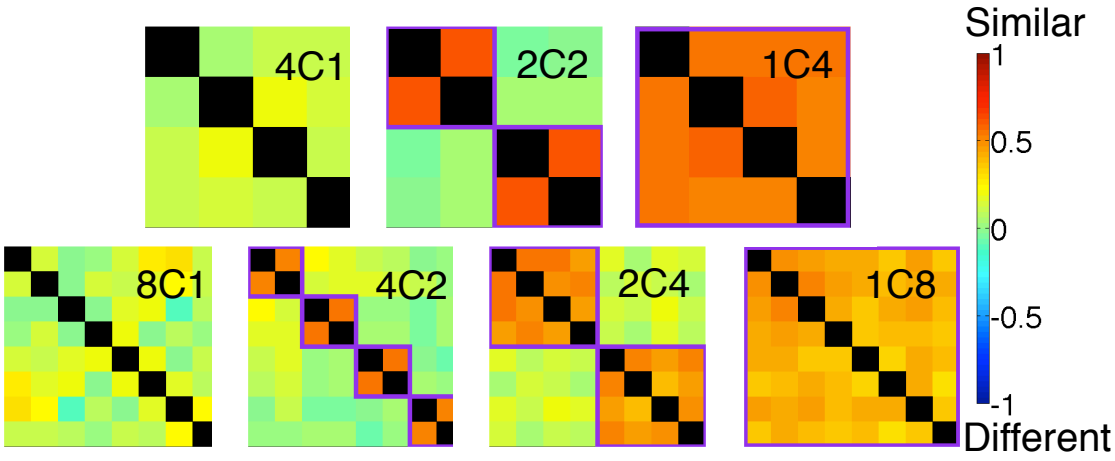

Figure 2.2 Error similarity heat maps with labels indicating the clustering structure superimposed. Warmer colors indicate more similar errors. Each square represents the error similarity between two different objects. Objects in the same cluster are outlined in purple. Objects in the same cluster were recalled with more similar errors

Did subject encode objects according to their clustering structure? If subjects did encode and utilize the clustering structure of objects instead of independently encoding objects, the errors for objects in the same cluster should be more similar (in the same direction) than expected by chance. We defined the similarity of the errors (q) in reporting the locations of two objects as:

$$q_{ij} = \frac{\boldsymbol{x_i}\boldsymbol{x_j}^T}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}$$

Where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are vectors containing the spatial translational error of the two objects' reported locations. The numerator is the projection of the translational error vectors, with positive values indicating vectors in the same direction and negative values indicating vectors in the opposite directions. The denominator normalizes the numerator such that q falls between -1 and 1. Thus, if the recalled locations of two objects were both shifted in

exactly the same direction *q* would be 1, if they were shifted in orthogonal directions *q* would be 0, and if they shifted in opposite directions *q* would be -1.

We calculated the translational error similarity (*q*) of objects in the same cluster for each environment (Figure 2.2). We excluded environments without clustering (4C1 and 8C1) from this analysis. For all clustering structures, subjects recalled objects in the same cluster with more similar errors than expected by independent encoding (*smallest t-value, t(34)=16.05, p<.001*). Subjects did not appear to encode the objects independently and instead used the clustering structure of objects.

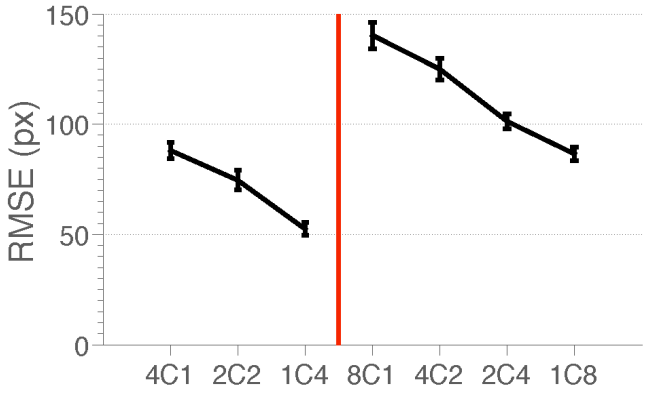

Figure 2.3 Raw performance measured in root mean square error (RMSE) for each of the clustering structures, arranged in order of increasing clustering. The red line separates the 4-object conditions from the 8-object conditions. Error bars indicate SEM. Performance improved as objects were arranged in fewer clusters containing more objects.

How did clustering structure affect recall fidelity? If subjects encoded objects independently, then clustering structures should not have affected how accurately subjects recalled locations. We assessed the effect of clustering structure upon the fidelity

of recall by calculating the root mean square error (RMSE[5]) of subjects' responses (Figure 2.3). We used a mixed effects model that included the number of objects, the number of clusters and their interaction as fixed effects and subjects as random effects to test whether object load and clustering structure affected recall.

RMSE was lower in the 4-object conditions compared to the 8-object conditions (*t(241)=12.47, p<.001 for the linear effect of number of objects*) and decreased as the number of objects in each cluster increased for both the 4-object and 8-object conditions (*t(241)=16.95, p<.001 for the linear effect of number of clusters*). Post-hoc Tukey's honest significant difference (HSD) pairwise comparisons confirmed that performance improved with every increment of cluster size in both the 4-object conditions (*smallest difference: 13.30, 95% confidence interval=3.69—22.92, p=.0042*) and the 8-object conditions (*smallest difference: 14.71, 95% confidence interval=3.55—25.88, p=.0046*). The decrease in RMSE with increasing cluster size seems constant across the 4- and 8-object conditions (*t(241)=.31, p=.76 for the interaction of the number of objects and the number of clusters; i.e., the difference in slope of RMSE as a function of number of clusters*). The effect of clustering structure on performance suggests that subjects did not encode the objects independently and that subjects used clustering to help remember objects more accurately.

Error model. Thus far, we have demonstrated that subjects did not encode objects independently. Given that subjects appeared to use the clustering structure of objects,

[5] We calculated RMSE using the formula $RMSE = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ where (x_1, y_1) and (x_2, y_2) are the true location of the object and the subject's reported location, respectively.

how did that structure constrain the locations of objects? Did subjects encode objects using hard chunking, a hierarchical generative model and/or a relative position tree? These encoding models predict different levels of reliance on (and bias towards) objects' hierarchical structure and different patterns of noise. To determine what type(s) of structured encoding subjects' errors were consistent with, we constructed an error model that estimates the extent of errors due to misassociations, bias and noise.

First, subjects may have had difficulty remembering which objects were in which locations. We estimated the probability of correctly matching an object to its location, p_T, and the probability of making a misassociation between an object and another object's location, $p_M = 1 - p_T$. The probability of misassociating to a particular location then was $\frac{p_M}{n-1}$, where n is the number of locations. To determine exactly which location each object was misassociated to, we assumed a bijective mapping of objects to locations (f), such that only one object could be paired with each location. $f^{-1}(i)$ denotes the inverse mapping from locations to objects.

Second, subjects may have been uncertain about objects' locations but used their memories of cluster locations to inform their responses. This would have resulted in objects being drawn towards their clusters. We accounted for two types of such "regularization" bias: the degree to which clusters are drawn towards the global centroid of all objects (cluster-to-global bias, β_c) and the degree to which objects are drawn towards their cluster centers (object-to-cluster bias, β_o). Here, a bias of 0 indicates the object/cluster is unbiased and a bias of 1 indicates the element is drawn completely towards its parent.

To parameterize how the locations of objects would be shifted by these sources of bias, we decomposed the true locations of objects, *t*, into their relative positions and then weighted the relative positions by the bias parameters. The decomposition of the true locations yielded a relative position tree in which the locations of objects were represented relative to their clusters (*x*), the locations of clusters were relative to the global centroid (*c*), and the global centroid (*g*) was the mean of the true locations (*t*). Conditional on the mapping $f(i)$ of the true locations *t* to response locations *s*, the position of an object *i*'s cluster relative to the global center was defined by:

$$c_i = C_{M(f(i))} - g$$

where *M()* maps objects to the clusters of which they are members and *C* is the absolute position of the cluster center, calculated by averaging the locations of all objects in that cluster. Similarly, the positions of objects relative to their clusters were defined by:

$$x_i = t_{f^{-1}(i)} - c_i - g$$

We then weighted the relative positions of clusters and objects by the cluster-to-global bias (β_c) and the object-to-cluster bias (β_o), respectively. Thus, the biased absolute positions of an object, b_i, were:

$$b_i = g + (1 - \beta_c) * c_i + (1 - \beta_o) * x_i$$

Finally, subjects may have remembered locations with some imprecision. To account for this, the model includes three levels of spatial noise that might induce correlations in errors across objects: that which is shared globally across all object locations (σ_g), for locations within the same cluster (σ_c) and individual object locations (σ_o). This decomposition of object positions induces an expected correlation structure on

the errors in reporting individual objects, which can be parameterized with a covariance matrix, Σ, of the form:

$$\Sigma_{i,j} = \begin{cases} {\sigma_g}^2 & i \neq j \;\&\; M(f(i)) \neq M(f(j)) \\ {\sigma_c}^2 + {\sigma_g}^2 & i \neq j \;\&\; M(f(i)) = M(f(j)) \\ {\sigma_o}^2 + {\sigma_c}^2 + {\sigma_g}^2 & i = j \end{cases}$$

where the three conditions reflect (in order) error covariance shared by all objects, error covariance for objects in the same cluster and error variance for individual objects.

Let Θ be the set of parameters $\{p_M, \beta_c, \beta_o, \sigma_g, \sigma_c, \sigma_o\}$. Altogether, for each environment, the likelihood of a set of responses given the targets and parameters was:

$$LIK(s|t, f, \Theta) = ({p_T}^{n_T}) \left(\frac{p_M}{n-1}\right)^{n_M} \mathcal{N}(s|b, \Sigma)$$

where s denotes the response locations, n is the number of objects, n_T is the number of objects correctly mapped to their locations by f, and n_M is the number of objects incorrectly mapped to their locations by f. We estimated these parameters (f, p_M, β_c, β_o, σ_g, σ_c, σ_o) for each environment across subjects using a Markov chain Monte Carlo algorithm (see Appendix 2.C for more details concerning our Markov chain Monte Carlo algorithm and Appendix 2.D for all parameter fits).

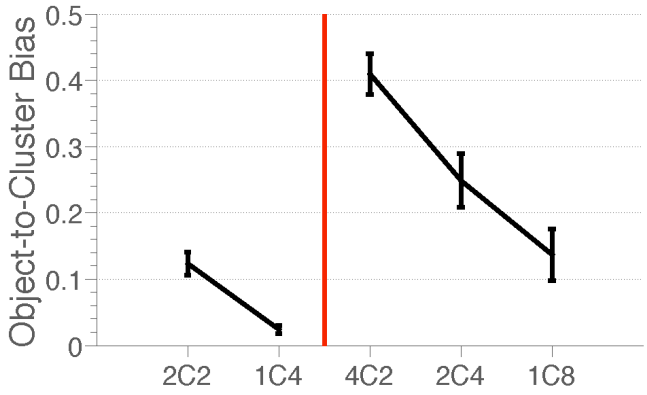

Figure 2.4 The extent to which objects were drawn towards their clusters (β_o) for each clustering structure. Larger object-to-cluster bias indicates objects are drawn more towards their clusters. 0 indicates the object is not biased towards the cluster and 1 indicates the object is drawn completely to the cluster. The red line separates the 4-object and 8-object structures. Error bars indicate SEM. Object-to-cluster bias was generally low, suggesting subjects did not solely encode chunks (thus forgetting relative object position within a cluster) and contrary to the predictions of a hierarchical generative model, the bias of objects towards their clusters decreased as clusters contained more objects. Nevertheless, in all conditions objects were drawn towards their clusters to some degree.

