The Role of Stimulus Form in Visual Working Memory for Orientation

By

Young Eun Park

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

September 30, 2017

Nashville, Tennessee

Approved:

Frank Tong, Ph.D.

Randolph Blake, Ph.D.

Geoffrey F. Woodman, Ph.D.

Joo-Hyun Song, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In everyday situations, we often need to keep multiple pieces of information briefly in mind for as long as they are needed to carry out a task, such as rehearsing a phone number while dialing it or completing a math problem through multiple steps of calculation. The ability to temporarily store and manipulate information in the service of ongoing cognition is called working memory. Working memory is thought to provide a foundation for many high-level cognitive functions, including language comprehension, learning, planning, reasoning, and general fluid intelligence (Baddeley, 1986; Daneman & Carpenter, 1980; Engle, Kane, & Tuholski, 1999). Working memory is a multifaceted concept that has evolved over time since the beginnings of psychology, as reflected in the various names this system has been called: "primary memory" emphasizing the phenomenology of conscious access to its contents (James, 1890), "short-term memory" emphasizing the transient nature of the storage (George Miller, 1956; Atkinson & Shiffrin, 1968), and finally, "working memory" emphasizing the active mental operations on the stored information (Baddeley & Hitch, 1974).

The most influential view of working memory is that developed by Baddeley and colleagues (e.g., Baddeley, 1986, 1992, 2000, 2012; Baddeley & Hitch 1974), in which working memory consists of separate buffers for storing different forms of information, controlled by executive attentional processes. While this multi-component model of working memory emphasizes the dynamic role that this system plays in complex cognitive activities, another aspect of working memory that researchers have focused heavily on is its limited capacity (e.g., Cowan, 2001; Luck & Vogel, 1997), which has led to prominent advances in our understanding of the fundamental limits in mental storage capacity. While the multi-component model views the overall capacity limit as an emergent property of the coordinated operation of multiple cognitive resources or subsystems, this latter approach seeks to determine the 'true' capacity limit of working memory after removing contaminating variables, such as articulatory rehearsal, chunking, and long-term retrieval. Based on an extensive number of studies, Cowan (2001) estimated that up to 3~4 chunks of information can be simultaneously held in working memory.

Historically, working memory research was dominated by verbal paradigms, perhaps because of the prominent role of verbal ability in everyday cognition. Over the past two decades, however, research on visual working memory has seen tremendous theoretical and methodological advances, shedding new light on the nature of the capacity limits in the storage of visual information.

The capacity limitations in visual working memory (VWM)

One of the critical findings that spurred considerable research on visual working memory (VWM) was a pioneering study by Luck and Vogel (1997). While earlier psychophysical studies examined short-term memory for a single visual feature (e.g., Kinchla & Smyzer, 1967; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Nilsson & Nelson, 1981; Regan, 1985) or a complex visual pattern (e.g., Miyashita & Chang, 1988; Phillips, 1974), the storage capacity for such stimuli could not be precisely quantified. Luck and Vogel devised a paradigm that involved the brief presentation of a sample array containing a variable number of simple features (e.g., colored squares or oriented bars), followed by a blank interval and then a test array. The test array either remained the same or one of the features was altered, and the task was to indicate whether a change had occurred across the two displays or not. This change detection paradigm revealed that observers could accurately detect a change in arrays containing a small number of items, but that performance suffered as the set size increased beyond 3 or 4 items. On the other hand, sample arrays of objects consisting of multiple task-relevant features (e.g., colored oriented bars) led to similar levels of performance as was found for single-feature objects. These results suggested that the capacity of VWM was limited by a fixed number of integrated objects, such that performance was about equally good regardless of the number of individual features that had to be stored for each object. This finding led to an influential theory of VWM, which posits that individual visual objects are held in working memory using a small, fixed number of independent "slots" (Luck & Vogel, 1997).

Another major shift in research on VWM was prompted by the finding that the precision of VWM monotonically declines as a function of number of items to be remembered in the sample array (Wilken & Ma, 2004). The authors developed a delayed estimation paradigm, in which the observer is briefly shown an array of items whose

feature values can continuously vary, and asked to precisely report the remembered feature of a randomly probed item, by adjusting a test stimulus in a continuous feature space. Memory noise, as indicated by the variability (SD) of recall errors (i.e. difference between the correct value and the reported value), was found to increase in a continuous fashion as the set size increased, indicating a decline in memory precision. Based on this finding, Wilken and Ma proposed that VWM capacity is best understood as a continuous resource that can be flexibly shared among numerous items.

The proposal that memory capacity consists of a continuous and flexible resource (Bays & Husain, 2008; Wilken & Ma, 2004) substantially challenged the longstanding notion that items in working memory are either remembered or forgotten in an all-or-none fashion. More recently, Zhang and Luck (2008) introduced a novel procedure to separately estimate the precision of memory and the probability of forgetting in a delayed estimation task. Their mixture modeling approach provides a conceptual framework for dissociating response errors reflecting complete memory failure (i.e., random guesses) from those reflecting the noise in successfully remembered items (i.e., a normal distribution centered on the correct value). Through their work, the authors demonstrated that estimated rates of guessing substantially increased as the set size increased beyond three, implying that VWM has a fixed item-limit. To account for the superior memory precision observed at smaller set sizes, Zhang and Luck proposed that multiple copies of a single item can be stored in independent slots, which can be averaged to reduce the overall noise associated with the representation of that item.

Since then, researchers have shown that the continuous resource model can also be extended to account for these considerations of "apparent" guessing, by introducing stochastic variability of memory precision across items and across trials (Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012). In addition, some portion of the uniform error distribution may be attributable to accidentally reporting information about the wrong item in the display; such confusion errors may not necessarily reflect complete forgetting of the probed item (Bays, Catalao, & Husain, 2009). On the other hand, some of the variability in memory precision across trials has been shown to arise from systematic variation in precision of working memory across the stimulus space tested (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014; Pratte, Park, Rademaker, &

Tong, 2017), rather than from random fluctuations in the distribution of memory resources. Importantly, when these systematic sources of variability were taken into account, VWM performance revealed evidence of a discrete capacity limit in VWM (Pratte et al., 2017).

Thus, the formulation of each of these two competing views of the capacity limits of VWM has evolved in order to accommodate novel observations. This intense debate has led to a richer understanding of how visual information is stored in the short term, as well as continued refinement in the measurement and in the interpretation of recall errors.

Central goals of this thesis

In this dissertation, I will critically examine the assumptions of the discrete-slots view of VWM, which presumes that there is a fixed item-limit in VWM that is supposedly invariant across different types of stimuli. I will demonstrate that the physical appearance of stimuli can greatly impact the information capacity of VWM even within the same task-relevant feature dimension. Specifically, my studies demonstrate that people can hold far more line orientations in working memory than orientations defined by gratings, whereas visual precision is similar across the two stimulus types. These empirical findings cannot be readily explained by the existing models of VWM capacity assuming either discrete slots or a continuous resource. I will propose a new modeling framework that incorporates visual grouping into the discrete-slots model of VWM to account for these dramatic differences in the estimated memory capacity across different types of stimuli. By assuming that multiple items can be efficiently stored as a coherent memory unit, I will demonstrate that the apparent lack of fixed-item limit for certain types of stimuli can be explained within the discrete-slots framework, if one considers that "visual groups", rather than individual items, serve as the storage units of memory.

Orientation as a window to the architecture of VWM

Orientation is strongly represented in the visual system, and there is considerable cognitive neuroscience evidence that VWM for orientation relies on representations in the early visual cortex (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009; Ester, Serences, & Awh, 2009; Ester, Sprague, & Serences, 2015). Orientation stimuli are commonly used in studies of VWM (Luck & Vogel, 1997; Fougnie, Asplund & Marois,

2010; Fougnie, & Alvarez, 2011; Park, Sy, Hong, & Tong, in press; Rademaker, Bloem, De Weerd, & Sack, 2015; Rademaker, Tredway & Tong, 2012; van den Berg et al., 2012; Wilken & Ma, 2004). While VWM for color may benefit from verbal encoding (Donkin, Nosofsky, Gold, & Shiffrin, 2015), VWM for orientation appears to rely primarily on a visual code, as suggested by the engagement of the visual cortex during periods of working memory maintenance. A consideration of how orientation information is processed by the visual system has proven to be crucial for a better characterization of the nature of the capacity limits in VWM (Pratte et al., 2017).

While orientation stimuli have been widely used in studies of VWM capacity, most research has assumed that VWM performance arises from fundamental capacity limits, and therefore, should be invariant to the particular stimulus type or stimulus form to be remembered. Thus, the potential role of different stimulus forms, such as Gabor patches or bars, in the capacity of VWM has been largely neglected. In this dissertation, I will investigate the impacts of stimulus form in VWM performance for orientation, using a paradigm that allows estimation of the number and precision of orientations stored in VWM.

Overview

In this dissertation, I present a series of studies that examine the role of stimulus form in the efficient storage of orientation information in VWM, comparing the VWM performance between two stimulus formats: gratings and lines.

In the first study (Chapter 2), I report a large enhancement in VWM capacity for line orientation over grating orientation, as a starting point for an in-depth series of studies that explore the role of visual grouping in the capacity advantage for line stimuli. In the second study (Chapter 3), I develop a model of VWM capacity that incorporates visual grouping into the existing discrete-slots model in order to account for the stimulus-dependent changes in the information capacity of VWM, as well as the progressive advantage I observe in storage efficiency for line orientation as compared to grating orientation, with increasing set size. In the final study (Chapter 4), I examine whether VWM performance varies systematically across randomly generated line and grating displays. To determine whether this variability in performance across displays arises from

perceptual organization processes, I develop a model that quantifies multiple perceptual grouping cues in each display and predicts VWM performance based on the combined strengths of these grouping mechanisms for each display. This Gestalt-inspired model could successfully predict human performance, accounting for a major proportion of variance in VWM performance across displays.

Together, these studies demonstrate that multiple perceptual grouping mechanisms give rise to superior VWM capacity for line orientation, by combining behavioral experiments and computational modeling. The findings from these studies provide novel insights into the nature of capacity limits in VWM, and highlight the crucial role of stimulus factors and perceptual grouping mechanisms in determining the storage efficiency of VWM.

Chapter 2

Visual working memory capacity for orientation depends on stimulus form

Introduction

In everyday situations, we may need to remember important details of a visual scene when they are no longer in view, such as when comparing the face of a stranger with that from a wanted poster or when following a series of origami instructions across successive fixations. The ability to actively maintain and manipulate visual information for use in the task at hand is called visual working memory (VWM) (see review by Luck & Vogel, 2013; Ma, Husain, & Bays, 2014). While VWM provides a workspace needed for complex visual tasks, its capacity is severely limited, able to hold no more than 3-4 visual objects at a time (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008).

The capacity of VWM is typically measured using simple, unidimensional features, such as color or orientation. One of the advantages to using such stimuli is that the stored information is easy to quantify in terms of the number of discrete features (Luck & Vogel, 1997, 2013; Rouder, Morey, Morey, & Cowan, 2011), whereas the part-based structure of complex objects, such as Chinese characters or random polygons, would be more difficult to quantify (e.g., Alvarez & Cavanagh, 2004). Presumably, a random set of simple features cannot be organized based on any semantic relationship (e.g., Ericsson & Chase, 1982), the capacity measures are thought to be uncontaminated by grouping or rehearsal strategies (Cowan, 2010). Change detection tasks with simple visual features provide reliable measures of working memory capacity, which are highly correlated with broad measures of cognitive aptitude and fluid intelligence (Cowan et al., 2005; Fukuda, Vogel, Mayr, & Awh, 2010; Johnson et al., 2013). Thus, these capacity estimates are thought to provide relatively pure measures of one's core cognitive capacity, general across modalities and domains.

While a heavy emphasis has been placed on the link between VWM capacity and a general, cross-domain resource, a relatively neglected issue is the role of the visual system per se in mediating the capacity measures for specific types of visual stimuli. Although some studies have suggested that VWM capacity is diminished for perceptually complex

objects (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005; but see Awh, Barton, & Vogel, 2007), the ways in which VWM capacity for simple, unidimensional features might be affected by physical appearance of the stimuli have not been fully investigated.

Even within a common feature domain, VWM capacity and precision might depend critically on aspects of stimulus form. An important clue is provided by an earlier study by Alvarez and Cavanagh (2008), who compared change detection performance for orientations defined by Gabor gratings and lines. The authors found that VWM capacity for lines (4.0 items) was nearly twice as large as that for Gabor gratings (2.3 items), which was taken to suggest that the capacity of VWM depends on whether orientation information is defined by an object's bounding contour as in the case of line stimuli, or by its surface texture as in the case of grating stimuli. This finding led to the *boundary hypothesis*, which posits that an object's boundary feature (i.e., a line's tilt or a grating's round shape) is encoded into VWM first, followed by an encoding of the object's surface feature when necessary (i.e., a tilt of the grating's striped texture). Unlike the line stimuli, the grating stimuli additionally require their surface features to be encoded, thereby requiring extra "space" in VWM (Alvarez & Cavanagh, 2004).

Alvarez and Cavanagh's (2008) finding highlights the important role of stimulus factors in influencing VWM capacity estimates. However, given that change detection performance is determined not only by whether the probed item is stored in VWM, but also by whether the resolution of the stored representation is sufficient for reliable detection of a change (Awh et al., 2007; Wilken & Ma, 2004), it is unclear whether the impaired performance for gratings is due to a reduced VWM capacity or poorer resolution of stored orientation compared to that for lines. Another issue that has not been fully investigated is the possibility that VWM for lines could have benefited from enhanced perceptual grouping of line orientations. To account for the dramatic difference in VWM performance between gratings and lines, Alvarez and Cavanagh emphasized the inefficiency inherent to extracting a surface feature of an object, compared to a boundary feature of an object. While an isolated line orientation itself might be stored more efficiently than a grating orientation, a further VWM advantage could arise from visual grouping of multiple lines into coherent chunks and configurations.

In the present study, I wanted to determine whether VWM for gratings and lines differ in terms of the number or the resolution of stored orientations. I used a delayed estimation paradigm that required a precise report of the remembered feature value (Wilken & Ma, 2004). Applying a mixture model (Zhang & Luck, 2008) allowed me to obtain separate estimates of VWM precision and capacity for these stimuli. In addition, I sought to further examine the role of visual grouping as a potential mechanism for the efficient storage of line orientations.

General Methods

Participants

The participant pool consisted of healthy volunteers with normal or corrected-to-normal visual acuity, recruited from the Vanderbilt University community. Participants (except the author) received course credit or monetary compensation ($12 per hour) for participation. Written informed consent was obtained prior to participation. All aspects of this study were conducted according to procedures approved by the Institutional Review Board of Vanderbilt University.

Apparatus and stimuli

In all of the experiments, visual stimuli were generated using MATLAB 8.3.0 (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) running on a Mac Pro. Stimuli were displayed on a gamma-corrected CRT monitor (1280 x 1024 resolution; a 85Hz refresh rate; 27.2 cd/m^2 mean luminance). Experiments were conducted in a darkened room, and head position was stabilized using a chin rest with a forehead bar to maintain a viewing distance of 46 cm. Observers were asked to fixate centrally on a black bull's eye (0.5˚ visual angle in diameter) for the duration of each trial, throughout the experiment.

The grating stimulus consisted of a sine-wave grating (2˚ in diameter, 2 cycles/˚, 50% contrast) presented within a circular Gaussian envelope (σ = 2˚, truncated at 2˚ diameter). In Experiment 1A, the line stimulus was a white rectangle (0.25˚ x 2˚), smoothed with a Gaussian filter (σ = 0.15˚), with the same peak luminance as the grating stimulus (40.8 cd/m^2). In all subsequent experiments, the line stimulus was generated by

applying a rectangular aperture (0.25˚ x 2˚) to the central bar of the grating, which corresponded to half a cycle of the sine wave. All stimuli were presented on a gray background (mean luminance 27.2 cd/m^2).

A sample array typically consisted of six gratings or lines presented at randomized locations surrounding the central fixation point (4˚ eccentricity) with a minimum separation distance of 0.62˚ edge to edge. This circular arrangement was used in all experiments, except for Experiment 1B, in which stimuli were presented at random locations on a virtual 4 x 4 grid with spatial jittering (see Experiment 1B for details). Stimulus orientation was randomly chosen on each trial for each item, from 180 evenly spaced orientations (0-180˚).

For the delayed estimation task (Experiments 1–3), I used a PowerMate USB knob (Griffin Technology, USA) to allow the observers to continuously adjust the orientation of a test stimulus. In Experiment 4, I evaluated working memory performance using a change detection task.

Experiment 1A

The goal of this experiment was to determine whether stimulus form (gratings or lines) impacts the capacity or precision of VWM for orientation. I adopted a mixture-modeling approach to characterize the response error distribution as a mixture of two components: A von Mises distribution (the circular analogue of the Gaussian distribution) centered on the true feature value for trials in which the probed item was successfully held in memory, and a uniform distribution for trials in which a random guess was made (Zhang & Luck, 2008). The delayed estimation responses will likely be more accurate for lines than gratings, but it was of interest whether this advantage would be characterized by a more precise representation of orientations in memory or a reduced occurrence of random guess responses.

Two components of the mixture model are depicted in **Figure 1A**. Considering this mixture distribution as representing the VWM performance for grating stimuli, two possible forms of VWM enhancement relative to this distribution are presented in **Figure 1B**. If VWM has a greater capacity to store line orientations (**Figure 1B**, left), the height of the uniform component ($P_{failure}$, shown in red) should be reduced for lines compared to

gratings. By contrast, if VWM can store the same number of lines as gratings but at a much higher resolution (**Figure 1B**, right), the spread of the von Mises distribution (SD, shown in blue) should be narrower for lines while the uniform component should be equivalent for the two stimulus types.

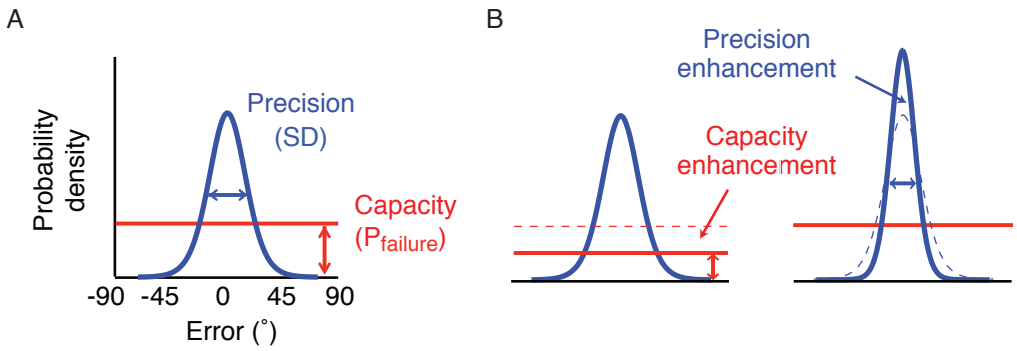

Figure 1. Zhang and Luck's (2008) mixture model analysis of delayed estimation data. (A) The distribution of errors is modeled as a mixture of two probability distributions: A von Mises distribution (blue) and a uniform distribution (red). The shapes of these two components are determined by the SD and $P_{failure}$ parameters, respectively. (B) Two possible forms of VWM performance enhancement, relative to the error distribution depicted in (A).

Method

Participants

12 participants (6 male; ages 19–30 years), including the author, took part in this study.

Stimuli and procedure

Each sample array consisted of 2 or 6 items of either stimulus type (gratings or lines) positioned at randomized locations, 4˚ from central fixation, with different stimulus types presented in separate blocks of trials.

A typical trial sequence for a delayed estimation task is depicted in **Figure 2**. Each trial began with a central fixation presented for 1 s, followed by a sample array containing 2 or 6 items for 200 ms. After a 1-s delay, a spatial cue (a black outline circle; 2˚ in diameter) appeared at one of the stimulus locations for 500 ms, followed by a test stimulus (a grating or a line) at fixation. Participants were asked to report the orientation of the cued

item from memory as precisely as they could, by rotating the test stimulus with a rotary knob. The initial orientation of the test stimulus was randomly determined on each trial, independent of the orientation of the original target stimulus. Participants were asked to press a space bar when they were satisfied with their response, after which they received visual and auditory feedback. Visual feedback consisted of a test stimulus showing the true orientation, along with an accuracy score (0-100), calculated as 100% minus the percentage of the absolute error magnitude relative to the maximum possible error (90°). Depending on the accuracy score, a pleasant rising tone (≥ 99%), a cash register sound (≥ 50%), or a low-pitched beep (< 50%) was played. For accuracy scores above 99%, participants received 100 bonus points. Participants received feedback about their average accuracy and the number of bonus points earned at the end of each experimental block. In general, participants appeared to find this feedback to be encouraging and motivating.

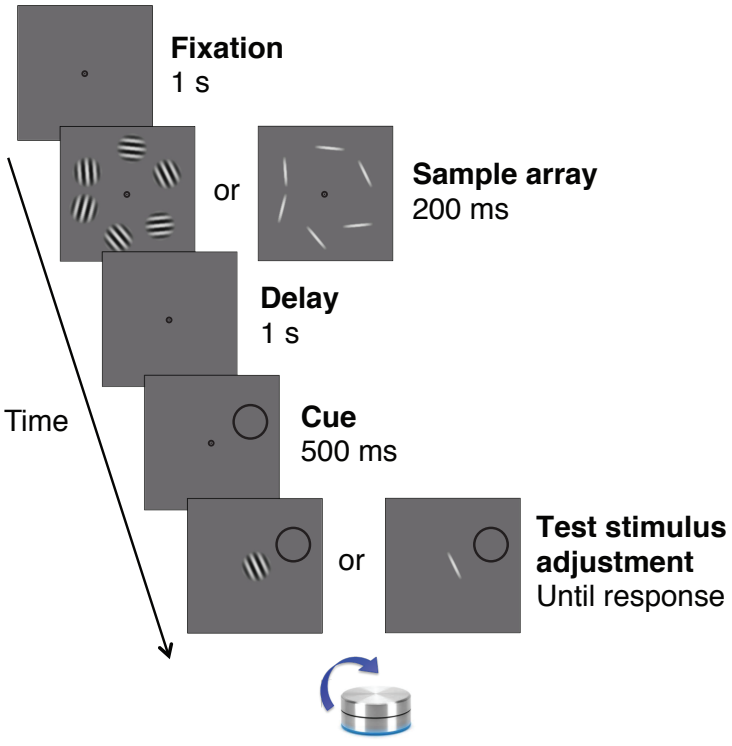

Figure 2. A trial sequence for the delayed estimation task in Experiment 1A. A sample array contained either gratings or lines (2 or 6 items). To highlight the visual impact of changing the stimulus form, the grating and line arrays are shown with matched orientations and locations. In the actual experiment, these values were randomly determined for each stimulus type.

Each participant completed two 1-hour sessions. Each session consisted of 10 alternating blocks of the two stimulus types (gratings and lines), 40 trials each. The order of the two stimulus types was counterbalanced across participants and sessions. Two set sizes (2 and 6) were randomly intermixed within each experimental block. For each participant, I obtained a total of 800 trials, resulting in 200 trials for each condition.

Data Analysis

The delayed estimation data were analyzed using the mixture model proposed by Zhang and Luck (2008). The model has two parameters: SD and $P_{failure}$ (see **Figure 1A**). The SD parameter determines the spread of the von Mises distribution, which is inversely proportional to the precision of the stored representation. The location (i.e., central tendency) of the von Mises distribution was fixed at zero, assuming no systematic bias away from the true stimulus orientation. The $P_{failure}$ parameter determines the height of the uniform component, representing the probability that the probed item is lost from memory. The mixture model was fitted separately to each experimental condition for each participant, using maximum likelihood estimation. The $P_{failure}$ and SD estimates were each submitted to a repeated-measures ANOVA, with stimulus type (line and grating) and set size (2 and 6) as within-subjects factors.

I calculated the number of stored orientations (K) for each condition using the formula, $K = (1- P_{failure}) \times N$, where N denotes the number of items in the memory array, or the set size. While I will plot the K values to facilitate comparison of capacity estimates across different set size conditions, the statistical tests were performed on the raw $P_{failure}$ data.

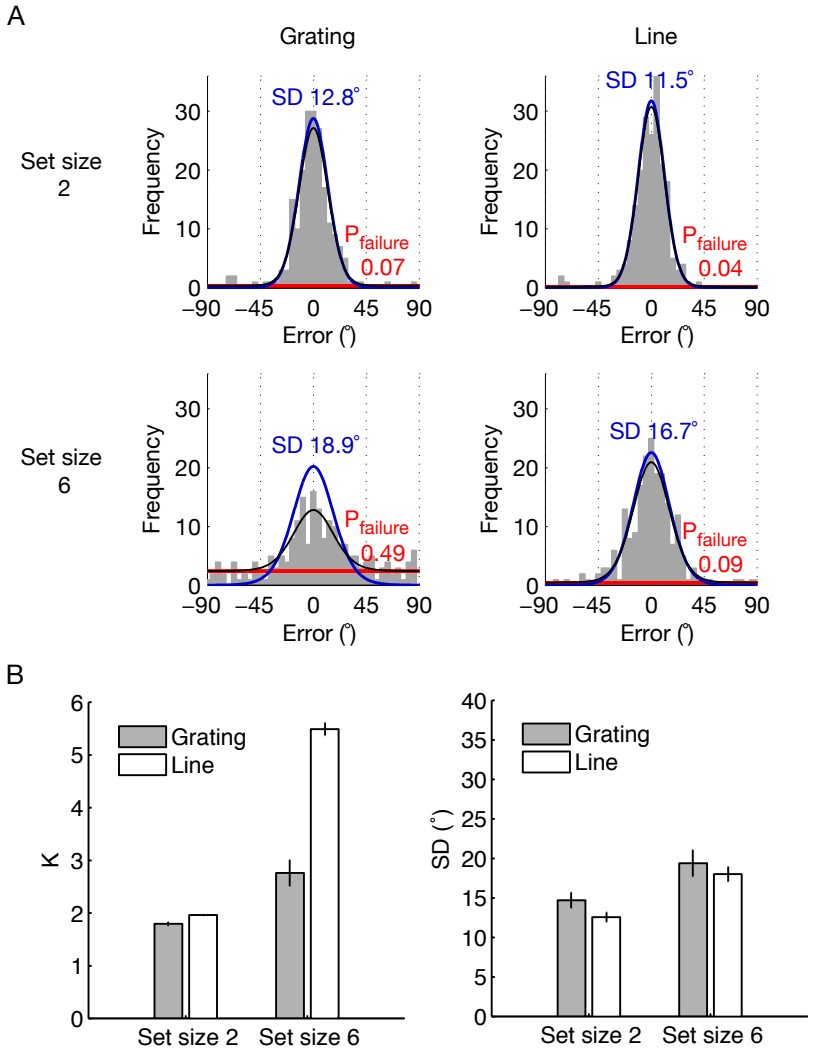

Figure 3. Results of Experiment 1A. (A) A representative participant's response error histograms for delayed estimation of gratings and lines, at set sizes 2 and 6. All histograms consist of 40 bins, with a bin width of 4.5˚. For each condition, the best-fitting mixture distribution for this participant is shown by a black curve. The fitted memory precision (SD) and guess rate ($P_{failure}$) components are indicated with the von Mises (blue) and uniform (red) distributions, respectively. (B) Results of the mixture model analysis (N = 12). The estimates of capacity (K) and memory precision (SD) are shown on the left and right panels, respectively. A *higher* K value represents a greater memory capacity, whereas a *lower* SD value represents greater memory precision. The error bars represent between-subject standard error of the mean (SEM).

Results

The distribution of response errors for a representative participant is presented in **Figure 3A.** As evident in this participant's response error histogram, VWM performance

was very similar for gratings and lines at set size 2, with a modest advantage in memory precision for lines and little guessing in general. On the other hand, there was a prominent elevation of uniform guessing responses for gratings compared to lines at set size 6, suggesting a reduced capacity for gratings. This trend was mirrored in the group-averaged results (**Figure 3B**), which showed that participants could hold an average of 2.8 grating orientations, but as many as 5.5 line orientations. The ANOVA confirmed that $P_{failure}$ was significantly lower for lines than for gratings, $F(1, 11) = 69.67$, $p < .001$, $\eta_p^2 = 0.86$. The effect of stimulus type on $P_{failure}$ significantly interacted with set size, $F(1, 11) = 102.46$, $p < .001$, $\eta_p^2 = 0.90$, indicating that the capacity advantage for lines over gratings was much larger at set size 6 compared to set size 2.

Compared to this dramatic difference in K at set size 6 (**Figure 3B**, left), memory precision (SD; **Figure 3B**, right) was very comparable between gratings and lines at both set sizes, with a modest precision benefit for lines. The SD for lines was marginally smaller than that for gratings, $F(1, 11) = 4.83$, $p = .0503$, $\eta_p^2 = 0.31$. For both gratings and lines, there was a general decline in memory precision at the larger set size, $F(1, 11) = 43.97$, $p < .001$, $\eta_p^2 = 0.80$, but this effect did not interact with the stimulus type, $F(1, 11) = 0.26$, $p > .250$.

In summary, I found that VWM precision was comparable for line and grating orientation, but that estimated capacity was about twice as large for lines (K = 5.5) as for gratings (K = 2.8). This capacity advantage for lines is similar in magnitude to that reported by Alvarez and Cavanagh (2008) (K = 4.0 for lines, 2.3 for gratings), relying on a change detection paradigm. The present results demonstrate that the VWM advantage for lines over gratings primarily lies in the number, rather than the resolution, of stored orientations.

Experiment 1B

The capacity estimate of 5.5 line items in Experiment 1A was surprisingly high, while the capacity estimate of 2.8 grating items was comparable to the estimates obtained in the previous studies using a similar paradigm and stimuli (e.g., K = 2.6~2.7 in Rademaker, Tredway, & Tong, 2012; K = 3.6 in Pratte, Park, Rademaker, & Tong, 2017) or that obtained in color VWM tasks (K = 2.2~2.5 in Zhang & Luck, 2008; 2009; 2011).

I considered whether the unusually high capacity for lines be due to the circular layout of each sample array. Such a circular arrangement of line orientations might allow for the perception of a global shape in some cases, such as a triangle or a diamond. To address this issue, I designed a follow-up experiment, in which 6 gratings or lines were presented at random locations on a 4 x 4 grid (see **Figure 12)**. If lines can be stored more efficiently than gratings regardless of the regularity of their spatial arrangements, the same capacity benefits should occur with this grid layout.

Method

Participants

12 participants (6 male; ages 19-26 years) took part in Experiment 1B. The stimuli and procedure were identical to Experiment 1A, except for the following changes.

Stimuli and procedure

A new line stimulus was used, created by isolating a central bar of a grating stimulus (see General Method). A sample array contained 6 gratings or lines, presented in random positions within a 4 x 4 square grid (15˚ x 15˚) (see Figure 4). The x, y coordinate of each item was jittered ±0.6˚ from the center of each cell (3.75˚ x 3.75˚). The minimum edge-to-edge distance between items was 0.55˚, and the maximum eccentricity of the center of an item was 8.80˚.

Each participant completed a 1-hour session, consisting of 8 alternating blocks of grating and line stimuli, 54 trials each. Only set size 6 was tested in this experiment to reduce the duration of the study and to ensure adequate power to test for differences in memory capacity. Each participant completed a total of 432 trials, resulting in 216 trials for each stimulus type.

Results

Results of the mixture model analysis are shown in **Figure 4B**. With the grid array, the estimated capacity (K) was 2.1 for grating orientations and 5.1 for line orientations, replicating the large capacity advantage for line stimuli. A paired-samples t-test (2-tailed) on the $P_{failure}$ data confirmed that this capacity difference was highly significant, $t(11) =$

15.91, $p < .001$, Cohen's $d = 4.59$. On the other hand, the precision of stored orientation (SD) was not significantly different between gratings and lines, $t(11) = 1.05$, $p > .250$. Thus, I conclude that the large VWM capacity for line orientations is not just limited to a circular stimulus layout, but generalizes to more irregular spatial arrangements.

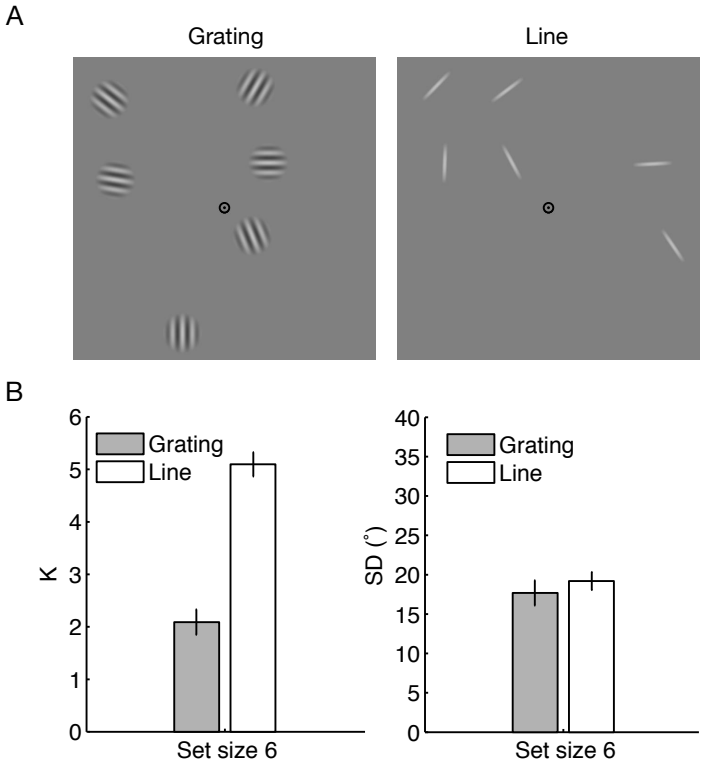

Figure 4. Example stimuli used in Experiment 1B and the mixture model results. (A) A sample array consisted of six gratings or lines, presented at various locations within a 4 x 4 grid. (B) The estimates of capacity (K) and memory precision (SD) are shown on the left and right panels, respectively (N = 12). The error bars represent between-subject standard error of the mean (SEM).

Experiment 2

While the results from Experiments 1A and 1B suggest that about twice as many line orientations as grating orientations can be stored in VWM, an alternative explanation is that VWM capacity might be the same for the two stimulus types, but gratings might take longer to encode into VWM than lines. A previous study suggests that two oriented gratings cannot be simultaneously consolidated into VWM as effectively as two sequentially presented gratings (Becker, Miller, & Liu, 2012). Specifically, orientation

judgment performance was found to be worse in the simultaneous condition than in the sequential condition when the two gratings were presented for 50–130 ms and then masked. Thus, it is possible that the current sample duration of 200 ms was not sufficient to encode all grating orientations that could be held in VWM.

The purpose of Experiment 2 was to rule out the possibility that the gross differences in the capacity estimates for gratings and lines might be driven by faster encoding of line orientations into VWM. The time allowed for VWM encoding was manipulated by presenting the sample array for 200 or 1000 ms, followed by a patterned mask to limit the potential reliance on iconic memory (Averbach & Coriell, 1961; Sperling, 1960). At stimulus onset asynchronies (SOAs) of 200 or 1000 ms, patterned masks are believed to interrupt VWM consolidation processes, without affecting the early sensory and perceptual processing of the sample array (Vogel, Woodman, & Luck, 2006).

If VWM encoding were to require more time for gratings than for lines, then an increased display duration should allow for more gratings to be stored in VWM, thereby reducing the difference in capacity between the two stimulus types. In addition, if a persistent iconic memory for lines contributed to the enlarged storage capacity in the previous experiments, the patterned mask should markedly reduce the estimated capacity for lines. By contrast, if VWM capacity is truly enhanced for lines over gratings due to differences in storage efficiency, the pronounced advantage in memory capacity for line orientation should persist across these manipulations.

Method

Participants

16 participants (4 male; ages 19-28 years) took part in Experiment 2. Stimuli and procedure were identical to Experiment 1A, except for the following changes.

Stimuli and procedure

Example stimuli in Experiment 2 are shown in **Figure 5A**. A sample array consisted of six gratings or lines presented at various locations around the central fixation at a 4˚ eccentricity. A line stimulus was taken from the central region of a grating (see General Method). The patterned mask stimulus consisted of a low-pass filtered uniform

noise (16˚ x 16˚) with a cut-off frequency of 4 cycles/˚ and a mean RMS contrast of 0.115 (s.d. = 0.005), contained within an annulus (mean diameter 8˚) covering the circular sample array. The annulus was contrast modulated using a single cycle of a cosine function: $\left\{\left(cos\ \frac{d\pi}{4} + \pi\right)/2 + \frac{1}{2}\right\}^3$, where d denotes the distance from the center ($0^\circ \leq d \leq 8^\circ$).

On each trial, a sample array was presented for either 200 ms or 1000 ms, and followed by a patterned mask for 200 ms. After an 800-ms delay, one of the items was randomly probed for report. The interstimulus interval (ISI) between sample and the spatial cue was kept at 1 s, consistent with Experiments 1A and 1B. The two stimulus types were presented in separate blocks of trials. Within each block, the encoding durations were randomly intermixed across trials. Each participant completed two 1-hr sessions, consisting of 12 alternating blocks of grating and line stimuli, 36 trials each. Each participant completed a total of 864 trials, resulting in 216 trials for each experimental condition.

Results

As shown in **Figure 5B**, observers exhibited a similar advantage in VWM capacity for lines over gratings regardless of whether stimuli were presented for 200 ms (2.2 gratings, 4.6 lines) or 1000 ms (2.8 gratings, 5.0 lines), followed by a patterned mask. The main effect of stimulus type on $P_{failure}$ was highly significant, $F(1, 15) = 67.37$, $p < .001$, $\eta_p^2 = 0.82$. The longer encoding duration led to a modest improvement in capacity, $F(1, 15) = 13.58$, $p = .002$, $\eta_p^2 = 0.48$, but the magnitude of this benefit was not different between gratings and lines, $F(1, 15) = .35$, $p > .250$. Thus, even with the five-fold increase in encoding duration, observers could successfully maintain about twice as many lines as gratings. This rules out the alternative hypothesis that the capacity advantage for line orientation arises from faster encoding of lines than gratings. As in Experiment 1A, there was a subtle advantage in memory precision for lines over gratings, but this effect did not reach significance, $F(1, 15) = 2.53$, $p = .133$, $\eta_p^2 = 0.14$. The increased encoding duration did not significantly improve memory precision for either stimulus type, $F(1, 15) = 3.04$, $p = .102$, $\eta_p^2 = 0.17$.

In summary, the capacity estimates for gratings and lines both improved modestly as the encoding duration increased from 200 ms to 1000 ms, but the large capacity advantage for lines was still preserved. Thus, the large difference in memory capacity for

the two stimulus types cannot be attributed to insufficient encoding time for gratings or a more persistent iconic traces for lines, but instead, seems to reflect a true advantage in VWM capacity for lines over gratings.

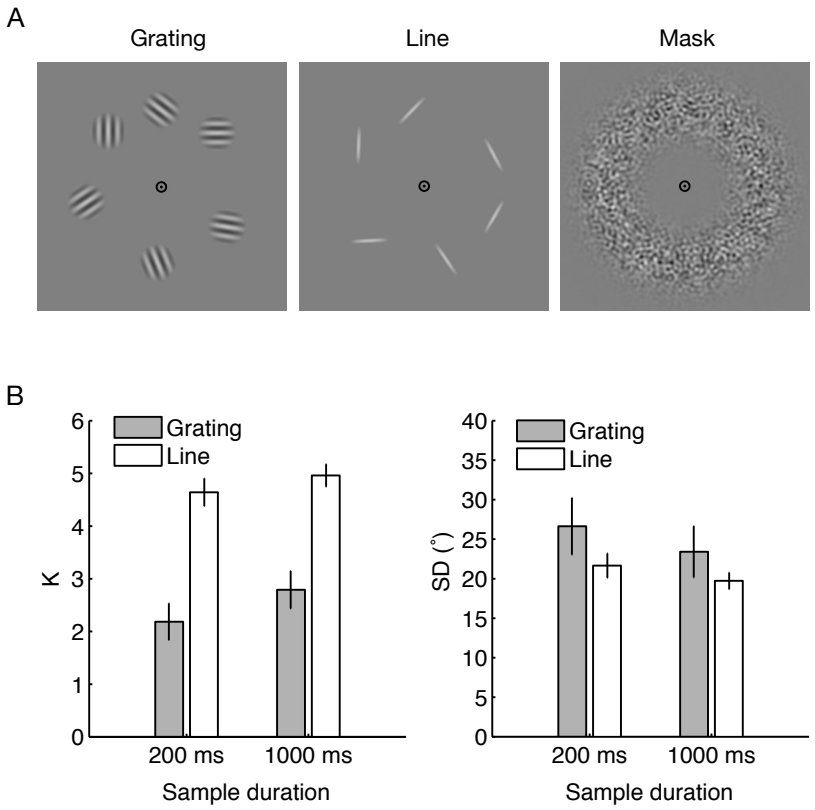

Figure 5. Example stimuli and mixture model results in Experiment 2. (A) Examples of sample arrays and a patterned mask. (B) Estimates of capacity and memory precision (N = 16) The error bars represent between-subject standard error of the mean (SEM).

Experiment 3

One possible mechanism underlying the enhanced VWM capacity for line orientation is that multiple line orientations might be readily grouped into a higher-order configuration, which can then be efficiently stored as a set of coherent units or chunks in VWM (Miller, 1956). In Experiment 3, I wanted to test this *visual grouping hypothesis*.

To disrupt visual grouping processes, six gratings or lines were sequentially presented at various locations, and sequential presentation performance was compared with the standard approach of simultaneous presentation (see **Figure 6A**). In the earlier study by Alvarez and Cavanagh (2008), the superior performance for lines over gratings was

observed even when the items were presented one at a time, suggesting that visual grouping was not necessary to show the advantage for lines over gratings. However, since the magnitude of the advantage for lines was not directly compared between the simultaneous and sequential conditions, it remains to be determined whether higher-order grouping among line items allows for an even greater storage advantage.

If the efficient storage of lines relies on visual grouping of simultaneously presented lines, sequential presentation should severely impair VWM capacity for lines, to a greater extent than it does the capacity for gratings. As a result, the capacity difference between the two stimulus types should become much smaller in the sequential condition compared to the simultaneous condition. Alternatively, if an isolated line itself can be stored more efficiently than an isolated grating, sequential presentation should not affect the capacity advantage for lines over gratings.

Method

Participants

16 participants (9 male; ages 18-30 years) took part in Experiment 3. Stimuli and procedure were identical to Experiment 2, except for the changes indicated below.

Stimuli and procedure

A schematic of the trial sequences for the simultaneous (Sim.) and sequential (Seq.) conditions are shown in **Figure 6A**. A sample array containing 6 gratings or lines was presented either simultaneously for 200 ms, or sequentially for 200 ms per item, with an interstimulus interval (ISI) of 400 ms. No patterned masks were presented. In the sequential condition, the items were presented one at a time along the virtual circle surrounding central fixation, in a clockwise or counter-clockwise direction. After a 1-s delay following the last item (or 4 s from the first item in the sequence), a spatial cue indicated which item to report. All four conditions (2 stimulus types x 2 presentation modes) were presented in separate blocks of trials. Within each block of the sequential condition, clockwise or counter-clockwise sequences were randomly intermixed across trials. The spatial position of the first item in the sequence was also randomly selected on each trial.

Each participant completed two 1-hr sessions. Each session consisted of 12 blocks, 25 trials each. The two stimulus types alternated across blocks, in a counterbalanced order. For each stimulus type, the two presentation conditions were randomly shuffled across blocks. Each participant completed a total of 600 trials, resulting in 150 trials for each condition.

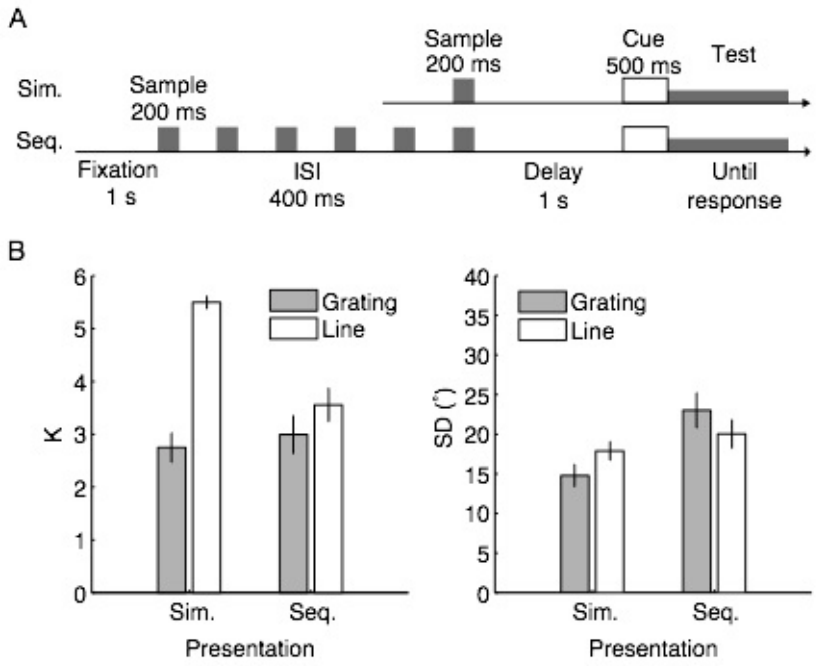

Figure 6. Timing of events and mixture model results in Experiment 3. (A) In the simultaneous condition (Sim.), an array of six gratings or lines were presented all at the same time for 200 ms. In the sequential condition (Seq.), six gratings or lines were presented one after another, for 200 ms each, in a clockwise or counter-clockwise direction along the circular array. (B) The mixture model results for Experiment 3 (N = 16). The error bars represent between-subject standard error of the mean (SEM).

Results

The mixture model results (**Figure 6B**) showed that capacity estimates for gratings did not significantly change between simultaneous (K = 2.7) and sequential (K = 3.0) presentation, $t(15) = 0.95$, $p > .250$, whereas capacity estimates for lines drop from 5.5 to 3.6 items with sequential presentation, $t(15) = 7.51$, $p < .001$, Cohen's $d = 1.88$. Lines showed a small, marginally significant capacity advantage over gratings in the sequential condition, $t(15) = 2.00$, $p = .064$, Cohen's $d = 0.49$, but the magnitude of this advantage

($\Delta K = 0.6$) was substantially smaller than that observed in the simultaneous condition ($\Delta K = 2.7$; $t(15) = 9.51$, $p < .001$, Cohen's $d = 2.38$), as indicated by a significant interaction between stimulus type and presentation mode on $P_{failure}$, $F(1, 15) = 76.37$, $p < .001$, $\eta_p^2 = 0.84$.

While the presentation mode did not affect the number of gratings that could be stored in VWM, memory precision for gratings was significantly impaired by sequential presentation, $t(15) = 3.68$, $p = .002$, Cohen's $d = 0.92$. On the other hand, memory precision for lines remained unchanged, $t(15) = 1.10$, $p > 0.250$. The selective impairment in gratings' precision due to sequential presentation was confirmed by a significant interaction of presentation mode and stimulus type on SD, $F(1, 15) = 11.95$, $p = .004$, $\eta_p^2 = 0.44$.

Sequential presentation severely impaired VWM capacity for line orientations, presumably because this manipulation impaired people's ability to group multiple line elements across successive frames. By contrast, the capacity for gratings remained unchanged with sequential presentation, with some impairment in memory precision. This led to a pronounced reduction of the capacity advantage for lines in the sequential condition, consistent with the hypothesis that visual grouping plays a crucial role in the efficient storage of line orientation. The small, residual capacity advantage for lines in the sequential condition might have occurred because sequential presentation severely interferes with the visual grouping process but might not entirely eliminate it.

These findings suggest that estimates of VWM capacity can be greatly inflated if the stimuli can be readily grouped into higher-order patterns. The degree of inflation might depend on an individual's ability to group multiple items into a chunk, as well as the degree to which the stimuli afford such visual grouping. It is reasonable to predict some degree of correlation between VWM performance across stimulus types and presentation modes given that individuals' core cognitive capacities would underlie performance in a broad range of cognitive tasks. However, if lines tend to be more readily grouped than gratings, and if different individuals have different proclivities to rely on visual grouping, then the capacity estimates for gratings and lines may not be strongly correlated, especially in the simultaneous condition. On the other hand, if sequential presentation disrupts visual grouping of lines, the capacity estimates for the two stimulus types may be more strongly

correlated in the sequential condition. To explore this potential relationship, I conducted correlation analyses comparing individual participants' VWM performance across conditions (stimulus type x presentation mode). Pearson's correlation was calculated for all pairs of conditions, on the K and SD estimates. The results of the key comparisons are presented in **Figure 7**.