Did subjects encode objects in addition to their hierarchical structure? Encoding objects as components of hard chunks or a hierarchical generative model should result in distinct patterns of object-to-cluster bias. If subjects encoded objects as hard chunks, they should have retained minimal information about the objects' locations and recalled the objects with a large bias towards their respective cluster centers. If subjects encoded objects in a hierarchical generative model, then they should have recalled objects with more bias towards their cluster centers when clusters contained more objects. Intuitively, subjects can more precisely estimate the centers of clusters that contain more objects and consequently should rely on those clusters more when they are uncertain about the locations of the individual objects.

The bias of objects towards clusters was consistently low (β_o: *M=.19, SEM=.02, max=.62*), suggesting that subjects remembered the locations of individual objects within

their clustering structure, rather than storing chunks and discarding their internal components. Additionally, contrary to the pattern of bias we expected to find if subjects encoded objects in a hierarchical generative model, as objects were arranged in fewer clusters containing more objects, the objects tended to be recalled with less bias towards their clusters (Figure 2.4; *$t(47)=7.14$, $p<.001$ for the linear effect of number of clusters on β_o in a model including fixed effects of number of objects and number of clusters*).

Post-hoc Tukey's HSD pairwise comparison tests confirmed that objects' bias towards their clusters varied with the number of clusters for the 4-object conditions (*smallest difference: .099, 95% confidence interval=.060—.14, $p<.001$*). With the exception of the 2C4 and 1C8 conditions (*difference: .11, 95% confidence interval=-.017—.24, $p=.098$*), the bias of objects towards their clusters also varied for the 8-object conditions (*smallest difference: .16, 95% confidence interval=.031—.29, $p=.01$*). However, even though the bias of objects towards their clusters was generally low, objects were consistently recalled with *some* bias. Together, this pattern of bias suggests that subjects encoded objects in a hierarchical generative model, but did not rely primarily on this form of representation.

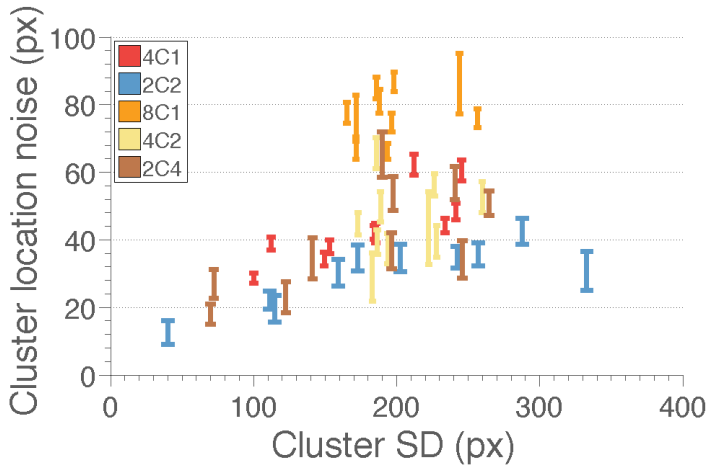

Figure 2.5 The noise of recalled cluster locations (σ_c) given the dispersion of clusters. Each point represents an environment estimated across subjects. Points are color-coded by clustering structure. Error bars indicate SD of the posterior distribution. As clusters were further apart, cluster locations were recalled less accurately.

Did subjects encode objects in a relative position tree? Subjects may have encoded objects in a relative position tree wherein object positions are coded as relative offsets from the cluster centers, and cluster centers are coded as relative offsets from the global center. At first glance this is no different from encoding the objects according to their absolute position. However, if relative positions are recalled with Weber noise (Tudusciuc & Nieder, 2010; Sims, et al., 2012), then larger relative distances will be more difficult to recall. Because the relative distances between objects decreases with more clustering, this could explain why subjects remembered more densely clustered objects more accurately.

Under such a relative encoding scheme, environments that happened to contain more dispersed clusters[6] require larger relative distances to represent positions.

[6] In our study, we held the standard deviation of objects within clusters constant, preventing us from analyzing the effect of relative distance on the accuracy of objects. We predict that this relationship between relative distance and accuracy should remain true for objects within the same cluster.

Consequently, as the dispersion of clusters in the environment increases, subjects should recall clusters less precisely (that is, σ_c should increase). The dispersion of clusters in an environment was significantly correlated with the precision with which subjects recalled cluster centers ($r=0.38$ $p<.01$) (Figure 2.5), consistent with subjects encoding objects according to their relative positions and having difficulty recalling larger relative distances.

Comparing Chunking, Hierarchical Generative, and Relative Position models

To directly test explicit formulations of different encoding theories, we designed three cognitive models that would encode a display and generate responses according to its biases: a hard chunking model that only remembers clusters, a hierarchical generative model that encodes absolute positions (similar to Orhan & Jacobs, 2013) and a model that encodes objects in a relative position tree and recalls relative positions with Weber noise. Each model uses a non-parametric Dirichlet process to determine the clustering of the objects (Ferguson, 1983). We evaluated how well these models could predict subject performance (measured in RMSE) in each environment.

Non-parametric Dirichlet process

We used a non-parametric Dirichlet process to determine the clustering structure of the objects (Ferguson, 1983). Although we used specific clustering structures to generate the locations of objects, the actual distribution of objects in a particular display may have been consistent with a clustering structure we did not design. Such impromptu clustering is especially likely in environments "without" built-in clustering (4C1 and

8C1). Non-parametric Dirichlet clustering assumes that each object's location is drawn from an isotropic Gaussian cluster with some position and standard deviation. Crucially, this clustering model estimates the number of clusters, the assignment of objects to clusters, and the breadth and locations of clusters that best explain the locations of the objects.

We used a Gibbs sampler (Geman & Geman, 1984) to estimate the clustering structure of objects and a concentration parameter. The concentration parameter captures a prior on the number of clusters and its average median value was .11 (*SD=.033*). The chunking, hierarchical generative and relative position models all use the maximum likelihood (MLE) clustering structures of the environments estimated by the non-parametric Dirichlet process.

Chunking model

The hard chunking model uses solely information about the clusters and which objects belong to which clusters to recall the locations of objects. Importantly, the chunking model knows nothing about the locations of the individual objects. Instead, the model recalls the location of an object by randomly sampling from the object's cluster based on the center and standard deviation of the cluster estimated by the Dirichlet process. The model has no free parameters.

Hierarchical generative model

The hierarchical generative model uses knowledge of clusters' locations to compensate for uncertainty in the individual objects' locations. This model is similar to the Dirichlet process mixture model used by Orhan & Jacobs (2013).

The hierarchical generative model noisily encodes the absolute locations of all the objects, as well as the properties of their clusters. Since the model pools memories of individual objects to determine the mean and dispersion of their respective clusters, each additional object in a cluster allows the model to estimate the position of that cluster more precisely. This model uses the same process to estimate the precision of the global center from the locations of the clusters. During recall, the model first recalls the locations of the clusters by averaging the positions of the clusters and global center, weighted by their precisions. The model then recalls the locations of individual objects by averaging the positions of the objects and their clusters, weighted by the precision of the encoded object locations and the posterior predictive spread of objects within a cluster, respectively.

This model has one free parameter: the noise with which objects are encoded. We set the noise parameter to the average object location noise (σ_o) estimated by our error model separately for the 4-object and 8-object conditions.

Relative position model

The relative position model remembers the relative positions of objects and clusters with Weber noise and uses clustering to reduce the magnitude of relative positions. Using the clustering structure inferred by the Dirichlet process, the relative position model remembers the positions of objects relative to their clusters and the

clusters relative to the global center. The model encodes relative positions via their distance and angle and recalls them with circular Gaussian noise on angle and proportional (Weber) noise on distance. The angular and distance noise are captured by two free parameters. We fit the model separately for the 4-object and 8-object conditions.

Can the models predict the difficulty of environments?

We tested whether the models could predict the difficulty, measured in RMSE, of each of the environments across and within clustering structures (Table 2.1). All models were able to predict the difficulty of the environments across clustering structures. However, the chunking model was the worst predictor of subjects' performance (*r=.55, 95% confidence interval=.37-.70*). The relative position model fit environments across clustering structures slightly better than the hierarchical generative model (*Hierarchical generative: r=0.70, 95% confidence interval=.56-.80; Relative position: r=0.89, 95% confidence interval=.82-.93*). Within clustering structures, the hierarchical generative model and relative position models generally predicted the difficulty of environments accurately. Notably, however, the hierarchical generative model matched subjects' behavior particularly poorly for 1C4 and 1C8 environments. This is most likely because when all the objects are in a single cluster, the hierarchical generative model tends to recall objects excessively biased towards the cluster centers. Instead, as our analysis of the bias of objects towards their clusters demonstrated, subjects retained a lot more information about the individual objects in these one-cluster environments. This pattern and the relative position model's better ability to predict behavior suggest that relative position encoding dominated subjects' errors.

Table 2.1 *r* values of the correlation between subject RMSE and model RMSEs for the environments within each clustering structure (4C1-1C8) and for all environments across clustering structures (All). Ch-Chunking model, HG-Hierarchical generative model, RP-Relative position model. *: p<.05, **: p<.01. The relative position model predicted the difficulty of environments within each clustering structure most accurately.

	4C1	**2C2**	**1C4**	**8C1**	**4C2**	**2C4**	**1C8**	**All**
Ch	.0095	.37	-.24	.38	.44	.70*	-.21	.55**
HG	.70*	.63	.16	.43	.54	.58	-.43	.70**
RP	.73*	.85**	.80*	.63	.67*	.61	.53*	.89**

General Discussion

People can encode more information about multiple objects if they exploit the objects' shared statistical structure, rather than encoding them independently. We considered several ways people might use this structure when encoding objects and found that in addition to using a hierarchical generative model to *infer* object properties, people also use the hierarchy to *encode* object properties as relative offsets from the central tendency of their group. Since relative positions seem to be recalled with Weber noise, hierarchical clustering reduces the number of large distances that subjects encoded and thus increases overall accuracy.

Implications for the structure of visual working memory

We found that people encoded objects in a relative position tree (Gershman, Jäkel & Tenenbaum, 2013; Mutluturk & Boduroglu, 2014), using clustering to reduce the Weber noise of relative distances. Even though the relative position model provided the

best quantitative account of our data, the qualitative pattern of results is not entirely consistent with the "pure" chunking, hierarchical generative model or relative position accounts. In contrast to the predictions of a chunking account, people retained more than just information about the hierarchical structure; they also remembered rich information about the individual object locations. Despite subjects recalling positions biased toward cluster centers in all conditions–consistent with subjects encoding positions via a hierarchical generative model (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013)–this bias decreased as clustering density increased, contrary to the predictions of such hierarchical encoding. Furthermore, although a relative position account could explain errors scaling with increasing relative distances, in isolation it does not predict the systematic biases toward cluster centers. Thus, our results suggest that human memory relies on some amalgamation of these structured representations. Indeed, encoding the relative positions of objects requires first determining the hierarchical clustering structure of the scene; and insofar as this is done under uncertainty, biases should be expected from such inference. Altogether, it seems that both hierarchical inference and relative encoding must play a role in human memory encoding.