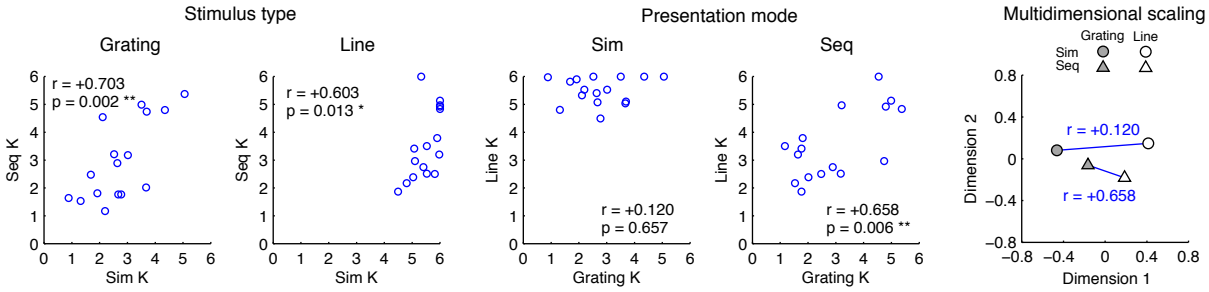

Figure 7. Correlation analyses of VWM capacity in Experiment 3. The scatterplots show the correlations of individual participants' VWM capacity (K) estimates across different conditions (N = 16). The first two panels show the relationship between simultaneous and sequential presentation performance within each stimulus type. The next two panels show the relationship between grating and line performance within each presentation condition. On the rightmost panel, the correlation strengths among the four conditions (presentation mode x stimulus type) are depicted using a multidimensional scaling (MDS) technique. The MDS was based on the dissimilarity score (*D*), measured as 1 minus the correlation (*D* = 0 for perfect positive correlation, 1 for no correlation, 2 for perfect negative correlation).

As shown in the first two panels of **Figure 7**, individual capacity estimates for gratings were strongly correlated between simultaneous and sequential presentations, $r(14) = .70$, $p = .002$, and a similar relationship was observed for the capacity estimates for lines, $r(14) = .60$, $p = .013$. On the other hand, the grating capacity showed little correlation with the line capacity in the simultaneous condition (third panel), $r(14) = .12$, $p > .250$, consistent with the possibility that the line capacity might be additionally mediated by the individual's visual grouping ability. Critically, a positive correlation between grating and line capacities emerged in the sequential condition (fourth panel), $r(14) = .66$, $p = .006$, conditions in which the capacity estimates for lines should be less affected by visual grouping strategies.

To illustrate the relative correlation strengths across all four conditions, I conducted a multidimensional scaling (MDS) analysis, using a 4 x 4 dissimilarity matrix (*D*),

calculated as $D = 1 - r$. The MDS plot (**Figure 7**, rightmost panel) revealed an overall separation of grating and line capacities, but the distance between the two stimulus types was substantially reduced in the sequential condition (filled and empty triangles) compared to that in the simultaneous condition (filled and empty circles). These correlation results suggest that the disruption of visual grouping processes for lines in the sequential condition make the capacity for lines more similar to that for gratings. Thus, visual grouping plays an important role in the enhanced VWM capacity for lines, and visual grouping abilities can be distinguished from one's core cognitive capacity.

Experiment 4

So far, I have relied on the delayed estimation paradigm to compare the VWM precision and capacity for gratings and lines. The fact that sequential presentation virtually eliminated the capacity advantage for line orientation (Experiment 3) supports my hypothesis that visual grouping mediates the capacity advantage for lines. In Experiment 4, I sought to provide converging evidence for visual grouping of lines using a change detection paradigm (Luck & Vogel, 1997). Change detection performance depends not only on how much information is encoded from the sample array, but also on how efficiently the stored representation can be compared with the probe array (Hollingworth, 2003; Makovski, Sussman, & Jiang, 2008; Simons, Chabris, Schnur, & Levin, 2002). I can capitalize on this requirement of memory comparison to tap into the organization of stored representations.

One factor that could facilitate the memory comparison process is the consistency of the global configuration between the sample and probe arrays, especially when the memory representation includes higher-order, relational information among items (Jiang, Olson, & Chun, 2000). If multiple items are stored as higher-order patterns, such representations would be more likely to be disrupted by an isolated probe item, compared to change detection involving a whole-array probe that preserves the original context. Thus, I predicted that change detection for lines should benefit from the whole-array probes, relative to the single-item probe condition. On the other hand, if individual items are stored as separate distinct representations, as would be the case for gratings, the whole-

array probe should not facilitate the comparison process, and thus, the whole-array benefit should be absent for gratings.

Method

Participants

Sixteen participants (8 male; ages 18-26 years) took part in Experiment 4. Stimuli were identical to Experiment 3. Participants performed a change detection task on these stimuli, instead of a delayed estimation task.

Stimuli and procedure

The experimental design (whole array or single-item probe) and stimuli for the change detection tasks are shown in **Figure 8A**. Each trial began with a fixation period of 1 s. A sample array containing 6 gratings or lines was presented for 200 ms, followed by a 1-s delay, and a probe array for 200 ms. For each stimulus type, the probe array contained either a single item presented at one of the six locations or a whole array of six items at the corresponding locations. For the single-item probe, the probed item was rotated by 90˚ on half of the trials, and remained the same on the other half. Participants were asked to press "1" on the number pad if the probed item remained the *same* as the sample array, and press "2" if it was *different*. For the whole-array probe, a change could occur in one of the items, and the rest of the array always remained the same. Participants were asked to press "1" if the sample and probe arrays were the *same*, and press "2" if one of the items was *different*. Participants were told that accuracy was more important than speed, and that only their accuracy data would be analyzed. All four conditions (2 stimulus types x 2 probe types) were presented in separate blocks of trials. Within each block, the change and no-change trials were randomly intermixed. No auditory feedback was provided on each trial. Instead, participants received the average accuracy score at the end of each block.

Each participant completed a 1-hr session, consisting of 8 blocks of 60 trials. The four experimental conditions were presented in a randomized order across blocks. I obtained a total of 480 trials for each participant, resulting in 60 change- and 60 no-change trials for each condition.

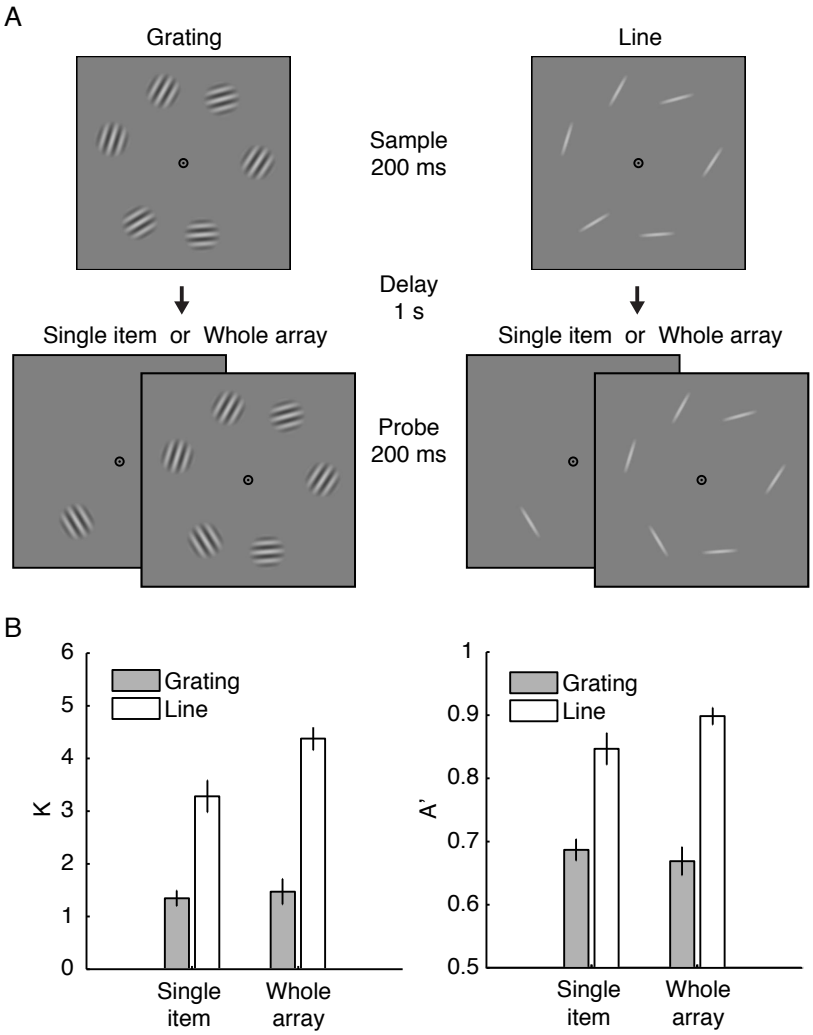

Figure 8. Example stimuli and results in Experiment 4. (A) Trial sequence for the change detection task with two types of probe. (B) Results from the change detection task (N = 16). Two measures of change detection performance, *K* (capacity) and *A'*, are plotted as a function of stimulus types and probe types (single item and whole array), on the left and right panels, respectively. The error bars represent between-subject standard error of the mean (SEM).

Data Analysis

I examined change detection performance using two different measures: The capacity (*K*) measure (Cowan, 2001; Pashler, 1988), which assumes discrete, all-or-none representation of items in working memory, and the measure of signal detection sensitivity (*A'*) (Grier, 1971; Pollack & Norman, 1964), which assumes graded representation. The results from the delayed estimation task indicated that VWM performance for gratings and

lines is mainly limited by the number of items that can be stored, and that the resolution of the stored orientation (SD = 15˚-20˚) would be sufficient to allow a reliable detection of a 90˚ change. Nevertheless, I examined both the discrete and continuous measures of VWM performance since different types of probe (single item or whole array) could affect the memory comparison process in an all-or-none or graded manner. While these measures are based on fundamentally different conceptualizations of VWM, their results should converge when there are large differences in performance accuracy across conditions.

To measure the capacity in the single-item and whole-array probe tasks, I used two different formulae, proposed by Cowan (2001) and Pashler (1988), respectively (see Rouder et al., 2011, for discussion on the choice between two measures). Cowan's K was calculated using the formula, $K = N (h - f)$, where N is the number of items on the sample array, h is the observed hit rate (i.e., reporting change when there is a change), and f is the observed false alarm rate (i.e., falsely reporting change when there is none). Pashler's K was calculated using the formula, $K = N (h - f) / (1 - f)$. The difference in the denominator reflects the fact that, with a single-item probe, guessing is needed only when the probed item is not in memory, whereas a whole-array probe requires guessing about the items that are not in memory *whenever* change is not detected among the stored items.

A' provides a non-parametric index of sensitivity that estimates the area under the receiver operating characteristic (ROC) curve (Grier, 1971; Pollack & Norman, 1964). *A'* ranges between 0.5 (chance performance) and 1.0 (perfect performance). I use *A'* instead of the more common measure d' (Green & Swets, 1966) since *A'* does not make strong assumptions about the underlying psychological mechanisms or the shape of the ROC curve. Moreover, *A'* can be defined even when hit rates are perfect or false alarm rates are zero. A' was calculated using the formula provided by Grier (1971): $A' = 0.5 + (h - f) (1 + h - f) / [4 h (1 - f)]$. If the false alarm rate (f) exceeded the hit rate (h), the following modification by Aaronson & Watts (1987) was used: $A' = 0.5 - (f - h) (1 + f - h) / [4 f (1 - h)]$.

Both K and *A'* utilize hit and false alarm rates, and provide a measure of information available for change detection, independent of guessing strategies or biases. These measures were calculated for each condition in each participant, based on the hit and false alarm rates in each condition.

Results

As shown in **Figure 8B**, performance was much better overall for lines than for gratings, consistent with my earlier experiments. Of particular interest, memory performance for line-orientation displays was significantly better when tested with whole-array probes, as compared to the single-item probe, $t(15) = 5.43$, $p < .001$, Cohen's $d = 1.36$ (*K*); $t(15) = 3.37$, $p = .004$, Cohen's $d = 0.84$ (*A'*). This could be contrasted with memory performance for gratings, which did not significantly differ for the two probe conditions, $t(15) = 0.60$, $p > .250$ (*K*); $t(15) = 1.01$, $p > .250$ (*A'*). As a result, the performance advantage for lines became even larger in the whole-array probe condition, $t(15) = 10.29$, $p < .001$, Cohen's $d = 2.57$ (*K*); $t(15) = 9.00$, $p < .001$, Cohen's $d = 2.25$ (*A'*). This was confirmed by a significant interaction of stimulus type and probe type, $F(1, 15) = 17.48$, $p < .001$ (*K*), $\eta_p^2 = 0.54$; $F(1, 15) = 10.61$, $p = .005$, $\eta_p^2 = 0.41$ (*A'*).

The improved change detection performance for lines with the whole-array probe relative to the single-item probe is remarkable if one assumes that multiple comparisons between remembered items and currently perceived items must be performed to detect a change. For the single-item probe, one must differentiate between change {1} from no-change {0} just for the probed item, so the number of comparisons that need to be made is 1 for both stimulus types. For the whole-array probe, on the other hand, one must differentiate five no-changes plus one change {0, 0, 0, 0, 0, 1} from six no-changes {0, 0, 0, 0, 0, 0}, resulting in a total of 6 comparisons. If each item were independently represented in VWM in an all-or-none manner, as assumed by the slot model, about 2.4 times as many comparisons would be made for a line array (Cowan's $K = 3.3$) as for a grating array (Cowan's $K = 1.3$), which would increase the decision noise more sharply for lines than for gratings. However, I found the opposite result: Line performance was improved with the whole-array probe (Pashler's $K = 4.3$), whereas grating performance remained the same (Pashler's $K = 1.5$).

It appears that, when it comes to VWM for lines, probing only one item, rather than the whole array, is detrimental to the memory comparison process. Such impairment is expected if a line item is stored as part of a larger pattern, and if the stored pattern is easily disrupted by a single-item probe, due to the large configural change. On the other hand, the memory comparison process for gratings was not disrupted by a single-item probe

presumably because the individual gratings were represented separately from one another, nor did it meaningfully benefit from the smaller number of comparisons to be made because no more than a couple of gratings could be stored in VWM. The present results are consistent with the idea that configural information plays an important role in VWM for lines, but has a much lesser role in VWM for gratings.

Discussion

In the present experiments, I investigated whether the precision and capacity of VWM for orientation differ between gratings and lines. Using a paradigm that required precise reproduction of a remembered feature value, I found that VWM capacity was greatly enhanced for line orientation compared to grating orientation (2.8 gratings; 5.5 lines) whereas the precision of stored orientation was comparable for the two stimulus types (Experiment 1A). The increased VWM capacity for line orientation was not specifically due to the circular placement of the stimuli, as the results were replicated using a grid array (Experiment 1B). Similar results were obtained when stimuli were presented for 200ms or 1000ms and followed by a patterned mask, ruling out the possibility that gross differences in capacity might be driven by faster encoding or more persistent iconic traces for the line stimuli (Experiment 2). These results demonstrate that the capacity of VWM for orientation depends critically on stimulus form.

Next, I evaluated the hypothesis that VWM capacity may be superior for lines because their random arrangements can be more readily organized into perceptual groups. I examined the impact of disrupting the visual grouping processes by sequentially presenting the items at various locations (Experiment 3). Compared to simultaneous presentation, memory capacity for lines was substantially impaired by sequential presentation (5.5 vs. 3.6 lines), whereas the capacity for gratings remained unchanged, suggesting that higher-order patterns formed by lines could be more efficiently maintained in VWM. To further investigate the difference in representational format between gratings and lines, I employed a change detection paradigm, presenting a probe item either in isolation or in context of the other items (Experiment 4). For lines, the single-item probe led to significantly worse change detection performance than the whole-array probe, suggesting that information about the item stored as part of a pattern was more likely to be disrupted by an isolated

probe item than by a whole-array probe. On the other hand, change detection performance for gratings did not depend on the probe, suggesting that the gratings were represented as isolated objects. These results provide converging evidence that visual grouping plays an important role in the efficient storage of line orientations in VWM.

The present experiments are certainly consistent with an earlier report of superior change detection performance for lines than for gratings by Alvarez and Cavanagh (2008). Equipped with a more recently developed modeling approach that allowed for a fine-grained characterization of VWM capacity and precision (Zhang & Luck, 2008), I show that VWM can store nearly twice as many line orientations as grating orientations, with just as good representational precision. Moreover, my experiments provide compelling new evidence that visual grouping underlies this dramatic mnemonic advantage for line orientation, whereas each grating appears to be represented in an independent manner. I find that the capacity advantage for line orientation is largely eliminated when the items are presented one at a time, inconsistent with Alvarez and Cavanagh's original account based on the distinction between boundary and surface features (Grossberg & Mingolla, 1985), which predicts that each line item, on its own, should be stored more efficiently than a grating item.

The *visual grouping hypothesis* is different from the boundary hypothesis in two important ways: First, it posits that a group of line items, rather than each individual line item, constitutes a coherent memory unit or a chunk. Second, it suggests that the capacity that people have for line orientation represents an overestimation of the upper limit on the number of items they can hold in working memory, because multiple line items can be stored as a coherent unit. This is quite different from the view that the reduced capacity for grating orientation reflects a trade-off between the amount of visual detail or complexity of each object and the maximum number of objects that can be stored (Alvarez & Cavanagh, 2004; 2008). Consequently, the *visual grouping hypothesis* predicts that a VWM capacity estimate obtained using gratings or sequentially presented lines might be a better predictor of an individual's true working memory capacity than that obtained using line stimuli, which would be additionally influenced by one's visual grouping abilities. My correlational analyses in Experiment 3 provide some support for this account. It would be an interesting

question for future research to compare the correlations of VWM capacities for gratings and lines with measures of aptitude and fluid intelligence.

In arguing that line stimuli lead to exaggerated K estimates, I are not suggesting that the K estimates for grating stimuli should represent the absolute value of *true* K. A broader implication of the present work is that the core aspect of working memory capacity, such as that represented by Miller's (1956) magical number 7 or Cowan's (2001) 4-chunk limit, might be only meaningful as a latent variable, which cannot be directly measured by any particular working memory paradigm. The *apparent* K estimates would be highly specific to the stimulus factors and task requirements. For example, in the present study, the average K values were 2.7-2.8 gratings and 5.0-5.5 lines in the delayed estimation tasks, but only 1.5 gratings and 4.3 lines in the change detection task with a whole-array probe, and 1.3 gratings and 3.3 lines with a single-item probe.

The K values obtained in the present study were generally lower in the change detection task than in the delayed estimation task. One possible reason is that, with change detection, one likely needs to encode the probe item into VWM in order to compare it with other items previously encoded from the sample array, which might lead to overwriting or impaired access to one of the concurrently held items. While the delayed estimation task also requires encoding of the test stimulus while matching it with one of the items held in VWM, I sought to protect the stored information from the potential interference from the test stimulus, with a careful use of spatial cuing. Specifically, instead of presenting the test stimulus at the same time as the spatial cue indicating the to-be-reported item, as is commonly done in many VWM studies (e.g., Zhang & Luck, 2008; 2009), I delayed the presentation of the test stimulus by 500 ms so that observers could fully focus their attention to accessing the cued item stored in VWM before they were prompted to adjust the feature value of the test stimulus. This procedure is akin to *retrospective cueing*, whereby attention is directed to the contents of VWM ahead of testing them, leading to an improved VWM performance for the cued item, which is called the *retro-cue effect* (see Souza & Oberauer, 2016 for a review). On the other hand, the change detection task did not involve such cueing, leading to a diffuse state of attention over the stored items when the probe stimulus appeared. The change detection K estimates for both stimulus types would have been generally higher had I used the same cueing procedure, which would

promote the access to the stored information before it was interrupted by the probe stimulus (see Landman, Spekreijse, & Lamme, 2003; Sligte, Scholte, & Lamme, 2008, for high K estimates for oriented bars).

Despite these differences in task requirements, the large capacity difference between gratings and lines was reliably observed across different paradigms. It appears that the amount of information stored in VWM for these stimuli can be more accurately measured with the delayed estimation paradigm with retrospective cueing, which minimizes the interruption of the stored information by subsequent visual processing. The delayed estimation paradigm might be more suited for studying the pure storage function of VWM than change detection, as the memory comparison process proves to be another important factor influencing change detection paradigm, in addition to the encoding and maintenance processes.

The present study challenges the common assumption that VWM capacity for simple visual features remains stable across variations in stimulus form. I show that, even within the same feature dimension, the amount of feature information that can be stored in VWM depends critically on the physical appearance of the stimuli carrying those features. I propose that VWM can store orientation information in a highly efficient manner if the stimulus form (i.e., a line vs. a grating) allows multiple items to be more readily perceived as higher-order patterns rather than as independent visual objects. Such visual grouping processes can help alleviate the severe limits of human working memory capacity, in similar ways as VWM's feature storage capacity can benefit from organization of multiple features into a coherent visual object (Luck & Vogel, 1997; Vogel, Woodman, & Vogel, 2001; Xu, 2002). The benefits of chunking have long been known in the case of verbal working memory capacity (Miller, 1956; Ericsson & Chase, 1982), and can lead to dramatic improvements in effective capacity. In comparison, there has been little emphasis on the role chunking in the domain of visual working memory, with the exception of perhaps a few studies that have introduced strong correlations or regularities into the visual displays (e.g., Brady, Konkle, & Alvarez, 2009; Orbán, Fiser, Aslin, & Lengyel, 2008). Given the dramatic impact of stimulus form on VWM capacity, future research on the nature of capacity limits in VWM should consider the perceptual factors that allow for more efficient usage of the limited storage capacity.

Chapter 3

Compression of working memory load by visual grouping: A modeling study

Introduction

In the earlier experiments, I found that VWM could store about twice as many line orientations as grating orientations, with little cost in the fidelity of the stored orientation. The large capacity advantage for lines crucially depended on the simultaneous presentation of the items, suggesting that the enhanced storage efficiency lies not in the representations of each isolated line item, but rather, in inter-item grouping. Based on these results, I proposed the visual grouping hypothesis: VWM can store multiple line items as a coherent unit, with just as good precision as storing each item as a separate, independent unit.

While I previously focused on comparing the upper limits of VWM capacity for gratings and lines at a given set size (i.e., 6 items), further insights into the mechanisms underlying VWM advantage for lines can be gleaned from the systematic relationship between the number and resolution of stored items across a range of set sizes. It is well established that VWM precision systematically declines as the number of remembered items increases (Bays & Husain, 2008; Wilken & Ma, 2004; Zhang & Luck, 2008), reflecting limited working memory resources that must be shared among items. If VWM is viewed as a representational medium that stores independent, noisy copies of perceptual input, which can be averaged to estimate stimulus attributes, then the fidelity of estimation should depend on the availability of "samples" taken from that stimulus (Palmer, 1990; Shaw, 1980; Wilken & Ma, 2004). The maximum fidelity, or the smallest possible variability of the sample mean, can be achieved by taking all of the samples from one stimulus. Assuming a finite supply of samples (n), increasing the set size (s) would proportionally reduce the number of samples allocated to each stimulus (n/s), resulting in an increased variability of the sample mean, as predicted by the central limit theorem ($\frac{SD}{\sqrt{n}}$ vs. $\frac{SD}{\sqrt{n/s}}$).

Current theories of VWM are divided as to whether these noisy samples are best characterized as a small, fixed number of *discrete slots* (Luck & Vogel, 1997; Zhang & Luck, 2008; Pratte, Park, Rademaker, & Tong, 2017) or as *a continuous pool of resources* that can be distributed among numerous items without a fixed upper limit (Bays & Husain,

2008; van den Berg et al., 2012; Wilken & Ma, 2004). Despite these divergent views regarding the upper limit of items that can be stored in VWM at a time, both theories predict a systematic trade-off between the number of items stored and the fidelity of stored representations, which is the result of dividing "a limited supply of representational medium" among multiple items (Ma, Husain, & Bays, 2014).

However, my empirical findings present a conundrum for current models of VWM. Neither the discrete-slots model and nor continuous-resource model predicts such a dramatic difference in VWM performance between two stimulus formats conveying the same task-relevant feature information, tested within the same individual. The same person would appear to have a vastly different number of memory slots or a different amount of memory resources depending on the stimulus format, negating the concept of working memory capacity as a finite and stable quantity. This suggests that current models of VWM capacity should not only consider the capacity-limited, central resources but also incorporate content-specific modulation of storage efficiency or memory resource requirements.

In the present study, I will incorporate the idea of visual grouping into an existing VWM model framework in order to better characterize an individual's VWM performance for grating and line orientations across multiple set sizes. Specifically, I will develop a slots-plus-averaging model with probabilistic grouping, and test whether this model can provide a suitable fit of an individual's working memory performance across changes in stimulus form.

For simplicity, visual grouping will be construed as *lossless data compression*, by which items in the memory array can be grouped into fewer memory units without sacrificing the representational precision of the items within a group. By allowing the degree of grouping to vary across stimulus types while keeping the core capacity of working memory constant within an individual, my model can account for the large difference in the presumed memory capacity for gratings and lines. Moreover, as I will describe below, the model predicts that visual grouping compresses the *effective* set size, thereby systematically altering the relationship between the *nominal* set size and VWM performance. Accounting for such compression should allow for a better characterization of VWM performance for a given stimulus type, especially if the stimulus form affords

strong visual grouping. I will test the fine-grained predictions of the visual grouping hypothesis by fitting the candidate models to the delayed estimation data obtained using a wide range of set sizes (1–8 items). Using a model comparison approach, I will assess the advantage of this new model over the existing model that does not incorporate visual grouping.