The extent to which relative encoding or hierarchical inference dominates the pattern of memory errors is likely to vary across circumstances, either due to strategy switching or even from a constant strategy that incorporates both mechanisms. Insofar as clustering structure or individual object properties may be apprehended more easily with brief presentations or other task constraints, different experimental protocols may yield errors that reflect the clustering structure, or the relative encoding. Similarly, stimuli designed with large variations in relative feature offsets will yield more error variability

captured by Weber properties of distance encoding while more homogeneous displays will not show such patterns. In short, while human behavior in our task was best described by the relative position model, we suspect that this result may vary with task parameters, and that uncovering this task-dependent variation in error structure may reveal more fine-grained details of visual working memory mechanisms.

Implications for visual working memory capacity

Our findings that subjects remembered the locations of many objects accurately, even in environments containing eight objects, is at odds with models predicated on a fixed number of slots in visual working memory (Zhang & Luck, 2008; Anderson, Vogel & Awh 2011). Additionally, neither such slot models nor flexible resource models (Bays & Husain, 2008; for review, see Ma, Husain & Bays, 2014) capture the effect of scene structure on memory fidelity. Instead, our results are consistent with recent work suggesting that visual working memory performance is constrained by both memory capacity and the encoded statistical structure of objects (Brady, et al., 2009; Sims, et al., 2012; Orhan, et al., 2014). By decreasing the relative distances between objects, clustering may have allowed a more efficient encoding of the objects, ostensibly increasing observers' capacity.

Limitations

Although we defined chunking as subjects retaining memories of clusters but not individual objects, there are other ways subjects could have encoded objects' structure while discarding information about the individual objects. Subjects may have encoded

sets of locations as familiar shapes such as squares, triangles, etc. (Yantis, 1992). They could have then used these remembered shapes, rather than the cluster centers, to constrain the locations of objects. Under this account, no information about individual objects would be preserved over and above the "chunk", but our analysis would still yield reliable information about the relative (within cluster) positions of individual objects.

Another ambiguity of our analysis arises from the assumption that subjects computed the centers of clusters and encoded individual objects relative to those centers (and reported objects with bias towards those center). An alternative possibility is that subjects encoded the positions of objects relative to *each other* with greater bias exerted by nearby objects (e.g., like gravity, with force dropping off with distance). Unfortunately, our results cannot distinguish whether objects were biased toward each other or toward inferred cluster centers.

Although our report focuses on people's memories of object locations, our model analyses revealed that subjects sometimes recalled locations correctly but matched the wrong objects to the locations (Appendix 2.D). Neither the relative position model nor the hierarchical generative model can account for this behavior. It is likely that subjects' real world priors caused them to expect the locations and identities of objects to be related; subjects may have consequently sought to connect the two. Because locations and identities were independent, the conflict between subjects' priors and the lack of structure in the stimuli may have even impaired performance (Orhan, et al., 2014). If the structure of locations and identities had been correlated—such as if all the objects in the same cluster were the same color or same type of animal—subjects may have used the structure of one to inform the other. Given that being able to perceptually group objects

based on proximity appears to improve the ensemble encoding of other features (Im & Chong, 2014), it is possible that objects in the same spatial cluster would have even been recalled with more similar features/identities. Future studies may examine how the hierarchical encoding of objects affects binding.

Other factors may have improved subjects' apparent memory capacity in our study. Unlike many prior studies, we used distinct objects that never repeated, which may have reduced interference between objects (Endress & Potter, 2014). Furthermore, many subjects reported using verbal strategies (e.g., "the pants are above the shoes") to help remember displays. We suspect that such strategies would have been only minimally helpful, both because they seem to play a minimal role in long-term memory using comparable encoding times (e.g., Brady, et al., 2013)[7], and because they seem insufficient to attain the precision exhibited by visual spatial memory. Since verbally encoded spatial relations (such as "above" or "left") offer only imprecise location information, we suspect that the main benefit of such verbal encoding was to reduce misassociations between objects (Lew, Pashler, & Vul, 2015), rather than encoding the locations themselves. Additionally, patterns of oculomotor movements and attentional shifts could have influenced performance by interfering with encoding in visual memory (Lawrence, Myerson, & Abrams, 2004). Although the uniform distribution of cluster centers in our study still mandates many changes of fixation, it is possible that clustering yields fewer eye movements and attentional shifts between objects in the same cluster,

[7] Although Brady, et al. (2013) assessed the influence of verbal strategies in long-term visual memory, they also found that both short-term and long-term visual memory rely on similar representations; thus it seems reasonable to apply their findings to short-term memories in our experiments. Moreover, the greater precision in short-term memory would seem to make verbal encoding even less effective here than in long-term memory.

improving the fidelity of memories. Our presentation times were also longer than most visual working memory studies, which may have given subjects more time to encode objects. Given that performance appears to asymptote with display times shorter than those used in the current study (Bays, et al., 2011), our results may reflect how people encode stimuli when given enough time to thoroughly observe all objects. Varying the encoding time, delay time or the environment statistics might reveal how people navigate the space of possible encoding schemes.

Finally, a relative position encoding scheme may have been particularly well suited for exploiting the structure of spatial positions. Computing relative positions is straightforward for spatial locations and most likely other features with Euclidean spaces such as size or aspect ratio. However, it is less clear how relative encoding would work in more complex, higher-dimensional spaces such as color or texture. For well defined but non-Euclidean features like hue or orientation, encoding relative positions will likely be helpful if the stimuli are constrained to a narrow range of the space (such that the space is effectively locally Euclidian), but it's not obvious what relative encoding would mean, or predict, if the features span the full range of a circular feature dimension. It is possible that for more complex object properties (such as face identity) people collapse those stimuli onto a small set of salient or trained dimensions (such as organizing faces according to race or gender; Hopper, et al., 2014). If so, relative memory encoding for such complex objects would be possible in this low-dimensional representation; however, finding evidence of such an encoding strategy would require solving a considerably harder problem: specifying the dimensions along which such stimuli are encoded.

Summary

We examined how people encode and use the hierarchical structure of objects under different object loads and structures. In addition to recalling objects biased towards their ensembles, people encoded objects in a relative position tree, using clustering to reduce the Weber noise of relative positions. Our findings are consistent with previous work suggesting that people select encoding schemes that allow them to efficiently represent a given set of stimuli with high fidelity and demonstrate a novel form of encoding.

Chapter 2, in full, is a reprint of the material as it appears in the Journal of Vision 2015. Lew, Timothy F., and Edward Vul. The dissertation author was the principal researcher and author of this material.

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, USA , 106* (18), 7345-7350.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience , 31* (3), 1128-1138.

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

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science , 321*, 851-854.

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision , 11* (10), 1-15.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review , 94* (2).

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 , 24* (1), 87-114.

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.

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

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science , 24* (6), 981-990.

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.

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General , 143* (2), 548-565.

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.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research , 126*, 232-241.

Hopper, W. J., Finklea, K. M., Winkielman, P., & Huber, D. E. Measuring sexual dimorphism with a race—gender face space. *Journal of Experimental Psychology: Human Perception and Performance , 40* (5), 1779–1788.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review , 98* (3), 352.

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

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics , 14* (2), 201-211.

Lawrence, B., Myerson, J., & Abrams, R. (2004). Interference with spatial working memory: An eye movement is more than a shift of attention. *Psychonomic Bulletin & Review , 11* (3), 488-494.

Lew, T. F., Pashler, P. E., & Vul, E. (2015). Fragile associations coexist with robust memories for precise details in long-term memory. *Journal of Experimental Psychology: Learning, Memory & Cognition , 42* (3), 379-393.

Lew, T., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision , 15* (4).

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience , 17* (3), 347-356.

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.

Mutluturk, A., & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, & Psychophysics , 76* (8), 2276-2285.

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., 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.

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

Tudusciuc, O., & Nieder, A. (2010). Comparison of length judgments and the Müller-Lyer illusion in monkeys and humans. *Experimental Brain Research , 3* (4), 221-231.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology , 24* (3), 295-340.

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

Appendix

Appendix 2.A: Mechanical Turk Replication

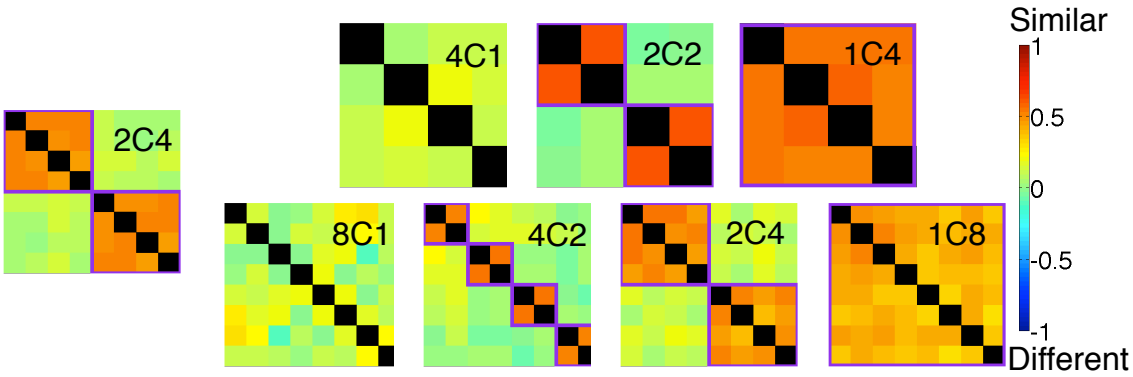

Figure 2.A.1 Error similarity heat maps for the Mechanical Turk replication (single 2C4 heat map on the left) and the main experiment (7 heat maps on the right). The format of the figure is identical to Figure 1.2. Subjects recalled objects in the same cluster with similar errors.

To test whether our results generalized when screen sized was uncontrolled and in an online sample, we replicated our in-lab experiment using Amazon Mechanical Turk for 10 new environments that contained two clusters each composed of four objects (2C4). 59 subjects participated, receiving a monetary bonus based on their performance. The stimuli were identical to our main experiment, except we decreased the size of the environments to 600 x 1100 px due to smaller space in Mechanical Turk's interface.

We again used our error similarity measure (q) to measure whether subjects recalled clustered objects with more similar errors. The error similarity of objects in the same cluster was consistently greater than 0 ($t(58)=23.83$, $p<.001$) (Figure 2.A.1), indicating that memory errors did not accumulate homogeneously for all objects. Instead, subjects' responses respected the clustering structure of the objects.

Appendix 2.B: Did subjects encode objects based on their positions?

Subjects may have remembered objects using salient positions or landmarks. For example, subjects may have used the center or the axes of the environments or visible landmarks like corners and edges (Huttenlocher, 2001; Hollingworth, 2007) to help them recall objects. We expected that if subjects used salient positions or landmarks they would recall objects near such locations more accurately (given Weber noise on relative positions).

To evaluate these strategies, we examined the magnitude of errors in the X-dimension given the X-position and the magnitude of errors in the Y-dimension given the Y-position and binned the positions (Figure 2.B.1). There was no significant effect of position on the magnitude of errors in the X-dimension (*t(438)=1.56, p=.12 for the linear effect of X-position bin in a model including the fixed effect of X-position bin*). However, the Y-dimension of an object's position did affect the magnitude of errors (*t(438)=2.90, p=.003 for the linear effect of Y-position bin in a model including the fixed effect of Y-position bin*) such that errors in the Y-dimension increased towards the bottom of the environment. Given that the environments were symmetrical, this most likely reflects subjects initially dragging objects from below the environment to place them rather than subjects using salient positions or landmarks.

Because objects were arranged in clusters, encoding objects in a relative position tree may have been more effective than landmark-based strategies. Objects were typically very close to their cluster centers, making positions relative to clusters easy to remember. If our stimuli were reliably near salient position or landmarks, we expect subjects would have used those alongside the clustering structure of objects.

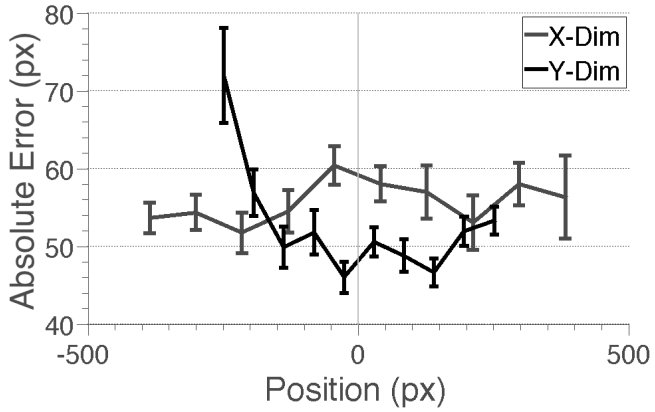

Figure 2.B.1 Absolute error in the X and Y dimensions based on X and Y positions. Lines indicate the binned results. (0,0) indicates the center of the environment, (-500,-350) indicates the bottom-left corner of the environment. Error bars indicate SEM. The locations of objects had little effect on subjects' errors, except when the object was located towards the bottom.

Appendix 2.C: Markov chain Monte Carlo error model fit

We used a Markov chain Monte Carlo algorithm to fit the parameters of our error model. Let $\Theta^{(i)}$ be the set of parameters $\{p_M^{(i)}, \beta_c^{(i)}, \beta_o^{(i)}, \sigma_g^{(i)}, \sigma_c^{(i)}, \sigma_o^{(i)}\}$ at iteration i and $f^{(i)}$. be the mapping of true locations to response locations at iteration i. In each iteration the algorithm samples the values of the parameters that compose Θ conditional on the current mappings of f and then samples the mappings of f conditional on the previously sampled value of Θ. The exact algorithm is:

1. Choose random starting values for the parameters $f^{(0)}$ and $\Theta^{(0)}$.
2. At iteration i, draw a candidate Θ^* from its proposal distribution $P\left(\Theta^* \middle| \Theta^{(i-1)}\right)$
3. Compute an acceptance ratio (probability):

$$a = \frac{LIK\left(s \middle| t, f^{(i-1)}, \Theta^*\right)}{LIK\left(s \middle| t, f^{(i-1)}, \Theta^{(i-1)}\right)}$$

4. Accept Θ^* as $\Theta^{(i)}$ with probability *min*(a,1). If Θ^* is not accepted, then $\Theta^{(i)} = \Theta^{(i-1)}$.
5. Draw a candidate f^* from its proposal distribution $Q\left(f^* \middle| f^{(i-1)}, \Theta^{(i)}\right)$.
6. Compute an acceptance ratio (probability):

$$a = \frac{LIK\left(s \middle| t, f^*, \Theta^{(i)}\right)}{LIK\left(s \middle| t, f^{(i-1)}, \Theta^{(i)}\right)}$$

7. Accept f^* as $f^{(i)}$ with probability *min*(a,1). If f^* is not accepted, then $f^{(i)} = f^{(i-1)}$.
8. Repeat steps 2-7 N times to get N samples of f and Θ.

For the proposal function $P\left(\Theta^* \middle| \Theta^{(i-1)}\right)$, we used truncated normal distributions for each parameter's proposal distribution (the truncation enforced the constraints that the noise parameters must be greater than zero and the bias and misassociation probabilities must be between zero and one). Noise proposal distributions had a standard deviation of 2.5 and bias and probability proposal distributions had a standard deviation of .1.

For the proposal function $Q\left(f^* \middle| f^{(i-1)}, \Theta^{(i)}\right)$, we sampled two unique objects based on the inverse likelihood that they came from their currently assigned locations. Intuitively, this selects the two objects that are currently least likely to be assigned to the correct locations. We then swapped the assignments of the sampled objects to create a new mapping proposal assignment. We set N to 3200 and treated the first 800 samples as burn-in.

Appendix 2.D: Error model parameter estimates

Table 2.D.1**.** Error model parameter fits for each clustering structure. Each cell indicates the mean parameter value and the values in parentheses indicate SEM. Cells containing "NA" indicate cases in which the parameter and clustering condition are not compatible (e.g., because objects are not clustered in 4C1 and 8C1, the model cannot measure objects' bias towards their cluster (β_o)).

	4C1	**2C2**	**1C4**	**8C1**	**4C2**	**2C4**	**1C8**
p_M	.082(.010)	.076(.017)	.071(.008)	.14(.017)	.11(.014)	.096(.028)	.13(.033)
β_c	.17(.014)	.14(.015)	NA	.21(.012)	.19(.015)	.11(.016)	NA
β_o	NA	.12(.018)	.024(.006)	NA	.41(.031)	.25(.040)	.14(.039)
σ_g	29.5(2.0)	22.2(3.3)	35.7(2.5)	33.6(2.1)	31.5(3.0)	27.5(4.0)	45.8(2.0)
σ_c	44.0(3.4)	29.8(2.8)	NA	77.7(2.3)	45.8(3.3)	40.0(4.9)	NA
σ_o	NA	28.9(1.9)	22.9(1.9)	NA	55.5(2.6)	47.8(2.4)	46.0(4.8)

Table 2.D.2. The linear effects of a model including the fixed effects of the number of objects and the number of clusters. DF indicates the degrees of freedom. Under "Number of objects" and "Number of clusters", values in the left and right columns indicate t and p-values, respectively.

	DF	**Number of objects**		**Number of clusters**	
		t	**p**	**t**	**p**
p_M	67	2.0	.042	.71	.48
β_c	47	5.3	<.001	5.2	<.001
β_o	47	9.4	<.001	7.1	<.001
σ_g	67	1.2	.23	2.7	.009
σ_c	47	9.2	<.001	7.7	<.001
σ_o	47	6.2	<.001	2.8	.007

To distinguish different forms of structured representations in visual working memory, our primary analyses focused on the extent to which subjects remembered objects biased towards their clusters and noisily remembered the centers of clusters. In addition, our error model allowed us to examine how the structure of objects influenced other types of errors in visual memories (Table 2.D.1). We used fixed effects models that

include the fixed effects of the number of objects and the number of clusters to examine how different conditions affected the types of errors subjects made (Table 2.D.2).

Subjects may have used the hierarchical structure of objects to help remember associations between objects and their locations. We found that although the rate of misassociations (p_M)[8] increased with the number of objects, it was unaffected by the clustering structure of objects. This suggests that subjects did not use the clustering structure of objects to minimize binding errors.

As objects were arranged in fewer clusters, subjects recalled the locations of clusters with less bias towards the global center (β_c). The decreasing bias of clusters towards the global center may suggest that subjects relied on a representation of objects' hierarchical generative model when remembering the locations of clusters, relying less on the location of the global center as the number of clusters decreased. However, it is unclear why this pattern did not extend to objects' bias towards their clusters.

Subjects also recalled the locations of clusters (σ_c) and objects (σ_o) more accurately. The decreasing noise of cluster and object memories is consistent with the relative position model—organizing objects into fewer clusters should decrease the magnitude of the relative positions needed to represent the objects' and clusters' locations. The clustering structure of objects had an unclear effect on the noise of the global center (σ_g), i.e., the error that is shared among all objects in a display. Subjects appeared to remember the global center more accurately as the number of clusters decreased but this benefit went away when objects were arranged in a single cluster. The

[8] The proportion of locations mismatched by object-to-location mapping function *f* gives similar misassociation rates.

sudden increase in the noise of the global center may reflect subjects focusing on encoding the locations of the individual objects at the cost of the global center when they do not need to remember the clustering structure of objects. Consequently, it is difficult to determine exactly how the objects' clustering structure influenced memories of the global center.

Appendix E: Did the non-parametric clustering process predict subjects' errors?

Our cognitive models used a non-parametric Dirichlet process to infer the clustering structure of objects. To determine whether subjects grouped objects like our cognitive models, we examined how well the groupings inferred by the Dirichlet process predicted the error similarities (q) of objects compared to the actual clustering structures used to generate the locations of the objects. For each condition, we found the average error similarity of objects in the same cluster (Figure A4). If no objects were in the same cluster, we calculated the average error similarity over all objects.

The groupings inferred by the Dirichlet process were either comparable to or better than the actual groupings at predicting the similarities of subjects' errors. The Dirichlet process was notably better than the actual clustering structures in unstructured conditions 4C1 ($t(34)=8.20, p<.001$) and 8C1 ($t(34)=14.20, p<.001$). This demonstrates that the Dirichlet process grouped objects like subjects did even when there was no intended clustering structure. In the other conditions, the error similarity of objects that were actually from the same cluster vs. that the Dirichlet process inferred were from the same cluster were similar, suggesting that both subjects and the Dirichlet process recovered the intended clustering structures.

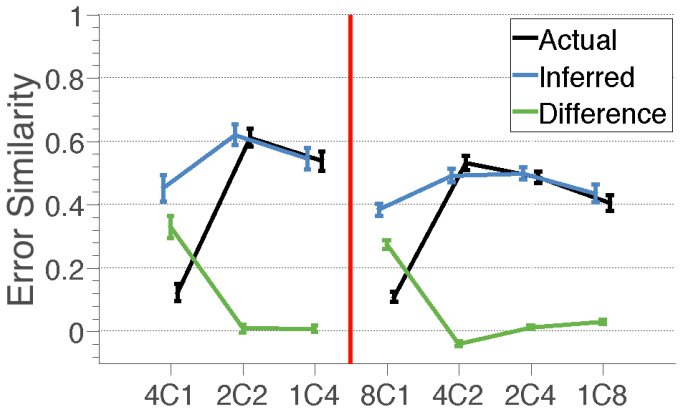

Figure 2.E.1 The mean error similarity (q) of objects in the same cluster. The black line indicates the error similarity of objects that were actually generated from the same cluster. The blue line indicates the error similarity of objects that the Dirichlet process inferred were generated from the same cluster. The green line indicates the difference between the actual and inferred clusters. Error bars indicate SEM. The error similarity of objects was indistinguishable or higher for objects using the inferred groupings compared to the actual groupings used to generate the objects.