Although the idea of data compression by visual grouping could be combined with either the discrete-slots or continuous-resource model, I adopted the former approach because of a straightforward theoretical interpretation of the model parameters (see Discussion for more on this issue). A recent version of the discrete-slots model, called the *slots-plus-averaging* (SA) model (Zhang & Luck, 2008), provides tightly constrained predictions on VWM performance using only two free parameters: The number of memory slots (K) and the precision of each slot (SD_{slot}). The number of memory slots, but not memory precision, has been shown to predict fluid intelligence (Fukuda, Vogel, Mayr, & Awh , 2010) whereas memory precision, but not the number of slots, has been shown to improve with perceptual expertise (Lorenc, Pratte, Angeloni, & Tong, 2014; Scolari, Vogel, & Awh, 2008). Thus, these two parameters seem to reflect distinct aspects of working memory, one that reflects domain-general, core cognitive ability and one that reflects domain-specific or perceptual abilities. Moreover, each of these parameters controls VWM performance in a way that is distinct from visual grouping, allowing for reliable parameter estimation for the combined model as well as a straightforward interpretation of the fitted parameters.

I will start by describing the existing approach of modeling VWM performance in the discrete-slots framework. Based on the known effects of stimulus form on VWM precision and capacity at set size 6, I will make predictions about how VWM performance should vary as a function of set size for each stimulus type if the capacity limit observed at set size 6 were taken to reflect an individual's *true* working memory capacity. I will then contrast these predictions with those based on the new model that incorporates visual grouping into the existing SA model.

The slots-plus-averaging (SA) model

The slots-plus-averaging (SA) model (Zhang & Luck, 2008) is a modification of

the classic item-limit model (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988), which postulates that VWM can store a small, fixed number of items with perfect precision. While the core assumption of a fixed item-limit is kept in the SA model, it integrates the idea that information stored in each memory slot is noisy, and that noise can be reduced by storing the same item in multiple independent slots and by averaging them. The standard SA model predicts the highest memory precision at set size 1 and declining precision as a function of increasing set size until the set size reaches the number of slots, after which the precision remains stable since an item would be either assigned a slot or not memorized at all. While memory precision would vary with increasing set sizes depending on the number of memory slots one has, the SA model constrains the total amount of information that an individual can store to be constant across all set sizes, by assuming a fixed number of memory slots for that individual.

The formal implementation of the SA model followed the previous literature (Pratte et al., 2017; van den Berg et al., 2012). The basic SA model has two free parameters: K (the number of slots) and J_{slot} (a single slot's precision). Precision (denoted J) of the observer's measurement (x) of a stimulus (s) is formally defined as Fisher Information (Cover & Thomas, 2001), which corresponds to the inverse variance, $J = 1/\sigma^2$, if x follows a Gaussian distribution with mean s and standard deviation σ. However, since the orientation space is circular, the measurement is assumed to follow a von Mises (VM) distribution, centered at s with the concentration parameter κ:

$$p(x \mid s) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-s)} \equiv \mathrm{VM}(x; s, \kappa) .$$

$I_0(\kappa)$ is a normalizing constant, denoting the modified Bessel function of the first kind of order zero. The concentration parameter of the von Mises distribution has a one-to-one relationship with Fisher information, given by

$$J = \kappa \frac{I_1(\kappa)}{I_0(\kappa)},$$

where $I_1(\kappa)$ is the modified Bessel function of the first kind of order one. For a given precision J, the corresponding κ can be computed numerically.

The SA model assumes that the observer's response, denoted r, follows a mixture of a von Mises distribution centered at the studied orientation s when the probed item is in

memory, and a uniform distribution when a random guess is made. The precision of in-memory responses (κ) and the probability of guessing (g) vary as a function of the set size (N), and they will be denoted as κ_N and g_N, respectively.

$$p(r \mid s) = (1 - g_N)\mathrm{VM}(r; s, \kappa_N) + \frac{g_N}{2\pi}.$$

When the number of memory items is equal to or exceeds the number of slots ($N \geq K$), an item will be successfully remembered with the probability of $\frac{K}{N}$, and a guess response will occur with the probability $g_N = 1 - \frac{K}{N}$. Each of the remembered items will receive one memory slot, leading to a stable memory precision $\kappa_N = \kappa_{slot}$ (calculated from J_{slot}). On the other hand, when there are *fewer* memory items than the number of slots ($N < K$), guessing will never occur, and an item will receive a variable number of slots (1~ K) depending on the set size. For example, if three slots are evenly distributed across two items, each item will receive one slot on half of the time, and two slots on the other half. The number of slots (S) assigned to an item can be formally expressed as a mixture of two possibilities:

$$S = \begin{cases} S_{\text{low}} = \left\lfloor \frac{K}{N} \right\rfloor, & \text{with probability } P_{\text{low}} = 1 - \frac{K \bmod N}{N} \\ S_{\text{high}} = \left\lfloor \frac{K}{N} \right\rfloor + 1, & \text{with probability } P_{\text{high}} = \frac{K \bmod N}{N} \end{cases},$$

where $\lfloor x \rfloor$ denotes the floor of x. These two possible slot assignments result in two levels of memory precision:

$$J_{\text{low}}(N) = J_{\text{slot}}\left\lfloor \frac{K}{N} \right\rfloor$$
$$J_{\text{high}}(N) = J_{\text{slot}}\left(\left\lfloor \frac{K}{N} \right\rfloor + 1\right),$$

which reflect the fact that the memory precision (J) for an item is directly proportional to the number of slots allocated to that item. $J_{high}(N)$ and $J_{low}(N)$ are then converted to the concentration parameters of the von Mises distribution, $\kappa_{high}(N)$ and $\kappa_{low}(N)$, respectively. Mixing these two von Mises distributions results in the following probability distribution

for in-memory responses:

$$p(x \mid s) \;=\; P_{\text{low}} \cdot \text{VM}(\text{x; s}, \kappa_{\text{low}}(N)) \;+\; P_{\text{high}} \cdot \text{VM}(\text{x; s}, \kappa_{\text{high}}(N)) \,.$$

The outcome distribution is no longer a von Mises distribution, but its expected precision can still be quantified in terms of the circular standard deviation of the response error vector x (Fisher, 1995):

$$circular\ SD(x) = \sqrt{-2\ln(\bar{R})}\,,$$

where $\bar{R}$ denotes the mean resultant length, calculated as:

$$\bar{R} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\cos(x_i)\right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n}\sin(x_i)\right)^2}\,.$$

For a continuous function $p(x \mid s)$, $\bar{R}$ can be computed using numerical integration:

$$\bar{R} = \sqrt{\left(\int_{-\pi}^{\pi}\cos(x)p(x)\,dx\right)^2 + \left(\int_{-\pi}^{\pi}\sin(x)p(x)\,dx\right)^2}\,.$$

While the parameter estimation will rely on the *latent* mixture of the two von Mises distributions without needing to estimate the *SD* of the outcome distribution, these *SD* estimates are useful for summarizing the memory precision at a given set size predicted by the model. Therefore, I will use these *SD* estimates to illustrate and contrast the predictions of the SA and SA-plus-grouping models.

For the present study, the standard SA model will be fitted separately to the grating and line data, requiring a total of four free parameters for each participant: *Grating's K*, *Grating's* J_{slot}, *Line's K* and *Line's* J_{slot}. Separately modeling each stimulus type is a typical approach in the literature when each participant's VWM is tested using one stimulus format and potential differences across different feature dimensions and stimulus materials are of little interest (e.g., Pratte et al., 2017; van den Berg et al., 2012; van den Berg, Awh, & Ma, 2014). I will demonstrate how this approach can lead to a suboptimal characterization of VWM performance when the model's fundamental assumption that items are stored as independent units does not hold. For a more intuitive understanding of the precision parameter (*J*), or Fisher Information, J_{slot} will be converted to the *circular SD* of the corresponding von Mises distribution, and will be reported as SD_{slot} within 0-180˚ space.

Incorporating visual grouping into the standard SA model

The *SA model with grouping* builds upon the existing discrete-slots framework, introducing an element of flexible data compression via stochastic grouping processes. The key difference is that the SA model with grouping treats the set size (N) as *nominal*, and instead relies on a latent variable G denoting the number of *groups* in a memory array as the *effective* memory load. As a result, the guess rate, as well as memory precision, is determined by the number of groups in the memory array relative to the number of slots. The underlying assumption is that the apparent capacity estimate based on the number of items is likely to reflect exaggeration of one's true working memory capacity due to visual grouping.

In addition to the fixed number of slots (K) and the precision for each slot (J_{slot}), the model relies on a free parameter (p) that controls the probability of visual grouping to allow for a flexible apparent capacity. Here, visual grouping is operationally defined as "a process of compressing an array of N items into G groups, or memory units, by stochastically linking a pair of items". One of the considerations in implementing the grouping process was that the effective memory load should continuously vary as a function of the grouping strength. In addition, this model makes an important prediction: since data compression via visual grouping arises from inter-item interactions, the data compression ratio or storage efficiency should increase as a function of the number of items in the memory array. To achieve these general goals, I make the following assumptions:

1) A pair of items are linked through a binomial process (see **Figure 9A**). The linking occurs independently across all pairs of items, with the likelihood of success defined by probability p.

2) For an array of N items, only up to $N - 1$ pairs will be considered, such that a maximum of $N - 1$ links (L) can be made. In theory, there are a total of $\frac{N(N-1)}{2}$ possible pairwise connections among N items, but calculating the independent success probabilities for all pairs of items (i.e., $2^{\frac{N(N-1)}{2}}$) and counting the numbers of groups for all these possibilities becomes intractable as the N increases. Thus, I make a simplifying assumption that N items are represented in a hypothetical one-dimensional space where only the

neighboring items can be grouped together, resulting in a total of $N - 1$ pairs that can be linked. Alternatively, one could arrange the items along a circle, allowing up to N pairs to be linked. In that case, linking either $N - 1$ or N pairs would result in one group of fully linked items (via a transitive rule, see below) although only $N - 1$ links would be required to achieve G of 1. To remove this redundancy as well as to simplify the calculation, I will adopt the former approach of considering only $N - 1$ adjacent pairs on a line. This is a reasonable simplification, resulting in only slightly different expected numbers of groups for a given p, compared to those based on linking N pairs arranged on a circle.

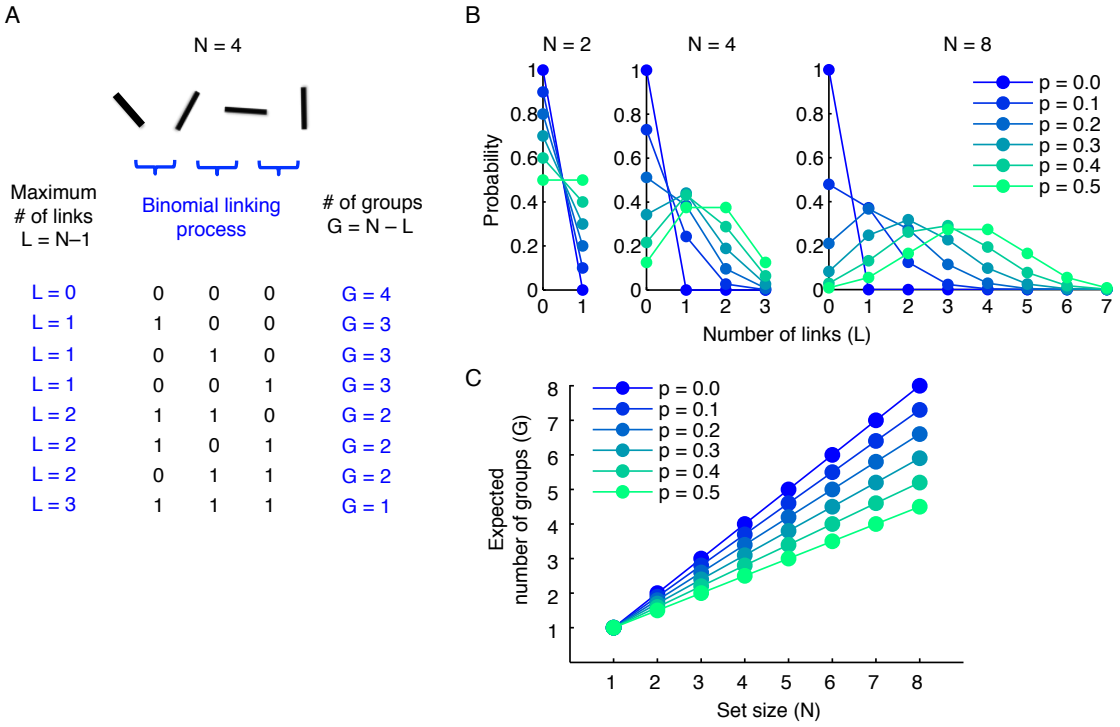

Figure 9. Probabilistic grouping and data compression. (A) An illustration of a hypothetical grouping process, using an example of a four-item array. (B) Predicted binomial distribution of the number of links (L) on arrays of 2, 4, and 8 items (from left to right) as the probability of linking a pair of items (p) varies from 0.0 to 0.5. (C) Expected number of groups (G) as a function of set size (N) and grouping probability (p).

3) Non-adjacent items can be linked through a simple *transitivity rule*: If A is linked to B, and B is linked to C, then A becomes linked to C. Transitive grouping has been proposed as a fundamental grouping mechanism in the visual system, which is essential for correctly binding together image features that belong to the same surface or

object, based on the smooth variations in the local image properties (Geisler & Super, 2000). This specific form of transitive grouping rule results in a straightforward calculation of the number of groups (G) by subtracting the number of links made (L) from the number of items (N): $G = N - L$. L follows a binomial distribution, $L \sim B(n, p)$, where n is the number of pairs that can be linked ($N - 1$) and p is the probability of successful grouping. As illustrated in **Figure 9B**, a higher value of p leads to a higher expected L, as given by $E(L) = (N - 1)p$, which in turn leads to a smaller expected G, $E(G) = N - (N - 1)p = N(1 - p) + p$. This simple linear relationship between the nominal set size (N) and the effective set size (G) is illustrated in **Figure 9C,** across a range of grouping probabilities. It becomes clear that increasing p leads to systematic compression of G relative to N. Moreover, at a given level of p, compression becomes more efficient as N increases, which can be described by the increase in the compression ratio (G/N) from 1.00 at $N = 1$ to 1.78 at $N = 8$ ($E(G) = 4.5$), with $p = 0.5$.

(4) The assumption of an independent linking process across pairs of items implies that there is no strict restriction on the *size* of a group (i.e., the number items that can be grouped together). A group can contain as many as N items when all pairs are linked, and as few as 1 item when an item forms its own group without binding to other items.

(5) Grouping results in *lossless data compression*. In other words, when multiple items are converted into a coherent memory unit, including more items in a group does not affect the precision with which each item is stored. Memory precision is solely determined by the number of slots allocated to a group, and not by the size of a group. For example, if six items are organized into two groups, the item precision remains the same for all different ways of partitioning the array, such as {1, 2, 3} {4, 5, 6} vs. {1, 2} {3, 4, 5, 6}. On the other hand, each of these groups is expected to receive more memory slots on average than if the array happens to be not grouped at all ($G = 6$), enjoying the advantage in memory precision as a result of the smaller effective set size ($G = 2$).

Similar to the SA model, the SA model with grouping assumes that the observer's response follows a mixture of in-memory responses and uniform guessing responses. However, the memory precision and guess rate at each set size are not defined by a fixed value N (i.e., κ_N and g_N), but by a latent mixture of G, comprised of all possible outcomes of the binomial linking process for $N - 1$ pairs, as shown in **Figure 9B**. For $N = 4$, for

instance, the model considers eight possible cases of grouping (*Ls* = 0-7; *Gs* = 1-8) resulting from a binomial distribution of the number of links made, $L \sim B(n, p)$, where n = $N - 1$. All possible grouping cases are combined with their associated probabilities, $\mathrm{P}(L = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Within a given set size *N*, memory precision varies across different grouping cases. For each case of *G*, memory precision is determined by the product of a single slot's precision and the number of slots assigned to a group, which is again a mixture of two possibilities:

$$J(G) = \begin{cases} J_{\text{low}}(G) = J_{\text{slot}} \left\lfloor \frac{K}{G} \right\rfloor, & \text{with probability } 1 - \frac{K \bmod G}{G} \\ J_{\text{high}}(G) = J_{\text{slot}} \left\lfloor \frac{K}{G} \right\rfloor + 1, & \text{with probability } \frac{K \bmod G}{G} \end{cases}$$

The final distribution of in-memory responses at set size *N* consists of multiple von Mises distributions with a variable concentration parameter $\kappa(G)$ corresponding to the *J*(*G*) provided above, weighted by their associated probabilities.

Similarly, even with the fixed number of slots (*K*), the probability of remembering an item varies depending on grouping, leading to a variable guess rate within a given set size *N*:

$$\mathrm{g}(G) = 1 - \frac{K}{G} \quad (\text{if } K \leq G)$$

The expected guess rate (g_N) can be calculated by the averaging the guess rates $\mathrm{g}(G)$'s, weighted by their associated probabilities. This g_N comprises the uniform component for the final response distribution at set size *N*.

In the SA model with grouping, a common capacity parameter will apply for lines and gratings, while the precision of each slot (J_{slot}) and grouping probability (*p*) will be separately estimated for each stimulus type. As a result, a total of five parameters will be fitted to each participant's data: *True K*, *Grating's* J_{slot}, *Grating's p*, *Line's* J_{slot}, and *Line's p*.

Model predictions: Standard SA vs. SA with grouping

The standard SA model allows one to use the VWM capacity and precision estimates previously obtained at set size 6 to generate predictions on the performance

across a range of different set sizes.

Let's assume that the capacity estimates were 3 items for gratings and 6 items for lines, with the same memory precision (SD = 20˚) for both stimulus types. If these estimates were obtained separately, one would assume that working memory consists of 3 slots for gratings and 6 slots for lines, with SD_{slot} of 20˚. This interpretation implies that line stimuli could be stored using twice as many memory slots as grating stimuli. Since the SA model assumes that memory noise (SD) is inversely proportional to the square root of the number of slots (n) allocated to an item ($\mathrm{SD} = \frac{SD_{\mathrm{slot}}}{\sqrt{\mathrm{n}}}$), this assumption of a different numbers of slots available for gratings and lines predicts that SD estimates should differ for these stimuli at set size 1. For example, averaging 3 independent estimates of a grating orientation would result in an SD of 11.5˚ ($\frac{20^{\circ}}{\sqrt{3}}$), whereas averaging 6 independent estimates of a line orientation would result in an SD of 8.2˚ ($\frac{20^{\circ}}{\sqrt{6}}$), which is a much improved precision (i.e., a twofold increase in Fisher information) compared to that of a grating.

The effects of a variable number of slots on the capacity and precision estimates across set sizes are illustrated in **Figure 10A**. The number of stored items (apparent K; left panel) increases with set size until it reaches the item limit set by the number of slots (K). At the same time, memory precision (SD; right panel) worsens as set size increases until SD reaches an asymptote set by the precision of a single slot (SD_{slot}). Importantly, the SD predicted at set size 1 is substantially smaller for $K = 6$ than for $K = 3$ reflecting the precision advantage due to averaging of twice as many slots.

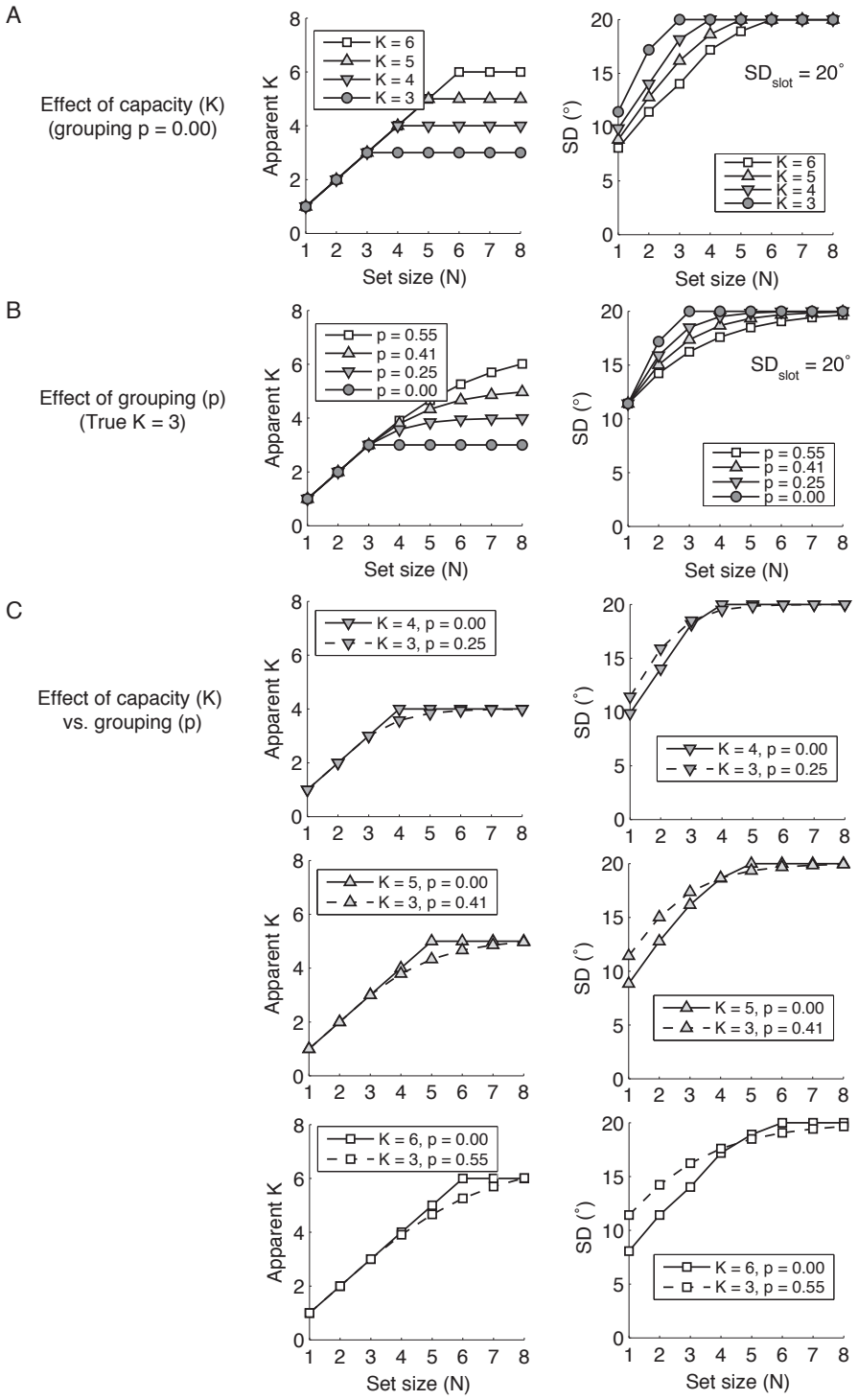

Figure 10. Comparison of predictions from the standard SA model and the SA model with grouping. The predicted VWM capacity (left column) and precision (right column) are shown as a function of set size (*N*). (A) Prediction from the standard SA model assuming a variable number of slots (*Ks* = 3-6) with a single slot's precision (SD_{slot}) of 20˚. (B) Prediction from the SA model *with* grouping assuming a fixed number of slots (*True K* =

3) with SD_{slot} of 20˚, and variable grouping probabilities (*ps* = 0.00-0.55) that result in apparent capacity estimates of 3-6 items. (C) The juxtaposition of the predictions with- and without grouping, pairing the parameters that lead to the same apparent *K* at the largest set size. Solid lines are the predictions from (A) assuming no grouping (*p* = 0) and dashed lines are the predictions from (B) with grouping *p*s of 0.25-0.55. Predictions for *K* = 3 and *p* = 0.00 are identical between panels (A) and (B), and thus not replotted in (C).

A new set of predictions can be generated using the *SA with grouping* framework by varying the grouping probability (*p*) while holding the number of slots constant (*true K*) (see **Figure 10B**). To highlight the deviations from the standard SA model's prediction, I chose a set of grouping probabilities (*p*s = 0.00, 0.25, 0.41, and 0.55) that result in apparent Ks of 3, 4, 5, and 6, at the largest set size (*N* = 8) when combined with *true K* of 3. When *p* is greater than zero, the number of stored items (apparent K; left panel) increases smoothly and continually as the set size increases, without a set upper limit. A higher *p* leads to a steeper increasing function, resulting in a greater capacity advantage when the set size is larger. With perfect grouping (*p* = 1.0), the apparent K would be identical to *N*, since the effective memory load would always be one group, and no items will be lost.

An important distinction between the grouping model and the standard SA model emerges in the predicted effects of memory precision (SD; right panel). A higher *p* tends to improve memory precision, resulting in smaller SDs, *except* at set size 1, where the SDs for all levels of *p* converge. This is due to the fact that the effective memory load (*G*) is still one for a single item, thereby precluding any benefits from multi-item grouping. At larger set sizes, on the other hand, a higher *p* reduces the average number of groups per memory array, thereby increasing the average number of slots devoted to each group. While increasing the number of available slots has an effect of *vertically* scaling down the memory SD (**Figure 10A**, right panel), increasing the grouping probability is akin to *horizontally* stretching the SD toward the right (**Figure 10B**, right panel), reflecting the compression of the effective set sizes. To contrast the predictions of two approaches, the curves in panels A and B are replotted in **Figure 10C**, by paring the parameters that lead to a matched apparent K at set size 8. It is evident that the deviations in the predicted SD become larger as the standard SA model (solid line) increases the number of slots (from top to bottom) to account for the enhanced capacity estimates.