Chapter 3 **Hierarchical encoding introduces structured illusions in visual memory**

Timothy Lew and Edward Vul

Abstract

Visual working memory can rely on the structure of objects to improve the fidelity of recall. However, these memory biases may impair observers' ability to identify scenes distorted by their structure. In a first set of experiments, we examined whether representing objects as components of hierarchical generative models biases the recall of objects towards their group centers during recognition. Participants increasingly remembered objects biased towards their clusters with shorter encoding and longer delay times, suggesting that visual working memory uses representations of hierarchical structure to compensate for slowly encoded, fragile memories of objects. In a second set of experiments, we examined whether representing objects as parts of a relative position tree resulted in objects inheriting their parents' rotational errors. Participants consistently remembered individual objects with correlated rotational errors in a location recall task. Although representing objects according to their structure can improve the fidelity of people's memories, doing so makes visual memory susceptible to changes consistent with tat structure.

Introduction

Visual working memory possesses a limited capacity (Cowan, 2001) but can compensate by exploiting objects' hierarchical structure. For instance, people often remember the mean size (Brady & Alvarez, 2011) or location (Lew & Vul, 2015) of groups of objects and recall their features around their average value. However, there are many ways observers might represent the structure of objects. Visual working memory,

for example, might solely encode what groups objects belong to or it might encode further information about how particular objects are arranged within their groups. Furthermore, different encoding schemes may allow observers to better retain information about individual objects and their grouping structures, yielding different patterns of degradation over time. How then do observers represent the grouping structure of objects?

Here we examine how different representations of objects' structure bias forgetting over time. Recent work has suggested that people encode objects as components of a hierarchical generative model (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013). In this scheme, people infer the latent structure that gave rise to the stimuli they observe. For example, if an observer sees several objects near each other, they might infer the objects came from a common cluster. As observers forget the locations of objects, they can compensate for uncertainty about objects' locations by relying on memories of the objects' ensemble statistics (Brady & Alvarez, 2011; Lew & Vul, 2016). However, integrating object and ensemble information in this way also has the potential to introduce biases towards objects' structure during forgetting.

People also appear to exploit the statistical structure of scenes by encoding objects as parts of relative position trees (Gershman, Jaekel & Tenenbaum, 2015; Mutluturk & Boduroglu, 2014; Lew & Vul, 2016). Rather than encode the absolute positions of objects, observers can encode their positions relative to their centroids (Mutluturk & Boduroglu, 2014; Lew & Vul, 2016) or landmarks (Huttenlocher, Hedges & Duncan, 1991) in a tree-like structure. In a relative position tree, clustering can improve memory

by decreasing the magnitude of distances and Weber noise during recall (Lew & Vul, 2016). Encoding objects in a relative position tree should also introduce distinct patterns of correlated errors over time: Because children are defined relative to their parents, as the parents decay in memory their children will inherit their errors.

We examined how visual memory forgets objects when it encodes objects as components of a hierarchical generative model (Experiments 1 and 2) and as parts of a relative position tree (Experiments 3 and 4). In each experiment, participants remembered displays of objects arranged in spatial groups. In Experiments 1 and 2, we tested participants' memory using a recognition task in which objects were either biased towards or away from their centers. In Experiments 3 and 4, objects were arranged in anisotropic clusters and lines and we tested whether participants remembered groups with rotational errors that were passed down to individual objects.

We found that inferring and relying on objects' hierarchical structure introduced biases into both recognition and recall. Participants were more likely to say they had previously seen a display if the lure contained objects shifted towards their cluster centers. Furthermore, this bias increased with shorter encoding times and longer delay times. Participants also remembered clusters and lines with rotational errors that were inherited by their constituent objects. Overall, representing objects as parts of hierarchical representations like hierarchical generative models and relative position trees can improve the overall fidelity of visual memory. However, doing so leaves visual memory vulnerable to errors that are consistent with the scene's statistical structure.

Experiment 1-Structured biases in visual working memory

Visual working memory can encode objects as components of a hierarchical generative model (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013). For example, an observer might infer that a group of objects comes from a common cluster and consequently recall objects shifted towards their cluster centers (Lew & Vul, 2015). Relying on a hierarchical generative model, however, may impair our ability to recognize changes consistent with the underlying model. In Experiment 1, we test whether encoding a hierarchical generative model leaves observers unable to discern when objects shift towards their clusters.

Participants

25 participants from the Amazon Mechanical Turk marketplace performed our task. We compensated participants with a base payment and a performance-based bonus.

Methods

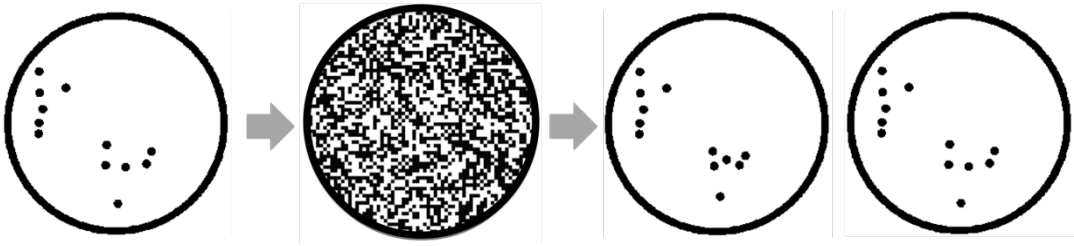

Figure 3.1 Example trial. A) Participants saw 2 clusters of 6 objects for 2 seconds. B) the display was then masked for 4 seconds. C) Finally, participants were shown two displays, (right) the original display and (left) a biased display in which the objects of one cluster were either biased towards or away from the cluster center by a certain magnitude. In this example, the biased cluster is shifted 20 pixels inward. Participants were instructed to select the display that was the same as the original display.

We generated 90 base displays, each composed of 2 clusters of 6 objects (Figure 3.1). The displays and objects had radii of 275 and 10 pixels, respectively, and we selected the location of each cluster from a uniform distribution over the display. We

treated each cluster as a 2-dimensional isotropic Gaussian distribution with a standard deviation of 50 pixels and sampled the object locations with the restriction that objects could not overlap.

For each display, we selected one cluster and shifted objects towards or away from the cluster center by a constant number of pixels. Each shifted display had one of nine shifts (positive and negative shifts indicates shifts towards and away from the cluster in pixels): -20, -15, -10, -5, 0, 5, 10, 15, 20. There were ten unique displays for each shift. Participants studied the base display for 2 seconds. After a 4 second mask, participants saw the base display and shifted display and reported which display was the original. We randomized the order of the displays and whether the target was on the left or the right.

Results

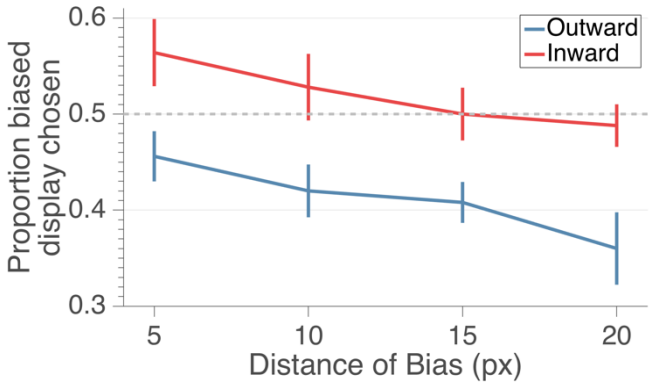

Figure 3.2 Proportion of trials in which the biased display was selected over the original display, given the magnitude of the bias. The red line indicates performance for displays with objects shifted inwards towards their cluster centers and the blue line indicates performance for displays with objects shifted outwards away from their cluster centers. Participants were able to reject the biased display more easily when shifts were larger, but had difficulty distinguishing inward shifted displays from the original displays, even for very large shifts.

Did participants remember objects biased towards their structure? For each shift, we

calculated the proportion of trials in which participants chose the biased display (Figure 2). Controlling for shift magnitude, participants were more likely to choose the inward than outward shifted displays ($t(3)=14.8, p<.001$). Furthermore, although participants reliably identified the strongest shifted-out displays better than chance ($t(9)=3.7, p=.0048$), performance for the strongest shifted-in displays was not significantly different from chance ($t(9)=.54, p=.60$). Participants' difficulty rejecting objects shifted inwards suggests that encoding objects' in a hierarchical generative model impaired visual working memory's ability to recognize when objects were biased towards their clusters.

Experiment 2-Structured biases over time

As memories of individual objects degrade, observers can compensate for their uncertainty by relying more heavily on their memories of objects' overall statistics (Brady & Alvarez, 2011). On this account, the more imprecise the memories for individual objects, the greater the bias towards their hierarchical structure. In Experiment 2, we aimed to test this prediction by manipulating encoding time and delay duration.

Participants

189 participants from the Amazon Mechanical Turk marketplace performed our task for payment.

Methods

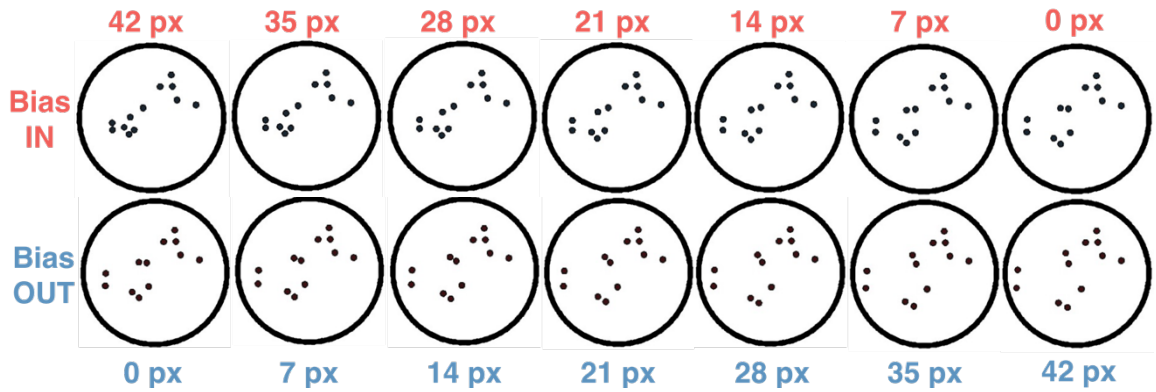

Figure 3.3 Example displays showing the different magnitude and directions of biases. The top-row shows inward biased displays and the bottom-row shows their corresponding outward biased displays. For example, participants would be asked whether the 35 px inward biased display or the 7 px outward biased display was more similar to the original base display. Displays with a bias of 0 px were identical to the base display. In the actual experiment, participants would only see one of these pairs for a given base display.

We generated 70 base displays, each composed of 2 clusters of 6 objects with the same sizes as Experiment 1 (Figure 3.3). For each display, we generated a shifted-in and shifted-out version, keeping the total magnitude of the shifts constant at 42 pixels. For example, if objects in the shifted-in display were shifted 28 pixels inward then objects in the shifted-out display were shifted 14 pixels outward. Each base display had one of seven different inward (and corresponding outward) shifts: 0, 7, 14, 21, 28, 35, 42. We used the same displays for all encoding and delay times.

Participants studied the base display, which was then concealed by a mask. We then presented the shifted-in and shifted-out displays side-by-side and participants reported which was more similar to the original display. We varied either the encoding and delay interval durations across participants. In the encoding conditions, the delay time was 2 seconds and the encoding time was 2, 4 or 8 seconds. In the delay conditions, the encoding time was 2 seconds and the delay was 2, 4 or 8 seconds.

Results

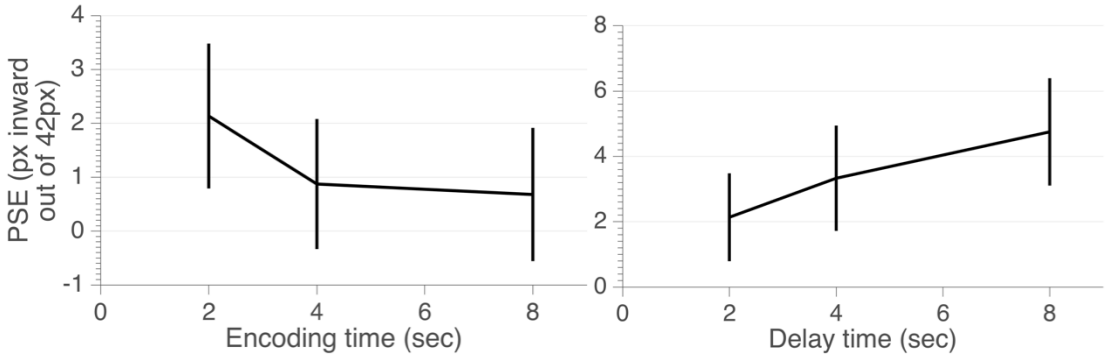

Figure 3.4 The point of subject equality (PSE) at which the inward biased and outward biased displays are equally similar for different (A) encoding times and (B) delay times. A PSE of 2 px, for example, indicates the magnitude of the inward display's bias was 2 px greater than the magnitude of the outward display's bias. With very short encoding times and long delay times, participants were biased towards choosing the inward biased displays.

How did encoding time affect structural biases? For each encoding duration, we found the proportion of shifted-in selections as a function of the difference between the outward and inward shifts (Figure 3.4). To measure participants' bias, we fit sigmoid functions and found the point of subjective equality (PSE) to identify the difference at which the inward and outward displays were equally similar to the base display.

With two seconds to encode objects, participants exhibited a significant bias towards selecting the shifted-in display (Figure 3.4A; *M=2.1, 95% CI=.8—3.5*). With more time to encode objects, this bias decreased (*mixed effect model treating 2 second vs. 4 & 8 second encoding time, shift and their interaction as fixed effects and display as a random effect, main effect of delay time: t(206)=2.37, p=.019*) such that there was no longer a significant bias (*4 seconds: M=.9, 95% CI=-.3—.21; 8 seconds: M=.7, 95% CI=-.6—1.9*). The smaller bias with longer encoding times suggests that participants quickly encoded objects' statistical structure but encoded the locations of individual objects more slowly.

How did delay duration affect structural biases? For each delay time, we calculated the proportion of shifted-in selections and fitted sigmoids to find the PSEs (Figure 3.4B). With increasing delay times, participants remembered objects more biased towards their clusters (*mixed effect model treating delay time, shift and their interaction as fixed effects and display as a random effect, main effect of delay time: $t(206)=5.04$, $p<.001$*). These patterns suggest that participants retained relatively precise memories of ensemble statistics while memories of the individual objects rapidly decayed. To compensate for increasing imprecision of individual objects, visual working memory relied more heavily on objects' latent hierarchical structure model.

Experiment 3-Rotational errors for clusters

Representing objects as parts of groups may have not only allowed observers to use different sources of information during recall but also encouraged observers to encode the positions of objects as vectors with distances and angles relative to their group centers. Lew and Vul (2015) previously demonstrated that clustering can reduce the magnitude of relative distances and thus improve the fidelity of recall by reducing Weber noise. Here we examine how imprecision in the recall of clusters' *orientations* affects the fidelity of memories. We test whether encoding objects as parts of a group can also influence forgetting by adding correlated rotational errors that are passed down from clusters to their components.

Participants

46 participants from the Amazon Mechanical Turk marketplace performed our task. We compensated participants with a base payment and a performance-based bonus.

Methods

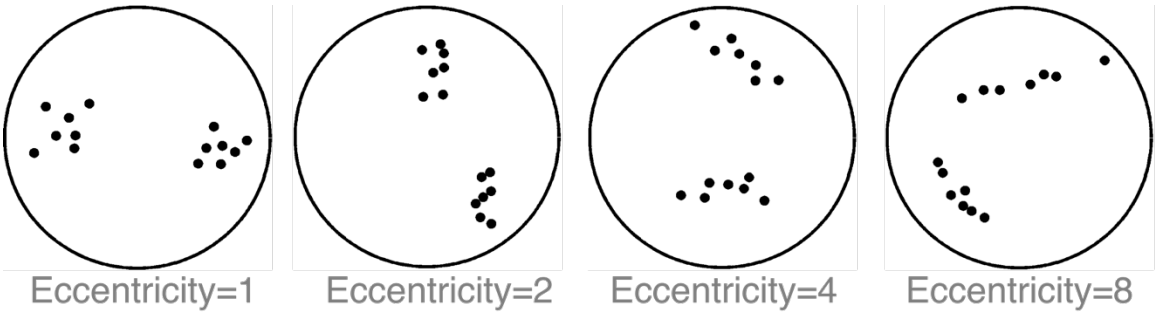

Figure 3.5 Displays containing clusters of each eccentricity level.

We generated 60 displays, each composed of 2 clusters of 7 objects with the same dimensions as the previous experiments (Figure 3.5). In contrast to the previous experiments, we varied the eccentricity of clusters. The base standard deviation along each axis was 32 px. Each display contained clusters where the ratio of the variance of the major axis to the variance of the minor axis was 1 (isotropic clusters), 2, 4 or 8. The location and orientation of each cluster were sampled from uniform distributions. There were 15 displays for each eccentricity and each participant saw the displays from one eccentricity in random order.

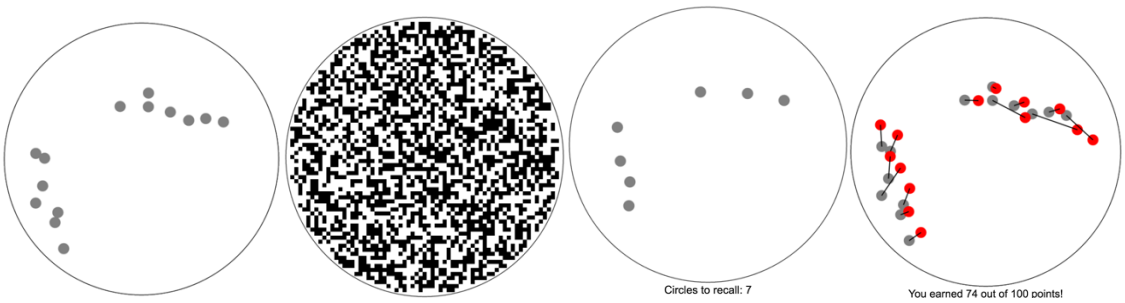

Figure 3.6 Example trial. (A) Participants first studied the target display which was then (B) masked. Once the mask was removed, participants (C) recalled the locations of objects by clicking within the environment. Participants were required to recall all 14 objects. (D) After recalling the objects' locations, participants received feedback in the form of the objects' true locations and a score out of 100.

We instructed participants to study and then recall the locations of objects in each display (Figure 3.6). Participants first studied a display for 6 seconds which was then covered by a mask for 1 second. Participants then saw a blank environment and were instructed to place objects in the locations they remembered by clicking on the environment. They had unlimited time to place the objects and could move them around afterwards, but had to place all 14 objects. Once participants were done recalling the locations of the objects, we gave them feedback by showing them their responses overlaid with objects true locations and giving them a score. We matched objects to responses by using a greedy search that minimized root mean square error (RMSE). Scores shown to participants were 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.

Results

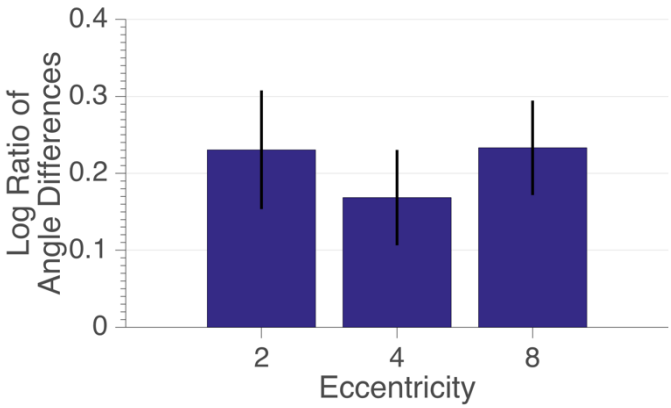

Figure 3.7 The log ratio of the difference between the original clusters' angles and the difference between the recalled clusters' angles for different cluster eccentricities (the eccentricity=1 condition is excluded because there is not a meaningful linear fit to circular clusters). Positive values indicate that clusters' orientations became more similar. Participants remembered clusters with more similar angles.

Were clusters recalled with more similar orientations? Previous work has suggested that visual memory encodes multi-level hierarchical representations of stimuli (Orhan & Jacobs, 2014; Lew & Vul, 2017) and uses information from higher levels to compensate for uncertainty at lower levels. For example, not only are objects arranged in lines recalled biased towards their lines, but the orientations of lines themselves are recalled biased towards their mean orientation (Lew & Vul, 2017). To evaluate whether visual memory also remembers clusters with rotational errors, we tested whether participants recalled the orientations of clusters biased towards their average orientation, such that the difference between the orientations of lines decreased.

To match the original objects to participants' responses, we used the Hungarian algorithm (Kuhn, 1955), minimizing root mean square error. We then used principal components analysis (PCA) to calculate the orientations of the original and the recalled clusters. To measure whether people recalled clusters with smaller differences in orientation, we took the log ratio of the difference between the orientations of the original

clusters and the difference between the orientations of the recalled clusters. Negative values indicate the clusters' orientations became more different, positive values indicate they became more similar and 0 indicates the difference between the orientations remained constant.

The log-ratio of clusters' orientation differences was consistently greater than zero, demonstrating that participants remembered clusters with more similar orientations (Figure 3.7) (*linear model treating eccentricity as a fixed effect, intercept: .23, 95% CI=.095—.37*). Thus, visual memory compensated for uncertainty about clusters' orientations by rotating clusters towards their ensemble statistics, biasing memories of clusters' orientations.

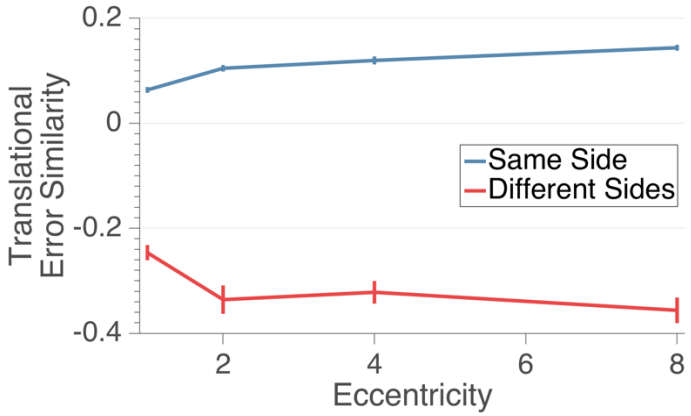

Figure 3.8 The translational error similarity of pairs of objects from the (blue) the same side of clusters and (red) different sides of clusters for clusters of different eccentricities. Participants remembered objects from the same side of clusters with more similar translational errors for clusters of all eccentricities.