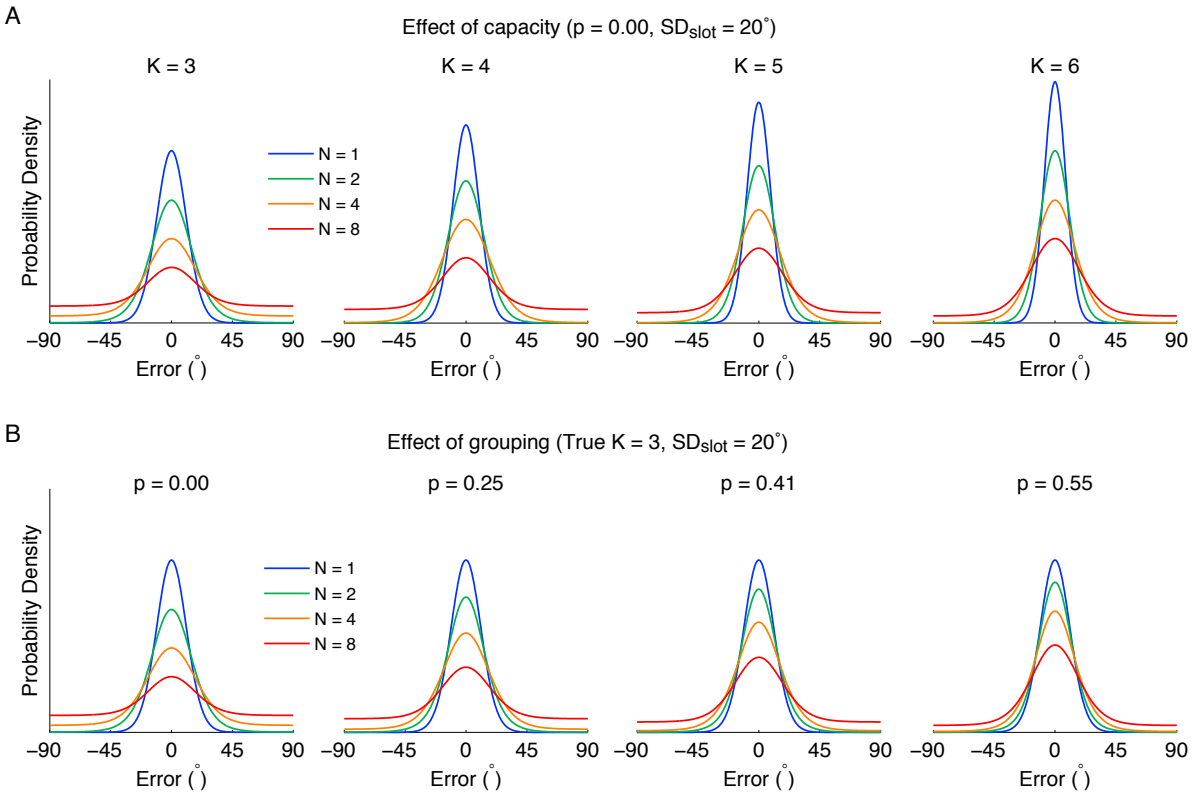

Figure 11. Visualization of the predicted response distribution. In each panel, the predicted response distributions for a subset of set sizes (1, 2, 4, and 8) are overlaid. The model parameters are identical to those shown in **Figure 10**. (A) Prediction from the standard SA model as the number of slots (K) varies from 3 to 6 (from left to right). A greater amount of memory resources (K) leads to improved VWM performance across all set sizes, as reflected in the proportional enhancement in memory precision for a single item (blue curves). (B) Prediction from the SA model *with grouping* as the grouping probability (p) varies from 0.00 to 0.55 (from left to right) while the number of slots is held constant (*True* $K = 3$). While increased grouping leads to fewer guess responses at large set sizes (orange or red curves), memory precision for a single item (blue curves) remains constant across different grouping probabilities.

To give a sense of what the response distributions would look like after applying these different parameters, the predicted response distributions are presented in **Figure 11** for a subset of set sizes (1, 2, 4, and 8). The plots on the leftmost column (identical for panels A and B) would correspond to the predictions for gratings, assuming three slots and no grouping ($p = 0$). The plots on the rightmost column would correspond to the predictions for lines, assuming either six slots and no grouping ($p = 0$) (panel A) or three slots with grouping ($p = 0.55$) (panel B). The capacity-based predictions (panel A) show

markedly enhanced memory precision for set size 1 (shown in blue) with $K = 6$, compared smaller Ks, in proportion to the increasing number of available slots (from left to right). By contrast, the grouping-based predictions (panel B) show a stable memory precision for set size 1 (shown in blue) across increasing levels of grouping probability ($p = 0.00 \sim 0.55$; from left to right) while demonstrating a pronounced reduction of guess responses with a higher p at set size 8 (shown in red). Thus, increasing the number of slots and increasing the grouping probability provide distinct predictions for VWM performance as a function of set size, especially in terms of memory precision.

An advantage for incorporating grouping into the standard SA model is that memory load will be no longer rigidly determined by the nominal set size. As illustrated in **Figure 12A**, flexible grouping strengths can accommodate possible changes in the effective memory load at a given set size, across different stimulus types. Moreover, the current implementation of grouping can account for changes in the storage efficiency within a given stimulus type as a function of set size. As shown in **Figure 12B**, the efficiency of data compression, measured by the ratio of the nominal set size (N) to the effective set size (G), has a baseline of 1 at set size 1, and gradually improves as set size increases. This implies that the more items there are in the memory array, the greater the potential benefit of grouping by the visual system. Whether this implementation of grouping indeed provides a better characterization of observers' VWM performance is an empirical question. In the next section, I will turn to the behavioral experiments designed to test the advantage of this modeling approach.

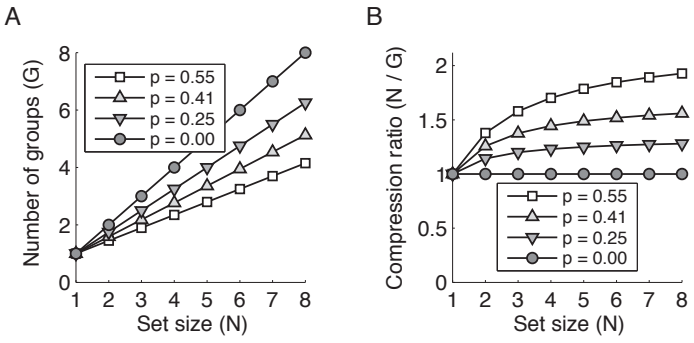

Figure 12. Compression of effective memory load via visual grouping. (A) The expected number of groups (G) as a function of the number of items (N) as the grouping probability (p) varies from 0.00 to 0.55. (B) Data compression ratio calculated as the ratio of the nominal set size (N) to the effective set size (G). A higher ratio indicates greater storage efficiency.

Experiments 1 & 2

The present experiments were conducted to examine how an individual's VWM performance for grating and line orientation varies as a function of set size, and to determine whether the SA model with grouping can provide a better account of the behavioral data than the standard SA model. VWM performance was tested in a delayed estimation task, by presenting sample arrays containing a variable number (1, 2, 4, or 8) of gratings or lines; these items were arranged in a virtual circle (Experiment 1) or on a jittered grid layout (Experiment 2). In these two Experiments, I sought to generalize the findings across different stimulus layouts, and introduced a visual masking procedure to control for the role of iconic memory (Experiment 2).

Two models (the standard SA model varying the number of slots across stimulus types and the alternative model that varies the probability of grouping while holding the number of slots constant) were fitted to the each participant's delayed estimation data. A statistical model comparison revealed a marked advantage for the model with grouping over the standard SA model in both experiments.

Methods

Participants

14 healthy volunteers with normal or corrected-to-normal visual acuity (6 male; ages 19-32 years) participated in Experiment 1, and another 14 volunteers (6 male; ages 19-30 years) participated in Experiment 2. Six participants (including the author) took part in both experiments. Participants (except the author) received course credit or monetary compensation ($12 per hour; four hours for Experiment 1, and two hours for Experiment 2) for participation. Written informed consent was obtained prior to participation. All aspects of this study were conducted according to procedures approved by the Institutional Review Board of Vanderbilt University.

Apparatus and stimuli

The experiments were programmed and controlled using MATLAB 8.3.0 (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) running on a Mac Pro. Participants viewed stimuli displayed on a gamma-corrected CRT monitor (1280

x 1024 resolution; a 85Hz refresh rate; 27.2 cd/m^2 mean luminance) in a darkened room, at a viewing distance of 46 cm. Head position was stabilized using a chin rest with a forehead bar.

The stimuli were generated using the same method described previously, using different spatial parameters and stimulus arrangements between Experiments 1 and 2. As in the previous experiments, Experiment 1 used a grating stimulus consisting of a sine-wave grating (2 cycles/˚, 50% contrast) presented within a circular Gaussian envelope (σ = 2˚, truncated at 2˚ diameter). In Experiment 2, the same sine-wave grating (2 cycles/˚, 50% contrast) was presented within a smaller Gaussian envelope (σ = 1.5˚, truncated at 1.5˚ diameter). In both experiments, the line stimulus was created by applying a rectangular aperture (0.25˚ x 2˚ in Experiment 1; 0.25˚ x 1.5˚ in Experiment 2) to the central bar of the grating stimulus, resulting in somewhat different aspect ratios between Experiments 1 and 2 (1:8 and 1:6, respectively). All stimuli were presented on a gray background (mean luminance 27.2 cd/m^2).

In Experiment 1, the gratings and lines (2˚ in diameter) were presented at randomized locations at 4˚ eccentricity with a minimum separation distance of 0.62˚ edge to edge (see **Figure 13A**). In Experiment 2, the stimuli (1.5˚ in diameter) were presented at randomized locations within a 5 x 5 square grid (15˚ x 15˚; central cell excluded) (see **Figure 13B**). Each item's position was jittered ±0.65˚ from the center of the cell (3˚ x 3˚), resulting in a minimum edge-to-edge separation of 0.20˚ between items. The maximum eccentricity of the center of an item was 9.40˚.

Stimulus orientation was randomly chosen on each trial for each item, from 180 evenly spaced orientations (0-180˚). Observers were asked to fixate centrally on a black bull's eye (0.5˚ in diameter) for the duration of each trial, throughout the experiment.

Experiment 2 used a patterned mask, consisting of low-pass filtered uniform noise (23˚ x 23˚) with a cut-off frequency of 4 cycles/˚ and a mean RMS contrast of 0.115 (SD = 0.005). The noise pattern was contained in a square (13˚ x 13˚), smoothed with a two-dimensional Gaussian kernel (σ = 12˚).

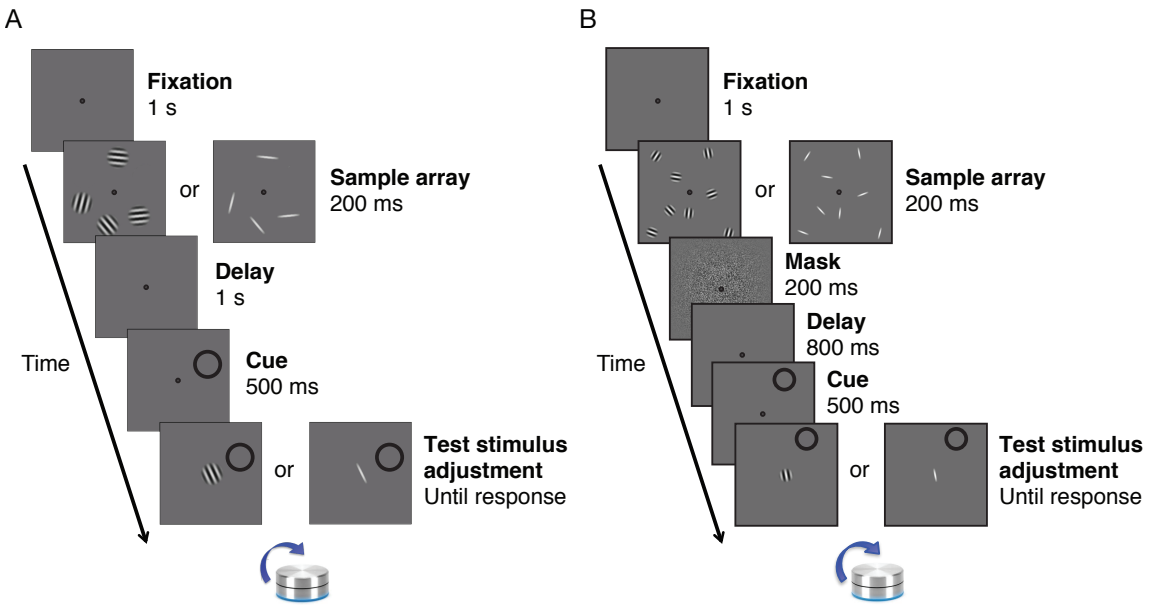

Figure 13. Trial sequences for the delayed estimation tasks in Experiments 1 and 2. In both Experiments 1 (A) and 2 (B), a sample array contained a various number of randomly tilted gratings or lines (1, 2, 4 or 8 items). For illustration, the grating and line arrays are shown with matched orientations and locations. In the actual experiment, these values were randomly determined for each stimulus type.

Procedure

A trial sequence for a delayed estimation in Experiment 1 is shown in **Figure 13A.** On each trial, a central fixation was presented for 1 s, followed by a sample array (200 ms) containing a various number (1, 2, 4, or 8) of gratings or lines, which varied randomly from trial to trial. After a 1-s delay, a spatial cue (a black outline circle; 2˚ in diameter) indicated the to-be-reported item for 500 ms, followed by a central test stimulus (a grating or a line) with a random initial orientation. Participants adjusted the orientation of the test stimulus by turning a computer knob. Participants were asked to report the remembered orientation as precisely as they could, and hit space bar to submit their response. On each trial, participants received auditory feedback based on accuracy, as described previously. The average accuracy score (0-100) for a block of trials was displayed at the end of each experimental block.

The trial sequence for Experiment 2 was identical to Experiment 1, except for the insertion of a patterned mask (**Figure 13B)**. After the sample array (200 ms), a patterned mask was presented for 200 ms, followed by a blank interval of 800 ms, and a cue. The

sample-to-cue ISI was kept at 1 s as in Experiment 1.

In both experiments, all conditions (two stimulus types, four set sizes) were randomly intermixed across trials. In Experiment 1, each participant completed four 1-hr sessions, divided into 9 blocks of 48 trials. This resulted in a total of 1728 trials, and 432 trials for each condition. In Experiment 2, each participant completed two 1-hr sessions, divided into 9 blocks of 48 trials, resulting in a total of 864 trials, and 216 trials for each condition.

Data Analysis

As a means for *summarizing* the capacity and precision for each experimental condition, the mixture model (Zhang & Luck, 2008) was fitted separately to each set size and stimulus type condition for each subject. This was to convey the general pattern of results without assuming a fixed underlying number of slots for each subject that applies across all set sizes.

For the *standard SA* approach, separate SA models were fitted to each participant's grating data and their line data, utilizing all set sizes. A total of four parameters were estimated per participant: *Grating's K*, *Grating's* SD_{slot}, *Line's K* and *Line's* SD_{slot}. For *the SA model with grouping*, the grating and line data from each participant were fitted concurrently, estimating a total of five parameters per participant: *True K*, *Grating's* SD_{slot}, *Grating's p*, *Line's* SD_{slot}, and *Line's p*. All models were fitted based on maximum likelihood estimation, using the genetic algorithms (GA) in MATLAB's Global Optimization Toolbox to seek the global optimal solution.

Statistical model comparison was conducted using the Akaike information criterion (AIC) (Akaike, 1974). The AIC is calculated as:

$$\mathrm{AIC} = -2\ln(L) + 2k,$$

where L is the maximum likelihood value of the fitted model, and k is the number of free parameters in the model. Among candidate models, the model that yields the lowest AIC score is selected as the best model. The AIC takes into account both goodness of fit and parsimony of the model, by rewarding high log-likelihoods and penalizing for extra free parameters.

The AIC score was calculated for each model for each participant. The AIC for the

standard SA model was simply a sum of the AIC scores for the two stimulus types. To obtain a total AIC score for each model, the AIC scores were summed across all participants from each experiment.

Results

Participants' delayed estimation data are shown in **Figure 14**. As is evident in the response error histogram (**Figure 14A, C**), the orientation reports for both stimulus types became less precise as the number of items increased. Importantly, memory precision for a single grating (shown in blue) was nearly as good as that for a single line. The difference between the two stimulus types became larger as the set size increased, characterized by a pronounced increase in random guess responses for gratings at set size 8 (shown in red) and a relatively stable performance for lines across set sizes.

The mixture model was separately fitted to each condition (**Figure 14B, D**); this revealed a subtle but reliable precision advantage for single lines over single gratings in Experiment 1 (ΔSD of 1.29˚, $t(13) = 4.51$, $p = .001$, Cohen's $d = 1.20$), and Experiment 2 (ΔSD of 1.30˚, $t(13) = 2.47$, $p = .028$, Cohen's $d = 0.66$). In fact, a similar magnitude of precision advantage for lines was observed across all set sizes, as the SD increased at a similar rate for both stimulus types as a function of set size. This is inconsistent with the standard SA model's prediction (see **Figure 10A**), in which the magnitude of the difference between the smallest and largest SDs depends on the number of available slots that can be averaged ($SD_{min} = SD_{max}/\sqrt{K}$). To illustrate, the average SD estimate for a single grating (10.6˚ in Experiment 1) is largely consistent with the SA model's prediction obtained by dividing the SD estimate at set size 8 (i.e., a single slot's precision) by the square root of the average K estimate at set size 8 (i.e., number of slots) ($19.4˚/\sqrt{3.5} = 10.3˚$). On the other hand, the SD estimate for a single line (9.3˚ in Experiment 1) is substantially larger than that predicted by averaging of 6.6 slots with SD of 17.6˚ ($17.6˚/\sqrt{6.6} = 6.9˚$).

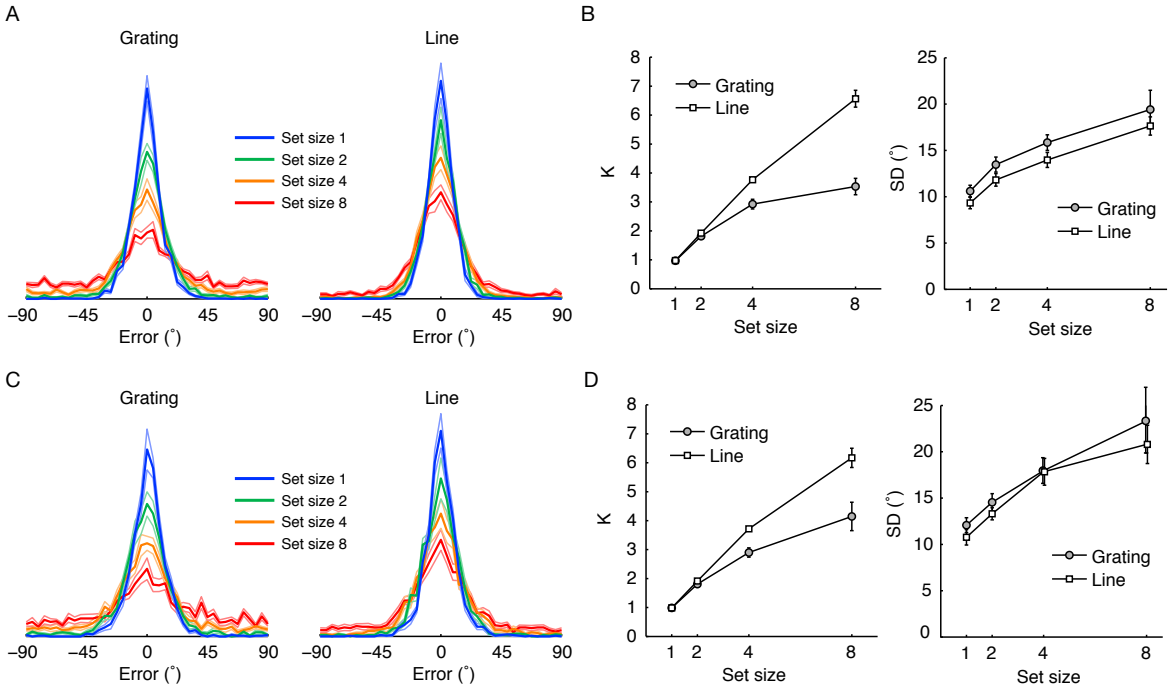

Figure 14. Delayed estimation results in Experiments 1 and 2. Top panels (A, B) show the results from Experiment 1 (the circular array), and bottom panels (C, D) show the results from Experiment 2 (the grid array). (A) and (C) show the response error histograms for gratings and lines, pooled across all 14 participants in each experiment. All histograms consist of 40 bins, 4.5˚ wide. Different colors indicate different set sizes (1, 2, 4, and 8), with the faint curves representing ±1 between-subject SEM of each bin's frequency. In (B) and (D), VWM performance in each condition is summarized in terms of the number (K) and precision (SD) of stored items. These K and SD estimates were obtained by fitting a separate mixture model to each set size and stimulus type condition for each participant. The error bars represent ±1 between-subject SEM.

This initial analysis provided a preliminary qualitative assessment of the SA model's prediction and its limitations, using the group-level, descriptive measures of VWM performance. Next, I will present the results from the formal model-based approach, in which each individual's VWM performance was characterized by two competing frameworks: The standard SA model and the new SA model with grouping.

Modeling results

When the *standard SA model* was fitted separately to each stimulus type, the difference between VWM for grating and line orientation was characterized as different numbers of memory slots. As shown in **Figure 15** (left two panels)**,** the mean *K* estimate

for lines was nearly twice that of gratings (*Grating's K* = 3.6, *Line's K* = 7.7 in Experiment 1; *Grating's K* = 3.4, *Line's K* = 7.1, in Experiment 2), while the precision estimate for an orientation stored in a single slot was not reliably different between the two stimulus types.

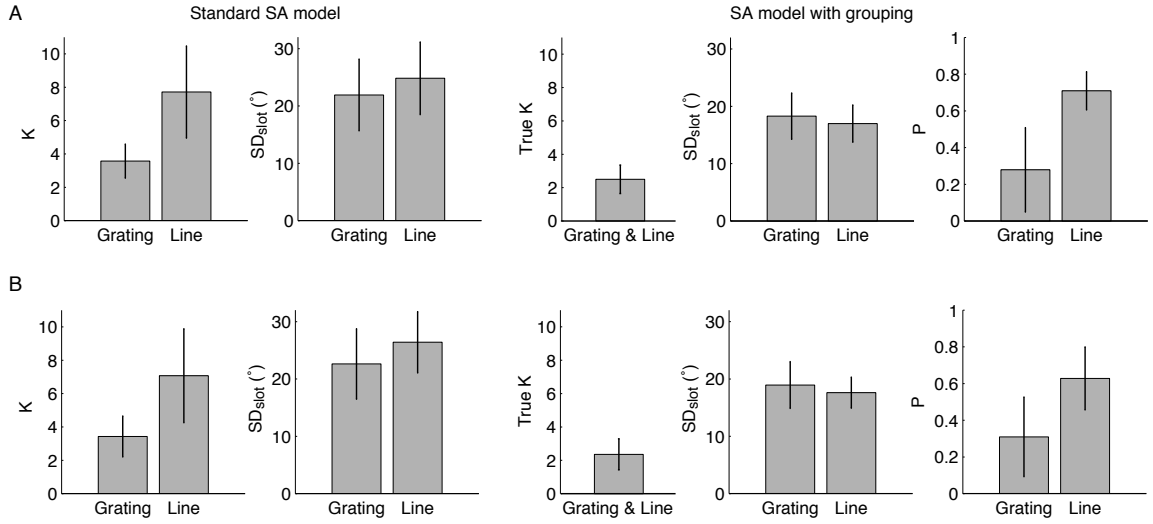

Figure 15. Modeling results in Experiments 1 and 2. (A) Results from Experiment 1. Left two panels show the mean parameter estimates for the standard SA model. Right three panels show the mean parameter estimates for the SA model with grouping. The error bars represent ±1 between-subject SEM. (B). Results from Experiment 2.

However, as discussed earlier, allowing a variable working memory capacity within an individual is not only theoretically unprincipled, but it can also lead to a mischaracterization of VWM if the storage efficiency (i.e., memory resource requirement) for a given stimulus were to change with set size. Since the averaging-based precision advantage at smaller set sizes is dictated by the upper limit of stored items at the largest set size, the parameters that maximize the likelihood of the data across *all* set sizes would systematically mischaracterize VWM performance, especially at the smallest and largest set sizes. This is demonstrated in **Figure 16**, which shows the response error histograms for two representative participants from Experiment 1, along with the predicted mixture distributions based on the best-fitting parameters of each model. While the standard SA model (shown in red curves) provides a reasonably good fit of the grating data across all set sizes, the model's prediction tends to deviate from the line data at two extreme set sizes, predicting a better precision at set size 1 and a worse precision at set size 8, compared to the actual data.

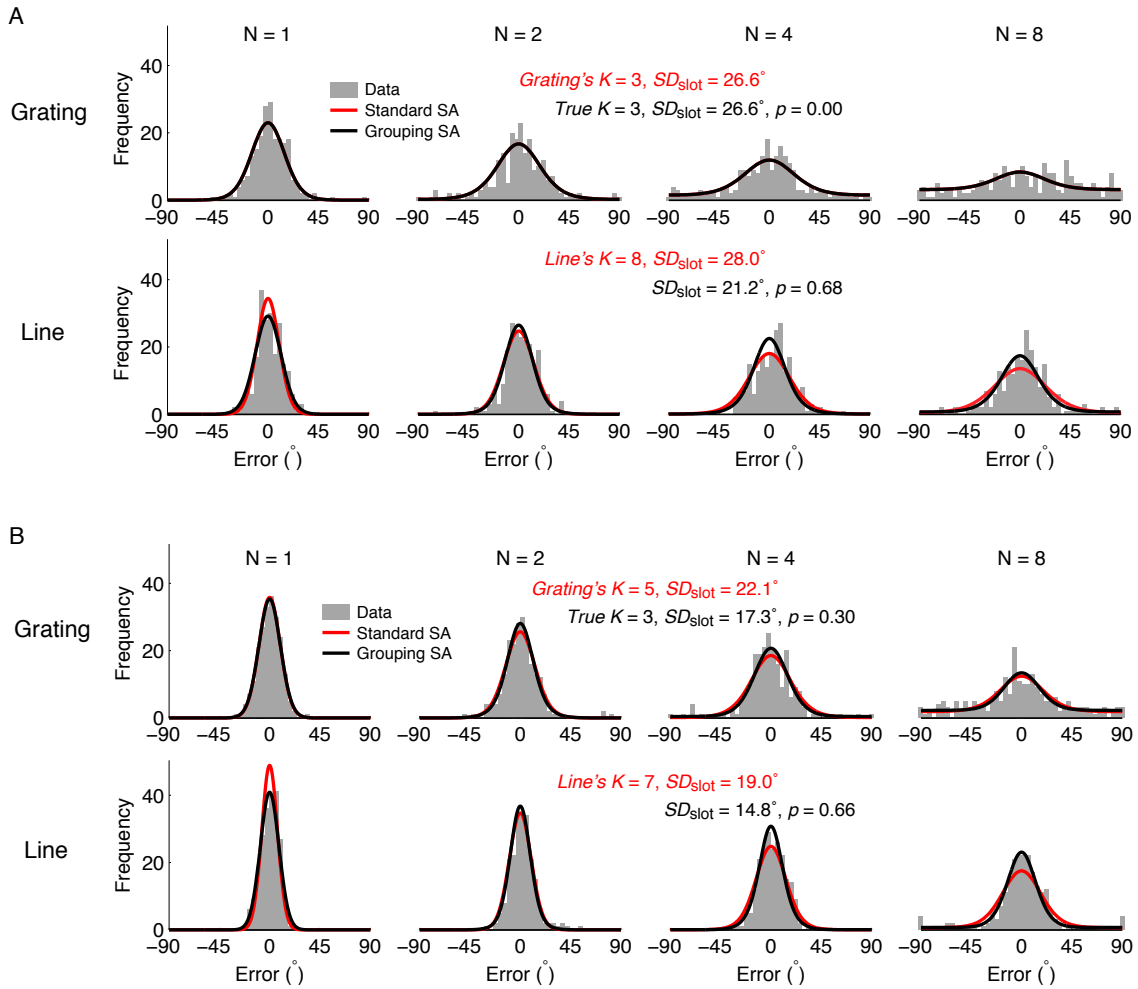

Figure 16. Example delayed estimation data and best-fitting model parameters from Experiment 1. Panels (A) and (B) show the response error histograms for two representative participants from Experiment 1, for each stimulus type and set size condition. All histograms consist of 40 bins, 4.5˚ wide. On each histogram, the mixture distribution based on the best-fitting parameters from the standard SA model (red curves) and the SA model with grouping (black curves) are overlaid.

When the grouping probability and single slot's precision were estimated separately for gratings and lines while holding the number of slots constant within an individual (*the SA model with grouping*; **Figure 15**, right panels), the estimated numbers of slots tended to be lower than those from the standard approach, with the mean *True K* of 2.5 and 2.4 in Experiments 1 and 2, respectively. The precision estimate for lines stored in a single slot was modestly better than that for gratings, $t(13) = 2.53$, $p = .025$, Cohen's $d = 0.68$ (Experiment 1); $t(13) = 1.96$, $p = .071$, Cohen's $d = 0.52$ (Experiment 2).

Importantly, the probability of grouping (p) was much higher for lines than for gratings (0.71 vs. 0.28 in Experiment 1; 0.63 vs. 0.31 in Experiment 2). While the estimated p was reliably above zero for both gratings and lines, ts(13) > 4.54, ps < .001, the difference in grouping probabilities between the two stimulus types was highly reliable across participants, $t(13) = 9.34$, $p < 0.001$, Cohen's $d = 2.50$ (Experiment 1); $t(13) = 8.18$, $p < 0.001$, Cohen's $d = 1.92$ (Experiment 2). Moreover, introducing the grouping factor improved the model fit for the line data by a much greater magnitude than it did for the grating data, as indicated by the log-likelihood increase of +160 for gratings vs. +291 for lines in Experiment 1, and +36 vs. +84 in Experiment 2. These results suggest that grouping played a bigger role in VWM for lines than that for gratings. This can be also gleaned from the example histograms presented in **Figure 16**, which demonstrates that the SA model with grouping (shown in black curves) provides a much better fit of these participants' line data than the standard SA model (red curves).

The formal model comparison based on the AIC statistic confirmed that the model with grouping provided a better account of the data (i.e., a lower AIC score) than the standard model in both experiments, after taking into account its increased complexity: $AIC_1 = 46950$ vs. $AIC_2 = 46076$ in Experiment 1 (AIC_1 is the standard SA model); $AIC_1 = 25522$ vs. $AIC_2 = 25308$ in Experiment 2. AIC difference scores (ΔAIC) were calculated by subtracting the grouping SA model's AIC score from the standard model's AIC score, such that a positive difference indicates how much worse the standard model performed in comparison to the grouping model. ΔAIC scores for individual participants' and the total difference score in each experiment are presented in **Table 1**, along with the number of participants showing positive ΔAICs (shown in parenthesis). In each experiment, the model with grouping outperformed the standard model in 12 out of 14 participants.

The relative fit of the standard model (AIC_1) compared to the grouping model (AIC_2) can be measured by the *relative likelihood* ($e^{(AIC_2 - AIC_1)/2}$). These measures indicated that the standard SA model was less than 10^{-189} times as probable as the grouping SA model in Experiment 1, and less than 10^{-46} times as probable in Experiment 2. Thus, the SA model's ability to account for the differences in VWM performance across set sizes and stimulus types was appreciably improved by allowing different grouping factors for the different stimulus types, while holding the number of slots constant within an individual.

Table 1. ΔAIC scores for individual participants

	Participant														Sum
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Expt. 1	147	62	71	17	30	165	24	62	62	44	-29	0	205	14	+875 (12)
Expt. 2	16	35	33	18	7	7	19	-9	33	5	45	6	-6	3	+213 (12)

Note. AIC = Akaike information criterion. Difference scores are calculated as: AIC(standard SA) – AIC(grouping SA)

Discussion

In the present study, I examined the systematic changes in VWM performance for grating and line orientation as a function of set size in order to better understand the resource allocation processes underlying the capacity advantage for line stimuli. Delayed estimation data showed that memory precision was very comparable for a single grating and a single line orientation. VWM performance for gratings sharply declined as the set size increased from 1 to 8, characterized by a marked increase in random guess responses. On the other hand, VWM performance for lines declined relatively smoothly as the set size increased, resulting in much fewer guess responses at the largest set size, with nearly twice as high of a maximum capacity as gratings. The observed relationship between the number and precision of stored items for line stimuli deviated substantially from the predictions of the standard *slots-plus-averaging* (SA) model, which assumed a greater number of memory slots for lines than for gratings. The alternative approach in which the number of slots was held constant while the probability of grouping (i.e., data compression ratio) was allowed to vary between stimulus types (the *SA model with grouping*) provided a much better account of VWM performance observed across different set sizes and stimulus types. This alternative model assumed that multiple items can be grouped into a coherent memory unit, reducing the *effective* memory load, and that the *apparent* capacity can be greatly increased depending on the degree to which the memory array is compressed into a fewer number of groups. The modeling results indicated that individual working memory capacity was limited to about 2.5 slots on average, but that the probability of grouping was nearly twice as high for lines as for gratings, leading to much higher estimates of apparent capacity for lines.

The present model of VWM with grouping provides a compelling account of the observed changes in the apparent information capacity of VWM across set sizes and stimulus types. In implementing visual grouping into the model, I made a few simplifying assumptions about how items are grouped together and how VWM resources are allocated among them. While these assumptions serve as a good starting point for characterizing the grouping process, they could be modified to more accurately capture the grouping mechanisms implemented by the brain. For example, I only considered grouping among $N - 1$ pairs of items comprising N items, instead of considering all N adjacent pairs along a circle or all possible pairwise connections within the array. This was to simplify the calculation of the number of groups based on the set size and grouping probability. The absolute value of the estimated probability will be valid only under certain assumptions, and thus should not be taken literally. It is also possible that grouping of items is not completely independent across pairs of items. If there are some constraints on the number of items that can be effectively grouped together, the number of groups will depend on the complex interactions among the potential grouping schemes among the items.

In addition, the assumption of *lossless compression* might be an oversimplification. Grouping of multiple items into a coherent memory unit might incur some loss of information about individual items. For example, a set of collinear lines could be grouped into one smooth path, but the estimates of individual line orientations could be biased toward the smooth path (unless they happened to be exactly collinear), resulting in an increased response variability across items and trials. Such grouping processes would still be adaptive if only a small number of slots are available compared to the large array size, and if the benefit of saving a memory slot (i.e., reducing random guesses) outweighs a modest loss of memory precision.

Among the competing models of VWM in the literature, I chose the discrete-slots framework for the theoretical reasons discussed earlier. While the idea that memory load can be reduced by visual grouping can be equally well applied to the continuous-resources framework, the resource models tend to be too unconstrained to be useful for explaining the perceptual and cognitive mechanisms behind VWM performance.

A recent version of the resource model, called the *variable precision* (VP) model (van den Berg et al., 2012), relies on several free parameters that afford considerable

flexibility in fitting the data but at the cost of introducing ambiguity in the interpretation of these parameters. The VP model assumes that random fluctuations in memory resources lead to a variable precision varies across items and trials around the mean precision ($\bar{J}$), which is determined by the number of items on the memory array (*N*). This idea is implemented by the J_1 parameter for the maximum precision at set size 1, and two additional parameters that freely vary depending on the data: The power parameter α that determines how steeply the mean precision should decline with increasing set size ($\bar{J} = J_1 N^{-\alpha}$), and the scale parameter τ that interacts with the shape of the gamma distribution from which the precision values are drawn. It is unclear how these parameters can be linked with important psychological constructs, such as fluid intelligence, the scope of attention, perceptual expertise and visual precision. Moreover, the unconstrained α parameter means that the total information capacity can vary across set sizes, with $\alpha < 1$ indicating an increase in the total information with increasing set size, and $\alpha > 1$ indicating a decline in the total information with increasing set size. No theoretical justification for such a variable information capacity has been provided. Importantly, the flexible α can mimic the effect of visual grouping, complicating the fitting and interpretation of the model parameters. For the current goal of understanding how visual grouping can increase the amount of information that can be stored in a limited memory space, the SA model was thus considered more suitable than the VP model.

The present work highlights the importance of considering the perceptual factors that allow for a more efficient usage of the limited capacity of working memory. By isolating one's core cognitive capacity from grouping-related capacity enhancements, the fluctuating estimates of VWM capacity can be reconciled with the large body of literature that links individual differences in working memory capacity with the scope of attentional focus (Cowan, 2001, 2005; Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010; Gold et al., 2006) or executive attention (Burgess, Gray, Conway, & Braver, 2011; Bettencourt, Michalka, & Somers, 2011; Cusack, Lehmann, Veldsman, & Mitchell, 2009; McNab & Klingberg, 2008; Kane, Bleckley, Conway, & Engle, 2001; Vogel, McCollough, & Machizawa, 2005). Thus, the present work is an important step toward a fuller understanding of the constraints on information capacity of the mind.

Chapter 4

The role of perceptual organization in visual working memory for orientation

Introduction

The results from the previous experiments indicated that the *total* amount of information encoded from a line array, taking into account both the number and quality of stored orientations, continually increased as the set size increased. In contrast, the amount of information that could be encoded from viewing arrays of gratings remained relatively stable across the larger set sizes. These results suggested that the more line items there were in the memory array, the more efficiently the array could be compressed into a smaller number of units for storage, thereby allowing far more line orientations to be stored than would be expected based on standard capacity estimates of VWM. This implies that the total information capacity of VWM is *not* a constant, but instead, varies as a function of the stimulus format and the number of items in the memory array.

So far, I have used the term *visual grouping* to describe the process that enhances storage efficiency in VWM (i.e., the amount of stored information per unit resource, whether discrete or continuous). This mechanism appears to be responsible for the dramatic capacity advantage for line stimuli over grating stimuli. The enhanced storage efficiency appears to arise from a formation of higher-order patterns or visual groups, given that the advantage for line stimuli was markedly reduced by sequential presentation of the stimuli, and that the amount of information stored for a single line orientation was negligibly different from that for a single grating orientation. However, visual grouping has thus far been discussed in a rather abstract manner as a data compression process, without regard to the specific visual mechanisms that might facilitate such processes. If this data compression is mediated by specific grouping mechanisms that operate on sets of oriented elements on an array, then storage efficiency should be directly linked to the configural properties of these elements, which can be systematically examined.

The Gestalt principles of perceptual organization (see Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012, for recent reviews) provide important clues as to what stimulus properties might contribute to the efficient storage of orientation information in VWM. Early Gestalt psychologists introduced a set of principles that determine the

perceived grouping of image elements; these include the laws of *proximity*, *similarity*, *good continuation*, *symmetry*, and *closure* (Koffka, 1935; Wertheimer, 1923/1938). These grouping principles have been expanded upon and further tested by modern vision scientists, equipped with rigorous psychophysics and computational modeling. Given that these grouping processes determine the units of our perceptual experiences, in this work I propose that the same principles will likely determine the units of visual representations actively held in working memory.

Role of perceptual organization in VWM

The most prominent effect of perceptual grouping in VWM can be inferred at the item-specific level, in terms of the presumed object-based nature of VWM. Specifically, perceptual grouping can enhance the overall information capacity of VWM, as demonstrated by the fact that people can maintain more features in VWM if they belong to a common object than if they belong to separate objects (Luck & Vogel, 1997; see also Delvenne, & Bruyer, 2004; Xu, 2002). In a study by Xu (2006), the object-based storage benefit depended on both the *proximity* and the strength of *connectedness* between the two features (color and orientation) comprising parts of an object, suggesting that these grouping cues are crucial determinants of the units of storage in VWM. Similarly, grouping by *common regions* has been shown to allow more shapes to be stored in VWM (Xu & Chun, 2007). The behavioral benefit for the grouped shapes over the ungrouped shapes was accompanied by reduced fMRI responses in inferior intraparietal sulcus (IPS), reflecting a reduced number of discrete objects, and increased fMRI responses in superior IPS, reflecting increased information storage (Xu & Chun, 2007). Thus, perceptual grouping of multiple components into a common object allows discrete image elements to be organized into a smaller number of coherent perceptual units, substantially increasing the amount of information that can be simultaneously held in VWM.

Although VWM performance has been characterized as relying on the assignment of discrete slots to individual items, several studies have suggested that perceptual grouping *among* the displayed items can influence performance in VWM tasks to some extent. Previous studies have systematically manipulated various grouping cues in the sample arrays to examine its impact on change detection performance. Woodman, Vecera,

and Luck (2003) showed that perceptual organization could bias the priority with which items are encoded into VWM; for example, colored squares that were grouped together by *proximity* or *connectedness* (Palmer & Rock, 1994) tended to be encoded together. VWM encoding can be also biased in favor of items grouped by *similarity,* as shown by Peterson & Berryhill's (2013) study in which change detection accuracy was improved for a pair of same-colored items, compared to the rest of the items with non-repeated colors (see Lin & Luck, 2009, for a related finding of subtle improvement in display-wide performance for similar colors over dissimilar colors; cf. also Brady & Tenenbaum, 2013, for a contextual sensitivity in color change detection).

In these previous studies, the grouped and ungrouped conditions were purposefully designed, such that the grouped displays contained greater regularities than would be expected by chance. Such manipulations could have affected the visual and cognitive strategies used by the subject to perform the task. In contrast, for the studies in this dissertation, the grating and line arrays used were randomly generated on each trial and no regular structure was introduced in an artificial manner. Nonetheless, the degree of perceived randomness (Hsu, Griffiths, & Schreiber, 2010) could still vary across these arrays. Previous studies of color working memory provide evidence that observers can extract and utilize higher-order structures even from randomly arranged colored items. Using a change detection task, Brady and Tenenbaum's (2013) found that a change that altered the global texture of the display (i.e., the color similarity of neighboring items) was more easily detected than a change that did not, suggesting that observers were not just storing individual items as independent units but also storing higher-order summary information of the entire display. In a very different study, Brady & Alvarez (2015a) showed substantial display-specific variability in delayed estimation performance for displays containing three colors, which could be reasonably well captured by a model that integrated the item-level representations (individual colors) and cluster-level representations (i.e., the mean and variance of sets of colors in the display).

The representation of gist-like summary statistics postulated by Brady and colleagues can be linked to some extent with the idea of perceptual grouping because both involve the process of assigning display elements into distinct clusters or sets. However, gist-based memory can be conceptually distinguished from the specific local processes

required for perceptual grouping, since no spatially specific information between pairs of items needed to be processed in their model. While the clusters comprising the gist representation are primarily defined by the *similarity* of items, perceptual groups can be defined by a more diverse set of image properties, such as those mentioned above. Moreover, perceptual grouping allows information about individual items to be encoded within a higher-order perceptual unit, without necessarily reducing the item-specific information to an abstract summary statistic, such as the mean or the variance. Despite these conceptual differences, both of these proposed notions of gist representation and perceptual grouping provide a means to represent visual information in working memory in a more efficient or compressed manner, and can be used to predict the systematic effects of display properties on VWM performance.

Compared to color, VWM for orientation might benefit from the consideration of even richer spatial structures that can emerge from an array of oriented elements (see **Figure 17**).

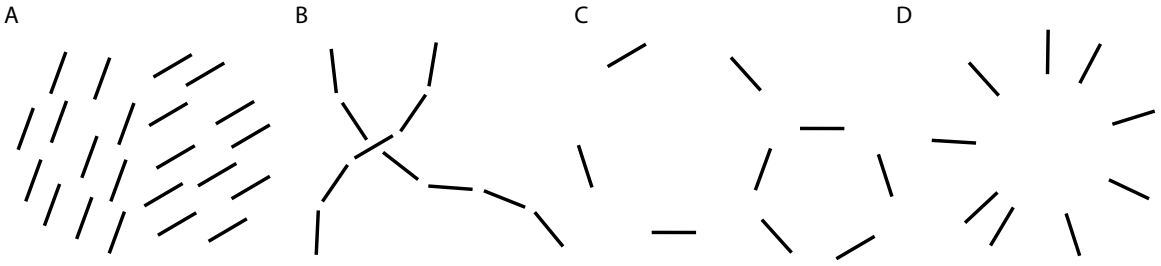

Figure 17. An illustration of perceptual grouping cues in oriented elements. (A) Similarity and proximity. (B) Good continuation. (C) Cocircularity. (D) A radial pattern.

Parallelism (or orientation *similarity*) and *proximity* could provide strong grouping cues, as a region of a display consisting of parallel gratings or lines can be economically represented as one oriented texture (**Figure 17A**). Even when the display items consist of heterogeneous orientations, they can still be grouped by *good continuation* (**Figure 17B**) if the orientation varies smoothly from one item to the next (i.e., *collinearity*) or if they are aligned along a common circle (i.e., *cocirciularity*, see **Figure 17C)**. Moreover, the entire display can be compressed into a *mirror-symmetric* polygon, such as a *radial pattern* (**Figure 17D**) or a *closed circle*, depending on the global configuration of the items.

In this study, I will evaluate the hypothesis that perceptual grouping is not only

limited to these rare situations where grouping cues are highly salient; instead, they may apply in a continuous manner to varying levels of grouping cues, modulating the probability of grouping among the items. If the visual system is adept at extracting regularities that emerge from these randomly oriented elements, and if these happen-chance patterns can be analyzed by grouping mechanisms to compress the memory array, then VWM performance should vary systematically depending on the spatial relations among the elements in any given visual display. Thus, orientation provides a unique and rich opportunity to test for the role of perceptual organization in VWM. This experiment addresses an important question, as no study to date has systematically examined these geometric relationships among the oriented elements and their influence on VWM performance.

Present study

The present study had two primary goals. The first was to determine whether the perceptual organization that occurs in randomly generated arrays of oriented items leads to systematic influences on VWM performance. The second goal was to determine what factors account for the superior VWM capacity that people show for lines over gratings. We hypothesized that certain grouping mechanisms might operate specifically on line stimuli or that the grouping strength would persist over larger spatial separations for line stimuli. However, an alternative possibility is that the magnitude or probability of grouping is much weaker for gratings than for line orientations, while the tuning width and spatial profile of these grouping mechanisms are shared for both stimuli.

To examine whether VWM performance varies systematically across displays, I conducted an online experiment in which the same set of randomly generated line and grating displays were shown to a large number of participants, using matched stimulus parameters between the two stimulus formats. I found substantial variability across displays in terms of how accurately the items were remembered, which was highly consistent across participants, as well as between the two stimulus formats. To gain insight into what spatial patterns observers might typically extract from a given display and hold in their VWM, I pooled the delayed estimation responses from all of the participants, and reconstructed the participants' memory representation for each item from each display.

This reconstruction hinted at the role of Gestalt grouping principles, as some of the most accurately remembered displays contained highly regular structures, consistent with the predicted effects of *parallelism*, *collinearity*, and *closure*. Moreover, the errors in the reported orientations were not random, but rather, were systematically biased toward a pattern that made the reconstructed display appear more coherent than the original display.

In order to quantify the strength of perceptual organization in these displays and test its relationship with VWM performance, I constructed a model of display-specific VWM performance, inspired by Gestalt grouping principles. The model considers a set of properties that can make an array of oriented stimuli more simple and unified, consisting of a set of perceptual grouping mechanisms that process information about the similarity, cocircularity and proximity of pairs of items, as well as the global radiality/circularity of the display. These perceptual grouping mechanisms are defined in a mathematically precise manner based on the stimulus variables (i.e., the positions and orientations of items in a display) and a small number of free parameters that are fitted to the participants' performance data. The overall pattern strength of a display is computed by a weighted sum of the strengths of the individual grouping mechanisms. The model estimates the parameters that maximize the correlation between the estimated pattern strength of every display and the accuracy with which the display is remembered.

Assuming this approach proves successful, my second goal was to determine whether the same perceptual grouping processes that enabled efficient storage of specific arrangements of oriented elements were also responsible for the overall capacity advantage for line stimuli over grating stimuli. It is conceivable that line stimuli can be stored more efficiently than grating stimuli because the patterns formed by oriented lines are more readily grouped by the visual system than those formed by oriented gratings. One way this could occur is if certain perceptual grouping cues operate more effectively on line stimuli than on grating stimuli, producing a larger modulation of VWM performance for lines. Then, the largest performance advantage for lines should be obtained in the displays with the strongest grouping cues, whereas the line displays that are weak in such grouping cues should produce a similar level performance as that for grating displays.

An alternative possibility, however, is that perceptual groups could be effectively formed in both grating and line displays, substantially modulating the storage efficiency in

each stimulus format, but the efficiency for storing line displays is enhanced over that for grating displays by a certain factor, across all levels of perceptual organization. For instance, perceptual grouping processes might provide the initial perceptual units for a given display that serve as blueprints for the organization of representations in VWM. The degree of facilitative interactions among these perceived units (i.e., gain modulation) might ultimately determine the representational units in VWM. It is possible that a set of local perceptual groups consisting of line items might be more readily apprehended as a whole than those consisting of grating items. This hypothesis predicts that VWM performance for both stimulus types should exhibit substantial variability across displays according to perceptual organization, and the performance for line displays should be enhanced relative to that for grating displays, to the same degree across all levels of perceptual organization.