Did clusters' rotational errors introduce correlated translational errors? If visual memory encoded the relative positions of within their clusters, then distortions in the orientations of clusters should have been passed down to the individual objects. During

recall, this should have resulted in objects on opposite sides of each cluster exhibiting translational errors in opposite directions.

To measure whether this was the case, we measured the similarity of objects' translational errors. We used the Hungarian matched responses to calculate the translational error $\boldsymbol{x}_i$ for each object i. For each pair of objects in the same cluster, we defined the similarity of their translational errors (**q**) as:

$$q_{ij} = \frac{\boldsymbol{x_i x_j}^T}{\|\boldsymbol{x}_i\|\|\boldsymbol{x}_j\|}$$

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.

To determine whether objects were on the same or opposite sides, we calculated the pairwise distance between each pair of objects and then found the median pairwise distance. If the distance between two objects was less than the median, we considered them to be on the same side. Otherwise we considered them to be on different sides. If objects inherited rotational errors from their clusters, then pairs of objects on the same side of the line should have had similar (positive) translational error correlations and objects on different sides should have had dissimilar (negative) error correlations.

Participants consistently remembered objects on the same side of a cluster with more similar errors than objects on opposite sides of a cluster for all eccentricities (Figure 3.8) (*linear model treating side and eccentricity as fixed effects, main effect of same side:*

.42, 95% CI=.39—.45). Additionally, objects on the same side had positive errors (*t(59)=11.53, p<.001*), indicating they were shifted in the same direction while objects on the opposite side had negative errors (*t(59)=-25.21, p<.001*), indicating they were shifted in opposite directions. While encoding objects as components of relative position trees can help visual memory efficiently represent distances, this form of encoding leaves memory vulnerable to systematic rotational errors.

Experiment 4-Rotational errors for lines

Experiment 3 demonstrated that representing objects as components of relative position trees can allow errors and biases to propagate from higher levels (i.e., clusters) to lower levels (i.e., individual objects), resulting in correlated patterns of errors. However, the amorphous structure of clusters made it difficult to account for objects' exact positions in clusters. Furthermore, objects being on the same or opposite sides was confounded with the distance between objects, making it difficult to tell whether correlated errors were the result of objects being on the same side or being placed right after each other. Consequently, in Experiment 4 we designed a new recall task with objects arranged in lines while also controlling the spacing between objects.

Participants

35 participants from the Amazon Mechanical Turk marketplace performed our task. We compensated participants with a base payment and performance-based bonus.

Methods

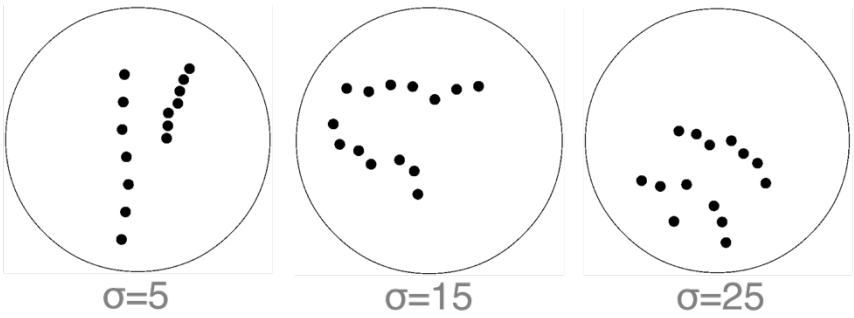

Figure 3.9 Example stimuli of displays for each noise level (σ).

We generated 30 displays, each composed of 2 lines of 7 objects with the same dimensions as the previous experiments (Figure 3.9). The locations and orientations of the lines were sampled from uniform distributions. The lengths of lines were sampled from the empirical distribution of line lengths observed in Lew & Vul (2017). Objects were evenly spaced along the lines and we added noise to each object orthogonal to its line. Displays could have orthogonal noise of 5, 15 or 25 px and was constant within each display. There were 10 displays for each noise level. Participants saw all 30 displays.

Study and recall were identical to Experiment 3.

Results

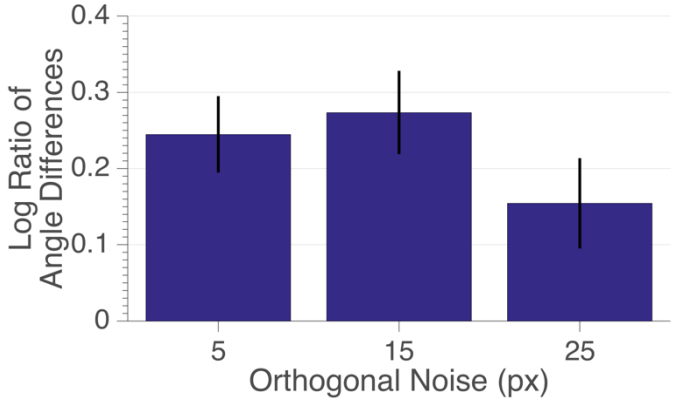

Figure 3.10 The log ratio of the difference between the original lines' angles and the difference between the recalled lines' angles for different levels of orthogonal noise. Positive values indicate that lines' orientations became more similar. Participants remembered lines with more similar angles.

Were lines recalled with more similar orientations? We first sought to confirm whether visual memory recalled lines' orientations biased towards their ensemble statistics as we observed with anisotropic clusters. Using the same procedure as Experiment 3, we matched objects to responses using the Hungarian algorithm, inferred the orientations of lines with PCA and calculated the log ratio of the difference between the recalled lines' orientations and the difference between the original lines' orientations (Figure 3.10). Once again, values greater than zero indicate that people remembered lines with smaller differences in orientation. The log ratio was consistently positive (*linear model treating orthogonal noise as a fixed effect, intercept: .24, 95% CI=.14—.35*), indicating that lines, like clusters, were recalled with more similar orientations.

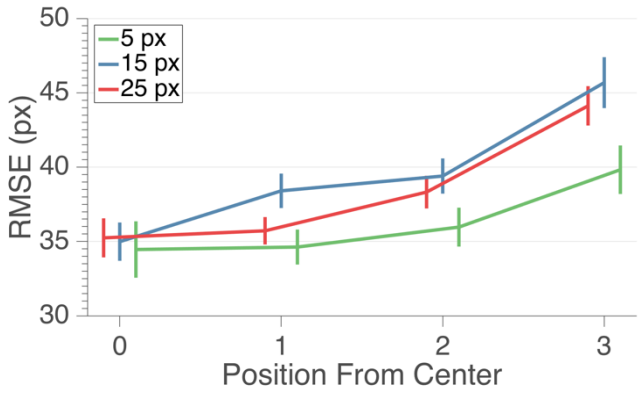

Figure 3.11 The root mean square error (RMSE) of responses given the position of the object in the line. 0 indicates objects at the centers of lines, 3 indicates objects at the ends of lines. Participants remembered objects at the ends of lines less accurately for all noise levels.

Did line position influence accuracy? Recalling the orientations of lines biased towards their ensemble statistics may have influenced how accurately observers recalled individual objects. Encoding objects in a relative position tree and passing along

correlated rotational errors should result in larger errors for objects on the ends of lines compared to objects close to the centers of lines.

To determine whether rotational errors had a greater impact on objects at the ends of lines, we measured the root mean square error of participants' responses for each object given its position in the line. Participants remembered objects at the center of lines most accurately and their accuracy decreased for objects towards the ends (Figure 3.11) (*mixed effects model treating noise and position and their interaction as fixed effects and environment as a random effect, smallest noise—position slope: 1.28, 95% CI=.19—2.36*). Relying on lines' ensemble statistics may have helped participants remember lines' orientations, but at the cost of introducing rotational errors that distorted the locations of objects at the ends of lines.

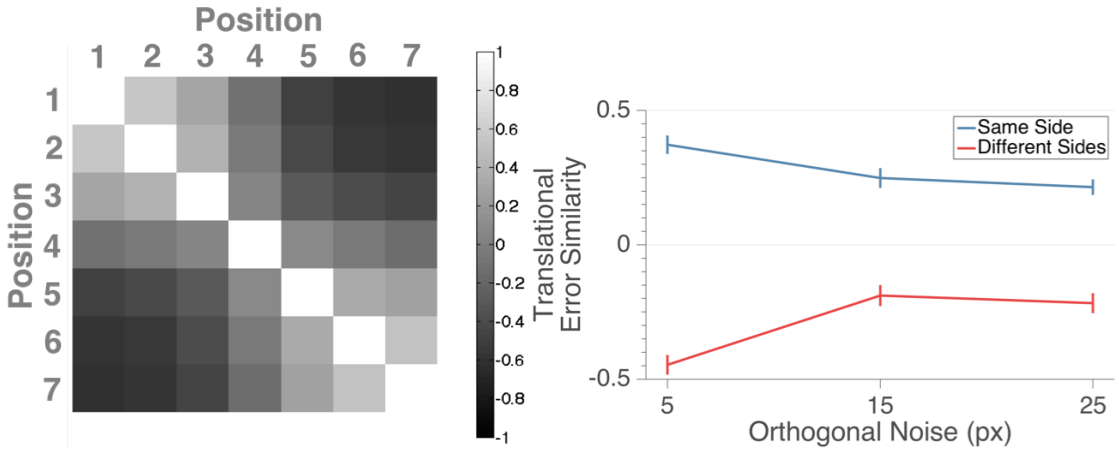

Figure 3.12 A) The translational error similarity of objects given their positions in their lines. 1 and 7 indicate objects at the ends of lines and 4 indicates the middle objects. Objects on the same sides of lines (the upper left and lower right quadrants) were recalled with translational errors in the same direction and objects on opposite sides of lines (upper right and lower left quadrants) were recalled with translational errors in opposite directions. B) The error similarity of objects two positions from the 3rd and 5th objects but either on the same side or different side of the line. Even controlling for distance between objects, objects on the same side of the line were consistently remembered with similar translational errors.

Did lines' rotational errors introduce correlated translational errors? As in Experiment 3, individual objects may have inherited lines' orientation errors, resulting in objects on opposite sides of lines being recalled in opposite directions orthogonal to the line. To evaluate whether objects on the same sides and different sides of lines had positively and negatively correlated translational errors, respectively, we calculated the translational error similarity (**q**) of each pair of objects in the same line.

Objects on the same sides of lines had very similar, positively correlated translational errors ($t(2)=5.42$, $p=.032$) while objects on different sides of lines had very dissimilar, negatively correlated translational errors ($t(8)=-13.70$, $p<.001$) (Figure 3.12A). However, the lack of error similarity between objects on opposite sides of lines may have arisen from time between when participants placed the objects.

To control for the distance between objects, we examined the error similarity of objects that were equidistant but on opposite sides of lines (Figure 3.12B). More precisely, we calculated the error similarity of the 1st and 3rd objects (objects on the same side of the line) and the 3rd and 5th objects (the same distance apart but on opposite sides of the line) and calculated the same measures for the 3rd, 5th and 7th objects. Even when we controlled for the distance between objects, objects on the same sides of lines had similar, positively correlated translational errors (*linear model treating noise orthogonal noise and line side as fixed effects, main effect of line side: .56, 95% CI=.50—.62*). Thus, rotational errors during the recall of lines not only impaired the accuracy of recall for objects at the ends of lines but also created systematic translational biases.

Discussion

Observers remembering objects arranged in clusters and lines exhibited systematic patterns of bias consistent with different forms of hierarchical representation. Encoding objects as components of a hierarchical generative model impaired visual memory's ability to recognize when objects have shifted towards their ensemble statistics. Encoding objects in a relative position tree allowed rotational errors to propagate from groups to individual objects, creating correlated translational errors. Although relying on objects' hierarchical structure can improve the fidelity of visual memory, it can also leave visual memory susceptible to errors consistent with the new structured encoding.

Structured illusions in visual memory

While hierarchical encoding schemes can help the fidelity of memory, our findings suggest they can come at the cost of information about individual objects. When objects are encoded as components of a hierarchical generative model, information is lost about objects' exact positions. When objects are encoded as components of a relative position tree, information is lost about their absolute positions in space. Consequently, visual memory's natural tendency to rely on objects' hierarchical structure may impair people's ability to discriminate fine details.

Overreliance on objects' hierarchical structure may have ramifications for stimuli in which experimenters did not intentionally introduce statistical structure or participants did not infer the structure the experimenter intended. For example, Im & Chong (2014) found that people consider groups of similar, nearby objects as a single unit when

calculating objects' ensemble statistics. Similarly, based on our findings if a subset of objects formed a line then their locations could be recalled with rotational errors, influencing any centroid estimation task. Given the wide array of structures people can infer and be biased by, knowing how people actually do group objects can improve our understanding of the limits of visual memory.

Structured errors as rational behavior

Despite the drawbacks of encoding objects' hierarchical structure, relying on objects' structure can improve the fidelity of memories in aggregate. Encoding the ensemble statistics of objects can help visual working memory efficiently and accurately represent stimuli despite its limited capacity (Sims, et al., 2012; Orhan, et al., 2014). And during recall, visual memory remembers objects biased towards their ensemble statistics, improving the overall precision of memories (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013).

Relying on objects' hierarchical structure may also improve the ecological validity of visual memories. Just as objects in a hierarchical generative model are drawn towards their center, a flock of birds circling a bagel will be drawn towards their center. The correlated translational errors introduced by encoding objects in a relative position tree can help capture how parts of objects rotate together, like a gymnast doing a cartwheel (Gershman, Jaekel & Tenenbaum, 2015). Although these structured biases impaired memories of static stimuli in our experiments, in the real world they may help visual memory compensate for how objects actually move and change over time.

Summary

In a series of memory tasks, we demonstrated that forming hierarchical representations of stimuli can impair people's memories of individual objects, yielding systematic illusions based on the inferred structure of a scene. Encoding objects as components of hierarchical generative models hindered people's ability to discern previously seen stimuli from new stimuli biased towards their ensemble statistics. Remembering objects as parts of relative position trees resulted in the loss of objects' absolute positions and introduced correlated rotational errors across objects. While relying on objects' hierarchical structure can improve the overall fidelity of visual memories, they impair people's ability to recognize errors biased towards that same structure.

Chapter 3, is currently being prepared for submission for publication of the material. Lew, Timothy and Edward Vul. "Hierarchical encoding introduces structured illusions in visual memory." The dissertation author was the principal researcher and author of this material.

References

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 , 24* (1), 87-114.

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.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research , 126*, 232-241.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review , 98* (3), 352.

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

Kuhn, H. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly , 2.1*, 83-97.

Lew, T., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision , 15* (4).

Mutluturk, A., & Boduroglu, A. (2014). Effects of spatial configurations on the resolution of spatial working memory. *Attention, Perception, & Psychophysics , 76* (8), 2276-2285.

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., 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.

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

Wertheimer, M. (1923). Laws of organization in perceptual forms. *A Source Book of Gestalt Psychology*.

Conclusion

My dissertation demonstrates how encoding objects according to their statistical structure can improve the fidelity of visual memory and the limitations of relying on structured memories. Visual working memory relies on complex priors about how objects in the world are grouped to compensate for uncertainty in memory (Chapter 1). These priors facilitate the organization of objects into groups, resulting in sophisticated structured representations in memory. In Chapter 2, we explored how these structured memories can improve the fidelity of recall. Encoding objects in a hierarchical generative model allowed visual memory to compensate for uncertainty about objects' locations by recalling objects biased towards their group centers. Encoding objects in a relative position tree reduced the magnitude of relative distances between objects, decreasing Weber noise. However, reliance on objects' statistical structure also leaves visual memory susceptible to errors and changes consistent with that structure (Chapter 3). Relying on a hierarchical generative model impaired people's ability to recognize new displays where objects were biased towards their clusters. Relying on a relative position tree introduced correlated rotational errors. Although using objects' statistical structure typically may improve the fidelity of memories, the loss of information about individual objects can still hinder people in many situations.

Investigating how people represent and group simple stimuli is a crucial step in understanding how visual working memory represents the world at large. Low-level hierarchies, like "objects, lines and pairs of lines" examined here, can act as the building blocks for representations of real-world hierarchies like "leaves, trees and forests". My work demonstrates that priors play a crucial role in determining how visual memory

infers these structures. Just as Gestalt priors about proximity and continuity yield clusters and lines in memory, statistical knowledge of how leaves and trees change with the seasons may bias people's memories of a forest. Uncovering more of visual memory's prior expectations can reveal what higher-level structures people infer and how those structures influence recall.

Although in Chapters 1 and 2 people exhibited behaviors consistent with classical Gestalt principles, these patterns may have been the result of task-specific demands rather than the innate structure of visual memory. Confronted with a difficult location recall task, for example, participants may have chosen to strategically rely on objects' groups rather than try to precisely remember the individual locations of a large number of objects. Participants' inability to distinguish objects biased towards their clusters during recognition tasks in Chapter 3 provides stronger evidence that basic Gestalt principles like the principle of proximity are more engrained properties in visual memory, and not just a strategic choice. However, the extent to which different patterns of structured encoding and bias are the result of an observer's conscious decision vs. the essential properties of visual memory require further investigation.

People's reliance on memories of objects' structure also lends insights into how visual memory goes from remembering objects as individual elements in a common ensemble to representing them as parts of whole objects. When visual memory remembers objects as components of clusters and lines, it also appears to lose information about objects' absolute positions and instead encode them according to their relative positions. People recalled relative position trees with object-like errors, introducing correlated rotational errors and remembering clusters and lines with similar orientations.

Hierarchical generative models, relative position trees and other types of ensemble representations may act as precursors for full-fledged object representations.

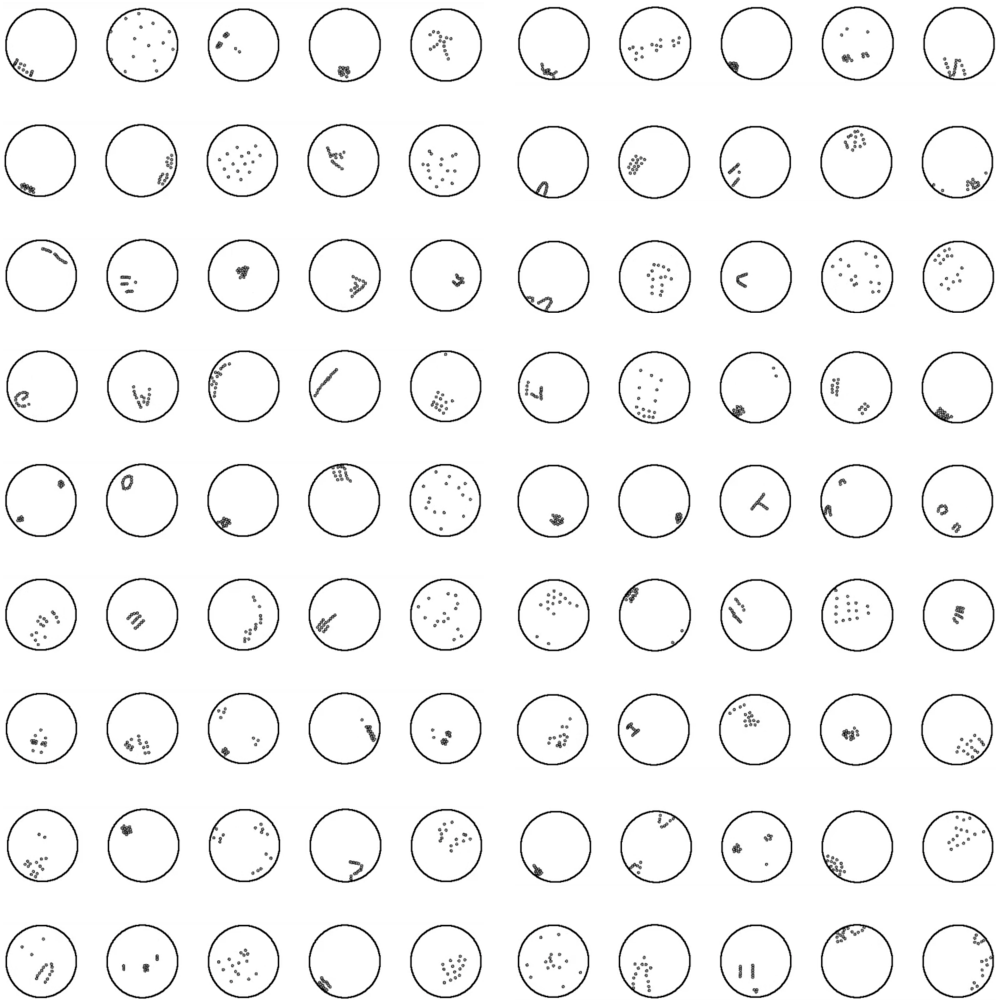

Figure 4.1 Final displays in Chapter 1's iterated learning chains. In addition to clusters, lines and parallel lines, participants converged towards a variety of structures exhibiting complex patterns of symmetry and often resembling more sophisticated objects. Each row is the result of the ten chains from a single seed.

For simplicity, these experiments have focused primarily on how people organize objects into relatively simple structures such as clusters, lines and parallel lines. Nevertheless, subjects often remembered objects in more complex or semantically meaningful arrangements. Examining the final iterations of each chain in Chapter 1

(Figure 4.1) for instance reveals that participants converged towards more complex structures such as contours with bilateral symmetry (row 1, column 5), nested arrows (row 3, column 4) and even a lowercase "e" (row 3, column 1). Similarly, in Chapter 2 participants occasionally reported relying on semantic associations like "The hat was above the suit". The clusters and lines observed here may act as the building blocks for basic geometric patterns and simple shapes, such as sets of parallel lines, perpendicular lines and grid-like patterns. Inferring semantically meaningful structures (i.e, the letter "e" or a person's outfit) may have allowed people to efficiently encode objects and introduced distinct patterns of bias consistent with those structures. For example, people might remember objects arranged in an "e" biased away from straightening the lower curve and turning the "e" into and a categorically different "p". Future studies may examine what factors influence when people infer more sophisticated, semantically-meaningful structures and what kinds of biases those structures introduce into visual memory.

In summary, my dissertation has sought to understand how we infer and utilize objects' hierarchical structure in our everyday experiences. Visual memory relies on objects' grouping as memories of individual objects decay, introducing distinct patterns of bias. Nevertheless, relying on objects' structure often improves the overall fidelity of visual memory and enables visual memory to build sophisticated, hierarchical representations that capture the real world.