Testing these competing hypotheses is crucial for understanding the role of perceptual organization in VWM for orientation, as well as how the organization processes might interact with the stimulus format. To achieve this goal, it is important to test VWM performance across a large number of randomly generated displays, producing a large range of natural variability in perceptual organization across displays. It is also crucial to formulate a model that can capture a substantial portion of the variability across displays by integrating multiple perceptual grouping principles. In the following sections, I will describe the online experiment that allowed me to collect a rich dataset using a large set of stimuli, followed by a description of the Gestalt-inspired model that I developed to explain the variability observed in these data.

Methods

Participants

600 participants were recruited on Amazon's Mechanical Turk (MTurk; https://www.mturk.com) for the line version of experiment, and another 600 participants were recruited for the grating version. All participants (ages 18 years or older) were from the United States, gave informed consent, and were paid 50 cents for approximately 20 minutes of their time. MTurk users are demographically diverse and more representative of the adult population in the United States than the college student population typically evaluated in experimental research (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang,

& Gosling, 2011). Previous studies have shown that VWM experiments run on MTurk replicate the results obtained in the laboratory (Brady & Alvarez, 2011, 2015a; Brady & Tenenbaum, 2013).

Stimuli and procedure

Line and grating stimuli were generated using the same methods as described in Experiment 1 of Chapter 3, except that a fixed set of orientations and locations were used for both stimulus formats. A sample array consisted of six gratings or lines arranged on a virtual circle (see **Figure 18A**). A total of 96 stimulus configurations were used. The stimulus orientations and locations were randomly determined for each configuration. These configurations were applied to each stimulus format, resulting in 96 grating displays and 96 line displays.

Figure 18B shows an example trial sequence of the delayed estimation task; this protocol was identical to that from Experiment 1 of Chapter 3, except the following changes in the experimental design.

The stimulus type was manipulated *between* participants. Each participant ran in either the grating version or the line version of the experiment, receiving all 96 displays of the same stimulus type. Each display was presented only once, in a randomized order. For each participant, the probed item in each display was determined pseudo-randomly, such that the six items were evenly distributed across participants, ensuring an equal number of trials (i.e., 100 participants) per item. This pseudo-randomization procedure was repeated for all 96 displays to ensure that there was no correlation between any two subjects in terms of which items were probed across different displays. This counterbalancing approach prevented any systematic biases that might result from repetitions of specific combinations of conditions across participants.

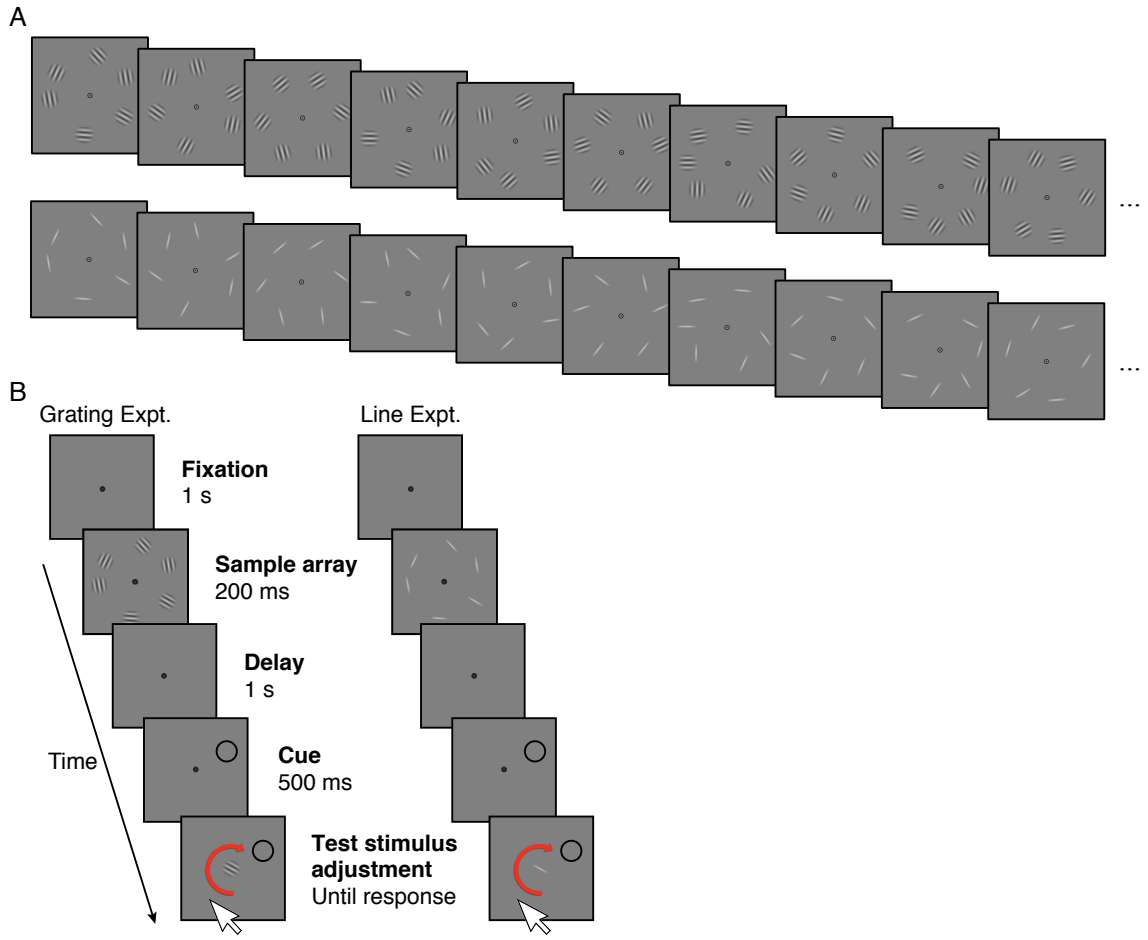

Figure 18. Example stimuli and trial sequence for the delayed estimation task used in the online experiment. (A) 96 grating displays and 96 line displays were generated using the same set of random arrangements of orientations. (B) A trial sequence for the two versions of delayed estimation task, which used gratings and lines, respectively.

Participants were asked to report the orientation of the cued item from memory by moving a mouse cursor in a rotational motion to modify the orientation of the test stimulus. As in all of my previous experiments, the initial orientation of the test stimulus was randomly determined on each trial, and thus it was independent of the orientation of the probed stimulus.

In addition to the 96 pre-generated displays, participants received 10 catch trials, in which only a single randomly oriented item was presented at a randomized location, for a total of 106 trials. Participants showing extreme error magnitudes (>2 s.d. from the mean) in either the catch trials or the experimental trials were replaced with newly recruited

participants (7.1% in the line experiment, and 11.5% in the grating experiment) to ensure data quality and to achieve the target number of trials per display and item.

For the main experimental conditions, a total of 57,600 trials were obtained from 600 participants (96 displays x 6 items x 100 trials) for each stimulus type. This large dataset allowed for reliable estimation of display-specific performance (600 trials/display) as well as item-specific performance (100 trials/display) for each stimulus type.

Measures of VWM performance

I applied two different sets of summary statistics to the delayed estimation data, to hone in on different aspects of the data.

The mixture model (Zhang & Luck, 2008) is useful for summarizing the overall performance level of each participant (across all of the displays) in terms of memory capacity (K) and precision (SD), as was done in my earlier experiments. This approach will be used for comparing the overall VWM capacity and precision between gratings and lines.

The mixture model approach can also be used to estimate the mean number of stored items in a particular display, by pooling the data from all 600 participants and considering them as obtained from a single observer. However, as I will demonstrate later, there was a modest degree of trade-off between the number of items stored in VWM and quality of information stored about each item, making display-specific K estimates an unreliable indicator of how well a display may be remembered (see *Behavioral results*). For this reason, the raw error magnitudes will be used in this study as a primary measure of *display-specific* VWM performance. The mean squared error (MSE) is inversely proportional to Fisher information, which provides an index of stored information per item. Since orientation is a circular variable, I will use the circular analog of the root mean squared error (RMSE), calculated based on the response error vector x as follows:

$$\text{circular RMSE}(x) = \sqrt{-2\ln\left(\frac{1}{n}\sum_{i=1}^{n}\cos(x_i)\right)}.$$

The RMSE is an aggregate measure of response error magnitude that reflects the variability as well as any bias in the responses. When the orientation estimation responses are unbiased, the RMSE is equivalent to the standard deviation of the responses, and the MSE

is equivalent to the variance.

In addition to producing display-specific effects in VWM performance, the visual system's sensitivity to perceptual grouping cues in a display might lead to systematic biases and variability in memory for individual items comprising the perceptual groups. To examine such *item-specific* effects, I used two additional summary statistics: The mean and the SD of responses. The circular mean of the error vector x is calculated as follows (Fisher, 1995):

$$\text{circular mean}(x) = \text{atan}\left(\sum_{i=1}^{n} \sin(x_i), \sum_{i=1}^{n} \cos(x_i)\right).$$

Any deviation of the mean from zero would indicate a systematic bias in the reported orientation relative to the true orientation.

The circular SD is calculated as follows:

$$\text{circular SD}(x) = \sqrt{-2\ln(\bar{R})},$$

where $\overline{R}$ denotes the mean resultant length, calculated as:

$$\bar{R} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} \cos(x_i)\right)^2 + \left(\frac{1}{n}\sum_{i=1}^{n} \sin(x_i)\right)^2}.$$

The SD of responses is similar to the RMSE, but it measures the response variability with respect to the mean reported orientation, not the true orientation. The measures of SD and RMSE may not agree if there is a large bias in the responses to an item. For example, an item's overall response accuracy could be very low (i.e., large RMSE) if the responses are tightly spread (i.e., low SD) around a wrong stimulus value.

Behavioral results

To assess individual participants' overall VWM performance across all of the displays, the mixture model was fitted to each participant's data, separately for the catch trials (set size 1) and experimental trials (set size 6). The group-averaged capacity and precision estimates are presented in **Figure 19**. The catch-trial results indicated that the participants included in the final dataset (N = 1,200) performed reliably on the delayed estimation task, precisely reporting the single item's orientation in nearly all of the catch

trials. The average guess rates were only 0.06 (SEM = 0.005) for gratings, and 0.04 (SEM = 0.004) for lines. Memory precision for a single item was comparable to that found in the lab, with the average SD of 12.36˚ (SEM = 0.22) for a single grating orientation, and SD of 10.61˚ (SEM = 0.18˚) for a single line orientation. Importantly, the online experiment replicated the pronounced capacity advantage for line stimuli over grating stimuli found in the lab experiments. At set size 6, participants could remember an average of 2.3 grating orientations (SEM = 0.07), and 4.4 line orientations (SEM = 0.06). The precision of the stored orientation (SD) was modestly better for gratings (20.30˚; SEM = 0.57˚) than for lines (22.7˚; SEM = 0.39˚). Despite the small disadvantage in memory precision, overall VWM performance for line orientation was considerably better than that for grating orientation, allowing me to examine the role of perceptual organization as a potential mechanism of VWM advantage for lines.

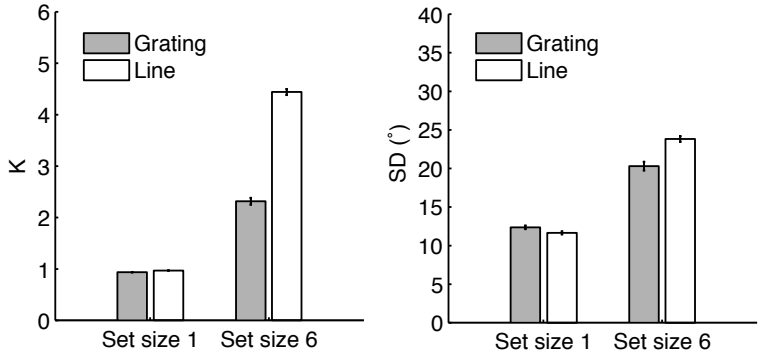

Figure 19. The mixture model results in the online experiment. The mixture model was fitted to each participant's data (N = 600 for each stimulus type), using all of the trials from each set size condition (i.e., regardless of the displays). The error bars represent ±1 between-subject SEM.

When the delayed estimation responses were pooled across all 600 participants for each display and each stimulus type, I found substantial variability in RMSE scores across 96 displays, which ranged from 38.2˚ to 55.3˚ (mean = 48.7˚, s.d. = 3.4˚) for grating displays, and from 24.0˚ to 40.0˚ (mean = 33.4˚, s.d. = 2.9˚) for line displays. The observed distribution of display RMSEs is shown on the leftmost column of **Figure 20** (red curve). To compare the observed variability with the variability one would expect from random variations in performance for individual items arising from the noise within the data, I

randomly resampled 96 sets of 600 trials (with replacement) separately for each stimulus type (57,600 trials), blind to the display labels. The RMSE was calculated for each set of 600 trials, treating them as if they came from the same display. After this resampling procedure was repeated 2,000 times, the expected variability (s.d.) of the 96 RMSE scores was only 2.1˚ for grating data, and 1.2˚ for line data (shown as shaded gray in **Figure 20**). The results indicated that observed magnitudes of variability across displays (3.4˚ for gratings, 2.1˚ for lines) were unlikely to have occurred if all displays were assumed to be equal and interchangeable (*p*s < 0.001).

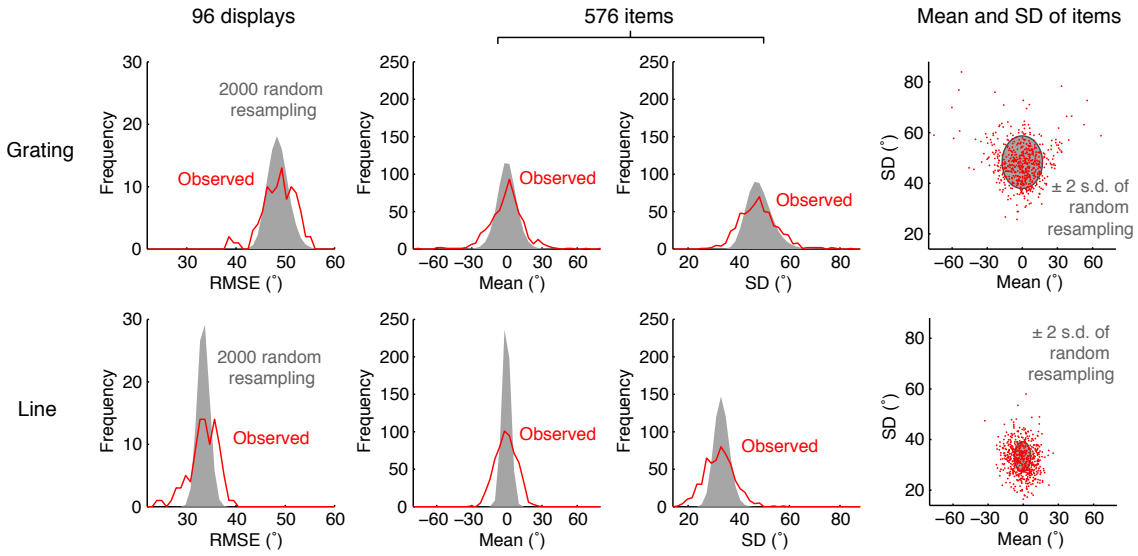

Figure 20. Variability of VWM performance across displays and items. The results from the grating experiment (top) and the line experiment (bottom) are shown separately. The leftmost column shows the distribution of RMSE values for 96 displays. The middle two columns show the distributions of the mean and the SD values of 576 items, which are displayed as a 2-dimensional scatter plot on the rightmost column. On each panel, the distribution of actual VWM performance is shown in red, and that based on the randomly resampled data is shown in gray.

An examination of VWM performance for individual items (96 displays x 6 items = 576 items) revealed substantial biases (i.e., shifts of the mean away from zero), as well as varying levels of response variability (SD) across items (see **Figure 20**, middle two columns). The same resampling procedure was performed at the level of item, randomly sampling 576 sets of 100 trials from each stimulus type and calculating the mean and SD of each set of 100 trials. Again, the observed variability across items was significantly larger

than expected by noise alone (ps < 0.001). The scatter plot on the rightmost column of **Figure 20** illustrates the deviations of individual items' observed mean and SD from those expected based on all of the data.

To test whether these large display-specific and item-specific effects on VWM performance are reliable across participants, I performed split-half correlation analyses on these data. For each stimulus type, the participants were randomly split into two sub-groups while ensuring that there were an equal number of participants assigned to each item (i.e., 50 participants) from each display (i.e., 300 participants) for the two groups. Pearson correlation was calculated between the two halves of the data, which was repeated for 2000 random splits. As shown on the top two rows of **Figure 21**, the effects of display and items on VWM performance were highly reliable across participants, as indicated by significant mean split-half correlations across 2000 random splits. The display and item effects tended to be less reliable for grating stimuli (r = 0.47 for display RMSEs; r = 0.30 and r = 0.35 for item means and SDs, respectively; all ps < .001) compared to line stimuli (r = 0.75 for display RMSEs; r = 0.72 and r = 0.60 for item means and SDs, respectively; all ps < .001). Nevertheless, it is noteworthy that systematic variations in performance were observed across displays and items for both types of stimuli.

If this variability in VWM performance across displays arose from a common perceptual grouping mechanism operating on the configuration of the display elements, there should be some correlation between grating and line performance. Indeed, I found a significant correlation in performance across the two stimulus types (r = 0.62 for display RMSEs; r = 0.33 and r = 0.51 for item means and SDs, respectively; all ps < .001), which is shown on the bottom row of **Figure 21**. While the RMSE varied greatly across both grating and line displays, the pronounced VWM advantage for line stimuli was preserved across all of the displays (i.e., all display RMSEs lie above the diagonal), with no stimulus configuration producing worse performance for lines than for gratings.

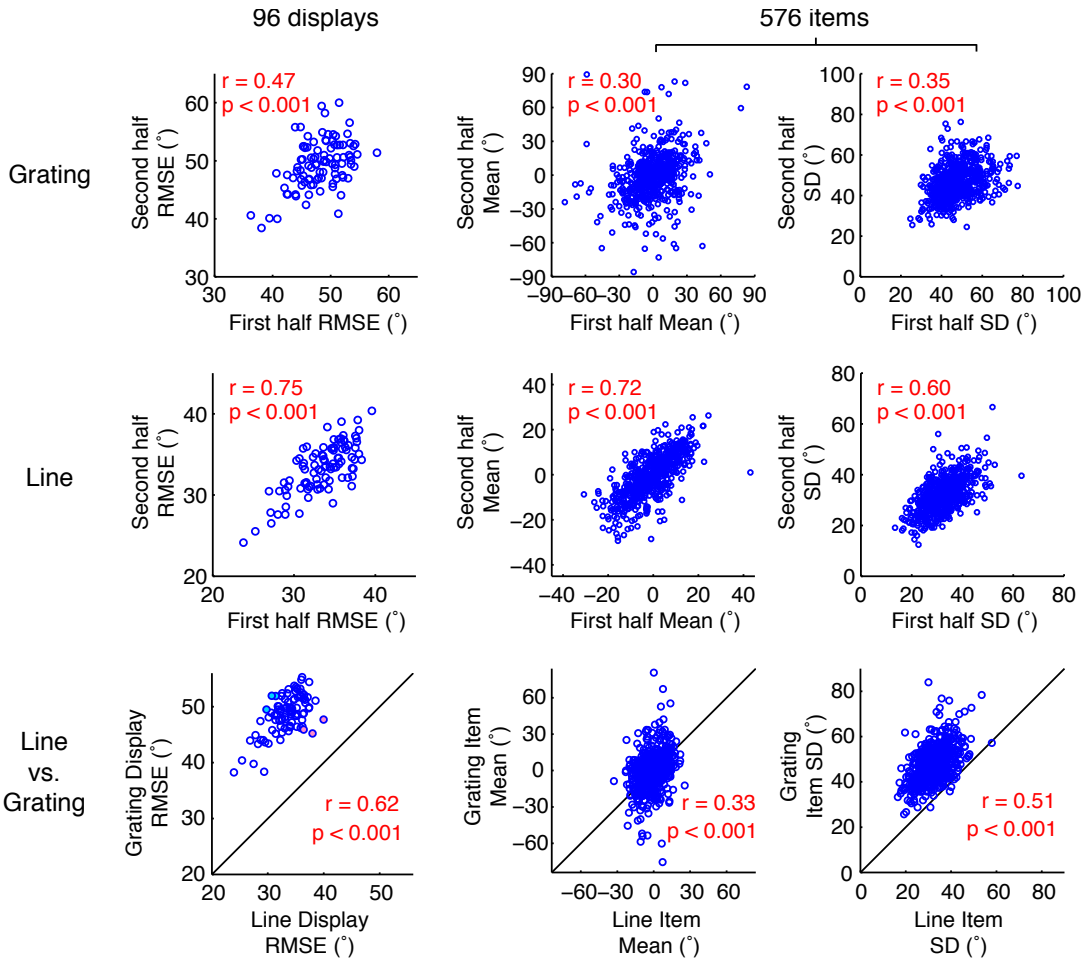

Figure 21. Reliability of display and item effects across participants and stimulus types. The random split-half correlations for the grating and line data are shown on the top two rows. On each panel, the mean Pearson correlation coefficient (*r*) across 2000 random splits is indicated in red. The scatter plot shows the results from one representative random split of the data. The correlations between grating and line performance are shown on the bottom row. On the bottom left corner, the displays with the largest discrepancy between line and grating performance are indicated with circles filled with pink (easy for gratings, hard for lines; IDs 46, 60, and 80) and light blue (easy for lines, hard for gratings; IDs 21, 26, and 29). These displays are presented in **Figure 24**.

This observed relationship between grating and line performance is inconsistent with the hypothesis that the enhanced VWM capacity for lines might arise from the perceptual grouping processes that operate specifically on line stimuli. This hypothesis predicts that RMSEs for line displays should vary substantially while RMSEs for grating displays should remain relatively stable, at a level comparable to the poorest performance (i.e., the largest RMSE) obtained with the line displays. On the contrary, VWM

performance appears to be similarly sensitive to the configural properties in grating and line displays.

The display-specific effects of VWM performance can be also characterized in terms of the number of items stored for each display (K) and their precision (SD), by fitting the mixture model to all of the data collected for each display. Interestingly, I found evidence of a systematic trade-off between the estimated capacity and precision across displays. As is evident on the top panels of **Figure 22**, higher K estimates were associated with larger SD estimates (i.e., poorer memory precision for successfully remembered items) on many of the displays. As a result, some of the displays with a high K estimate actually showed a large overall error magnitude (RMSE) (shown as red circles), and some of the displays with a low K showed a small RMSE (shown as blue circles). Despite this trade-off between the capacity and precision, there was a general advantage to storing more items, as the displays with higher K estimates tended to show lower RMSEs (see **Figure 22**, bottom panels). This relationship did not hold in a subset of displays, for which the increased K was not associated with any improvement in the overall performance, as reflected in the horizontal stripes of similar-colored circles (i.e., RMSEs of similar rank-orders). Such flat performance across displays would be expected if the increased capacity was achieved by simply averaging the orientations of multiple items on a display, leading to a direct trade-off between the number of items averaged together and the accuracy of these items.

The overall benefit of storing more items in a display is also well demonstrated by the large difference between grating and line stimuli in terms of their K estimates and RMSE scores. As reported above, average memory precision was subtly impaired for lines (SD = 22.7˚) compared to gratings (SD = 20.30˚), but the benefit of increased capacity greatly outweighed this precision cost, leading to a much smaller overall error for lines. As shown in the bottom right corner of **Figure 22**, the RMSE of a display is reasonably well predicted by the number of stored items in the display, across different displays and stimulus types. In the subsequent modeling analyses, I will assume that enhanced perceptual grouping would lead to an overall performance benefit, achieved by an optimal balance of the number and precision of stored items. Therefore, the RMSE will be used as a primary measure of display-specific VWM performance.

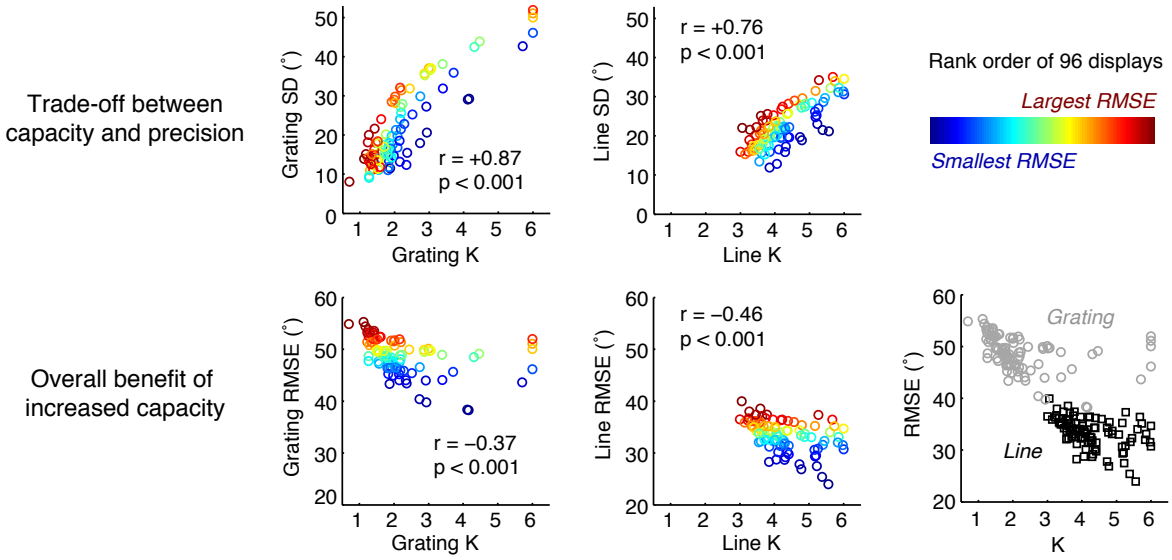

Figure 22. The relationships among different measures of VWM performance for specific displays. Estimates of the number (K) and precision (SD) of stored items, and the overall error magnitude (RMSE) are shown, for each stimulus type. On the top panels, the SD estimates for specific displays are plotted against the capacity estimates, based on the mixture model fitted to each display. Data points are color-coded based on the rank-ordered display RMSEs within each stimulus type (top right). On the bottom panels, the RMSE scores for specific displays are plotted against the same K estimates. The bottom right panel shows the data from the grating displays (gray circles) and line displays (black squares) on the same plot.

Together, these results demonstrate that specific arrangements of oriented elements led to substantial variability in VWM performance across randomly generated displays, and these effects were highly reliable across participants and between stimulus formats.

Reconstruction of remembered displays

Based on the display RMSE scores, I explored what kinds of spatial regularities might be present in the displays that were remembered most accurately. Moreover, given that the different stimulus configurations resulted in systematic and reliable biases in the reported orientations for individual items, I wanted to examine whether such biases were shaped by the configural properties of the display. The schematics of the stimulus configurations that led to highly accurate VWM performance are presented in **Figure 23**, along with a polar-plot histogram of reported orientations for each item in each display. These displays were selected based on the RMSE scores for *line* stimuli, for which a

majority of the items were successfully remembered, with an average guess rate of 26.6%. Compared to the grating displays, which elicited high rates of guessing responses (61.4% on average), reconstructions of delayed estimation reports for line displays were quite accurate, so that any systematic biases or patterns could be easily discerned.

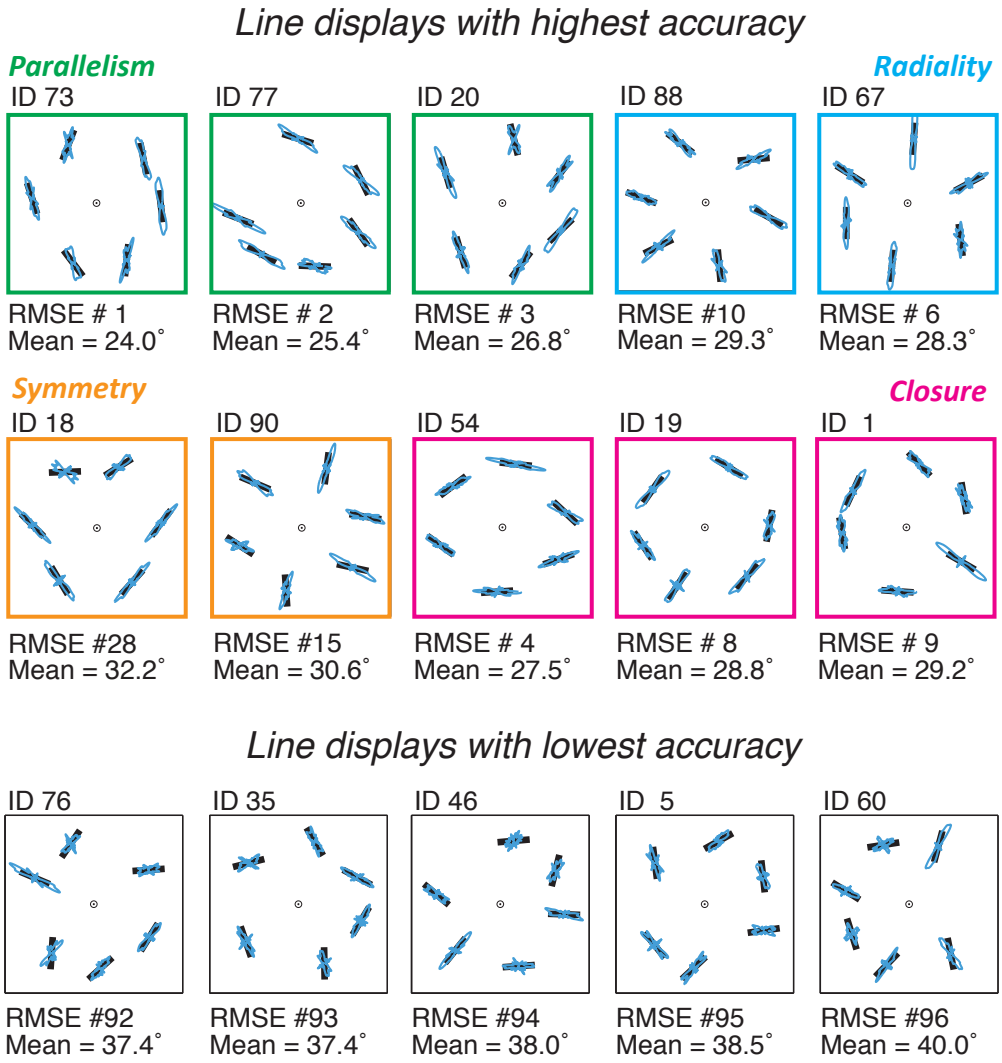

Figure 23. Reconstruction of remembered displays based on reported orientations. The examples of stimulus configuration that led to the most accurate (top) and most inaccurate (bottom) VWM performance for line stimuli. The actual stimulus orientation and locations are represented as black tilted bars. The blue curve overlaid on each stimulus represent the polar histogram of reported orientation for that item. The ID of each configuration is an arbitrary number assigned to 96 different configurations. The number next to # indicates the rank order of each display in terms of overall accuracy (#1 = smallest RMSE; #96 = largest RMSE).

As shown in **Figure 23**, the displays that were remembered most accurately tended to have highly regular structures, such as parallelism, symmetry, radial organization, and closure. Moreover, the reported orientations for individual items in these displays showed systematic biases in preference for configurations that exaggerated these regularities in the actual stimuli. For example, the reported orientations for items in display #73 and #77 were biased toward orientations that were parallel to the neighboring items. The oddball items in displays with a high degree of closure revealed a bias to report an orientation that completed a closed circle formed by the rest of the items, leading to extreme errors (i.e., reported orientations were orthogonal to the stimulus orientation) on some of the trials. Despite these item-specific biases, storing the regular structures from these displays seems to have been adaptive, as indicated by very low display-level RMSE values, compared to those from the most irregular displays (bottom row).

While there was generally good agreement between grating and line performance in term of which displays were easier to remember, some displays revealed highly discrepant performance across the two stimulus types. The discrepancy score for each stimulus configuration was calculated by subtracting the z-scored RMSE of the line display from that of the grating display, such that a positive score indicated a relative ease of remembering the line version of the configuration, and a negative score indicated a relative ease of remembering the grating version. **Figure 24** shows example configurations that produced the top three positive scores (left column) and top three negative scores (right column). The RMSEs for these six displays relative to all other configurations can be also seen on the scatter plot in **Figure 21** (bottom left corner).

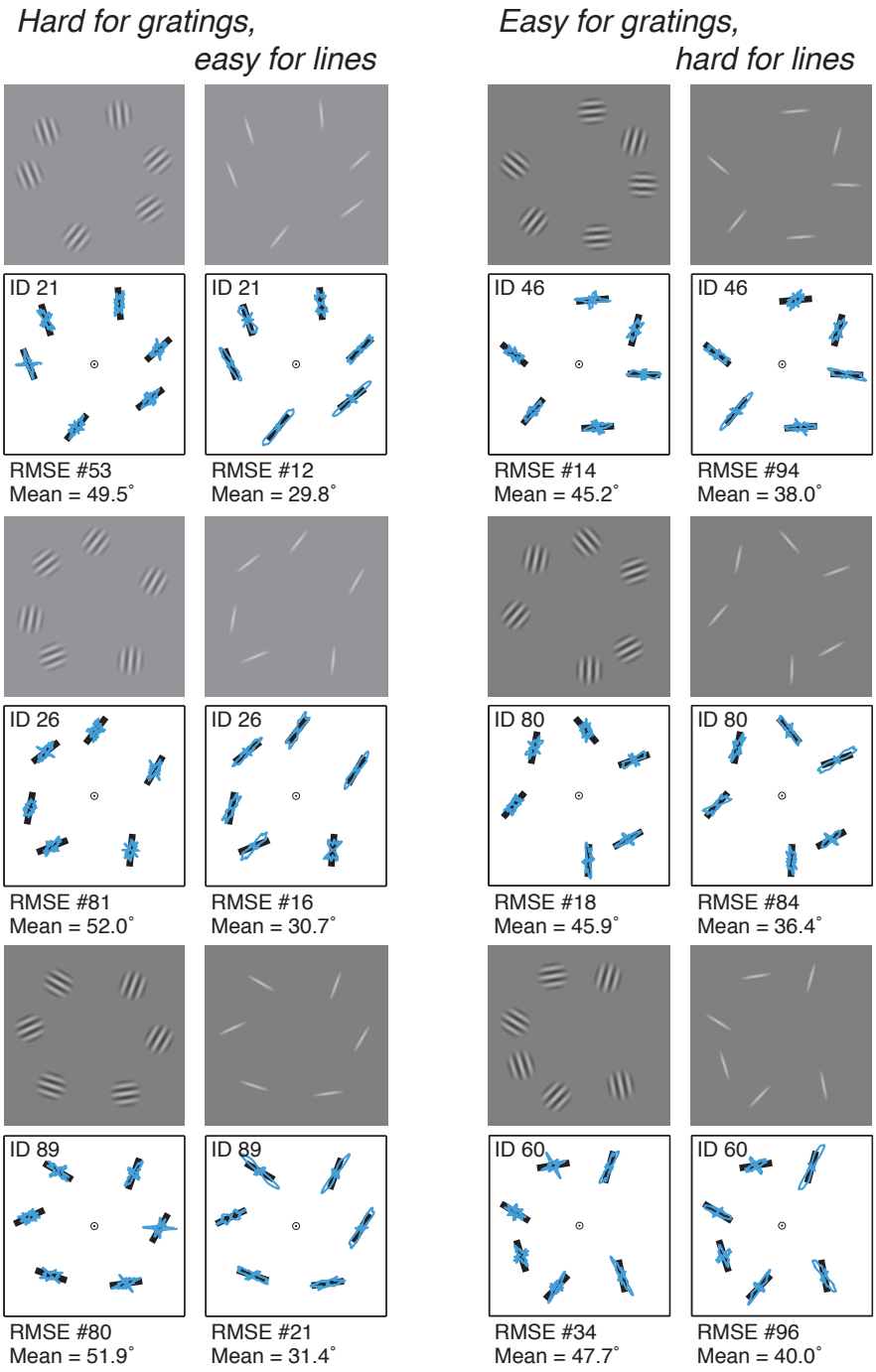

Figure 24. Examples of displays showing discrepant VWM performance between line and grating stimuli.

Visual inspection of the reported orientations for displays on the left column of **Figure 24** (i.e., easy for lines) suggests that the parallel structure in these line displays led to highly accurate responses, but presentation of the same arrangement of oriented

elements in the grating displays often led to extreme biases, aligned with the angle formed between the center and the item's location (i.e., the radial axis) or an angle tangent to the radial axis (i.e., aligned with the virtual circle). These extreme errors might reflect a guessing strategy, which happened to be suboptimal for these particular displays. On the other hand, the pattern of reported orientations for the displays on the right column (i.e., easy for gratings) appear very similar between gratings and lines. It is unclear what properties led to the relative advantage for grating stimuli or disadvantage for line stimuli.

While it would be important to determine what spatial properties might constitute uniquely strong grouping cues for gratings but not for lines, or vice versa, the broad correspondence of the display effects between the two stimulus types suggests a prevailing influence of common perceptual grouping mechanisms on VWM performance for grating and line stimuli. Thus, the present study will focus on the similarities in performance across the stimulus types rather than their differences.

Modeling the mechanisms of perceptual grouping

In order to determine how much of the variability in VWM performance across displays can be explained by Gestalt-like organization, I constructed a model that quantified the sensitivity of human observers to various spatial properties in the display, including similarity, cocircularity, proximity, and globally radial or circular organization (see **Figure 25**). The central idea of the model is that, in any given display, oriented elements exhibiting these properties can be perceptually grouped, allowing for more efficient representation of their orientations in VWM. Assuming that visual system can combine various configural properties detected in a display, the model combines the strengths of individual properties to quantify the overall pattern strength of a display; the model is then fitted and used to predict VWM performance across all displays.

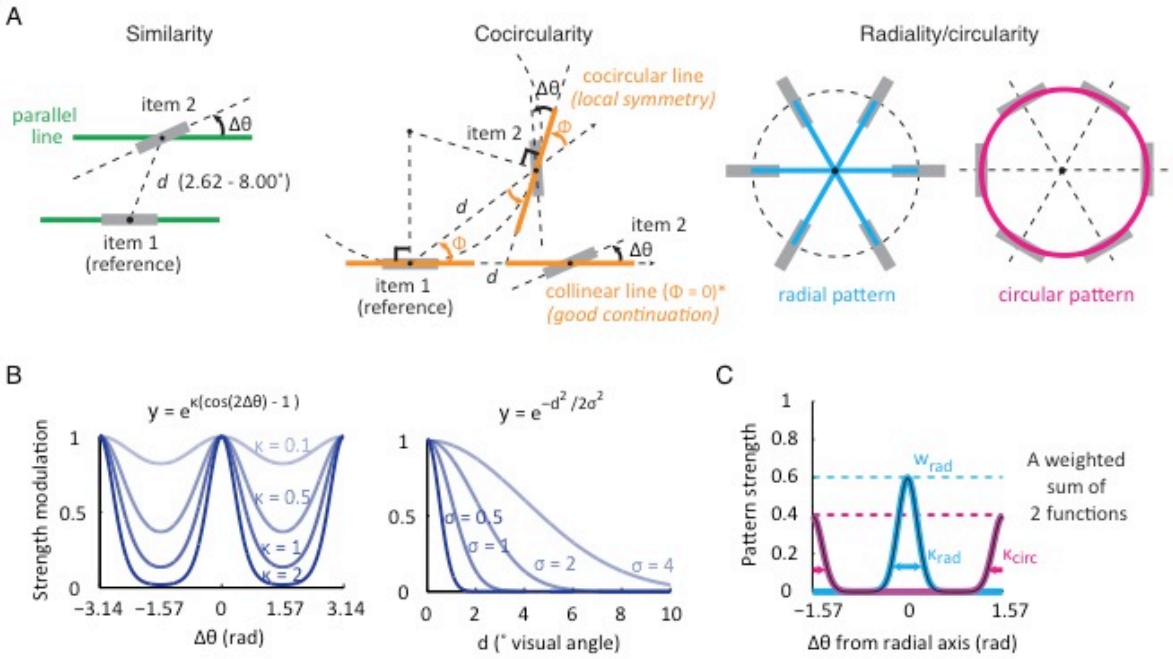

Figure 25. Components of the Gestalt-inspired model of display-specific VWM performance. (A) Illustration of the geometric relationships representing similarity, cocircularity, and radiality/circularity. Similarity and cocircularity are defined between pairs of items (items 1 and 2; gray bars). Radiality/circularity is defined as a global, higher-order pattern that is built upon a highly specific configuration of all items, which varies between the two extremes (radial vs. circular patterns). (B) Orientation sensitivity (left) and distance sensitivity (right) functions used for calculation of the pairwise similarity and cocircularity strengths. The strength of similarity (or cocircularity) is determined jointly by the angular deviation (Δθ) of a second item from the orientation parallel to (or cocircular with) the first item, and the distance (*d*) between the two items. In the model, the orientation sensitivity parameter (κ) and the distance sensitivity parameter (σ) are estimated based on the data. (C) The radiality/circularity strength varies as a function of the average absolute deviation (Δθ) of all six items from the radial axis. The sensitivities for radial and circular patterns are determined by κ_{rad} and κ_{circ} parameters, respectively. The peak strength of a radial pattern relative to that of a circular pattern is determined by the weight parameter, w_{rad}.

Each of these properties reflects a clearly defined geometric relationship among the items, which can be directly calculated from the orientations and locations of the items in a given display. The strength of each property is considered as a continuous variable, which has a maximum value if the arrangement of the items coincides with the strongest possible pattern that can be found among them, and gradually declines as the stimulus arrangement

deviates from the perfect pattern. This is characterized by a set of *grouping functions*, which relate the stimulus parameters to the perceived strength of each property in a display. As the perceived pattern strength would depend on both the physical stimulus parameters and the sensitivity of the organizational processes of the visual system, the parameters for the grouping functions will be estimated based on the VWM performance data.

In the model, *similarity* is defined as the degree of parallelism between two orientations. As illustrated in **Figure 25A**, one item can be treated as the reference, and the other item's orientation can be considered relative to the first item (shown as green lines). Similarity between a pair of oriented items is strongest when they are oriented parallel to each other, and decreases with increasing angular difference ($\Delta\theta$) between the two items. The model also considers the spatial *proximity* between a pair of items, which is inversely related to Euclidean distance (d) between two items. The model assumes that grouping by similarity depends on proximity (i.e., similarity strength declines as a function of increasing distance between the items), based on the idea that the parallel structure within a display might be better detected when these similarly oriented items are nearer to each other. This assumption is supported by the environmental image statistics literature, which shows that nearby parallel lines are more likely to arise from the same coherent object than are distant parallel lines (Brunswick & Kamiya, 1953; Geisler, Perry, Super, Gallogly, 2001).

Cocircularity occurs when two oriented elements can form a tangent on a common circle, as indicated by the orange lines in **Figure 25A**. Cocircular structure is prevalent in natural images (Elder & Goldberg 2002; Geisler et al. 2001; Sigman, Cecchi, Gilbert, & Magnasco, 2001) and psychophysical evidence suggests its role in contour integration (Feldman, 1997; Field, Hayes, & Hess, 1993). For a given reference item with orientation θ_1, a cocircular orientation (θ_c) can be defined based on the angular direction (Φ) of a second item with respect to the reference orientation, by rotating θ_1 in the direction of the second item by 2Φ ($\theta_c = \theta_1 + 2\Phi$). At a given distance ($d$), two cocircular orientations are smooth (i.e., collinear) when the radius of the common circle is large, making *collinearity* as a special case of cocircularity (Parent & Zucker, 1989). Cocircularity is also linked to *local symmetry*, as any two edges tangent to the same circle are rotated with respect to a

line connecting the centers of the two edges, by the same angle (Φ) but with the opposite signs. As in the calculation of similarity strength, the model assumes that cocircularity strength between a pair of items is influenced by both the angular deviation (Δθ) of the second item from the perfectly cocircular orientation and the distance between the two items (*d*).

Similarity and cocircularity can be independently calculated for any given pair of oriented items, although they will converge in some special cases. For example, two perfectly collinear items are automatically similar (i.e., parallel), but not vice versa: If the second item is displaced in an orthogonal direction without changing its orientation (e.g., as the green parallel lines in **Figure 25A**), the path between the two items is no longer as smooth as when they were on the same horizontal plane. Moreover, even highly heterogeneous orientations can have a strong cocircular relationship, depending on the alignment of their orientations *and* position.

The model applies a set of smooth functions to convert the degree of orientation deviation and the distance between two items to strength scores (see **Figure 25B**). Orientation values are wrapped around a circular space, such that the absolute deviation between two orientations increases as the Δθ changes from 0˚ to 90˚, and then decreases as Δθ changes from 90˚ to 180˚. To capture the circular nature of orientation space, I adopted a von Mises distribution (without the normalizing constant) as the orientation sensitivity functions for similarity and cocircularity strengths, both of which are defined based on Δθ.

The orientation sensitivity function produces maximum similarity or cocircularity value of 1 when a pair of items are perfectly parallel or cocircular. The strength declines as the orientation deviation (Δθ) increases toward 90˚. The shape of this orientation-tuned function is determined by the orientation sensitivity parameter κ, which corresponds to the precision parameter of the von Mises distribution. κ is estimated separately for similarity and cocircularity, based on the VWM performance data. Likewise, the distance sensitivity function (a half-Gaussian distribution[1] without the normalizing constant) leads to

[1] Previous studies have characterized the relationship between the stimulus separation and the strength of grouping by proximity using Gaussian (Elder & Zucker, 1994), exponential (Kubovy & Wagemans, 1995), and power law (Elder & Goldberg, 2002; Oyama, 1961) models. In the present study, these different functions produced comparable predictions of

decreasing grouping strength as the distance (*d*) between a pair of items increases. The steepness of the decline is determined by the distance sensitivity parameter σ, which corresponds to the standard deviation of the Gaussian distribution.

Mathematically, the final strength of similarity or cocircularity between a pair of items is defined as the product of the orientation- and distance-based strength scores, which is calculated for *all* possible pairs of items within each display[2]. The *display-level* similarity and cocircularity strengths are obtained by averaging these pairwise strengths. The pairwise calculation of similarity and cocircularity is based on the concept of a "local association field" (Field et al., 1993), which refers to the region around an image element in which local oriented elements are grouped together according to a specific set of rules. These rules, or local grouping functions, are instantiated by the geometric relationships and functions presented in **Figure 25**. The structure of the association field can also be represented as a two-dimensional map, as shown in **Figure 26**. Spatial properties of the psychophysically defined association field can be linked to how visual information is integrated by neurons in the visual cortex via feedback and lateral connections, giving rise to receptive field structures representing complex spatial relationships (Hess & Field, 1999). Since any coherent spatial relationships in a display would simultaneously activate the neurons sensitive to these structures, I assumed that all pairwise relationships among the items could contribute to the perceived pattern strength of the display, thereby influencing how accurately the display can be remembered.

VWM performance, with the half-Gaussian distribution providing a modestly better overall fit.

[2] A display containing six items results in a total of 15 pairwise relationships. Unlike

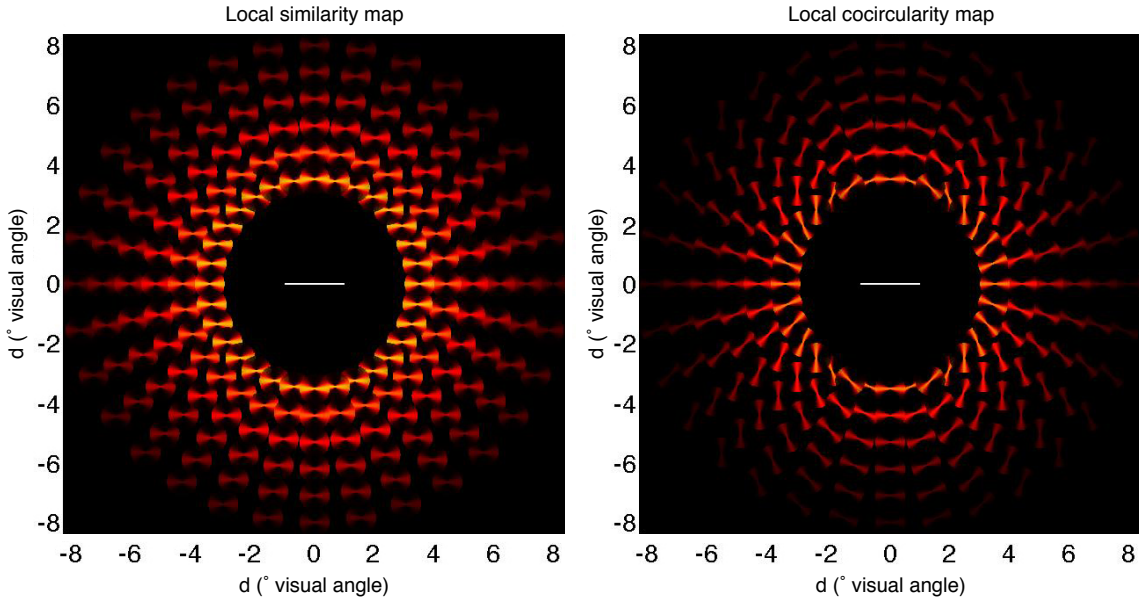

Figure 26. A graphical representation of local similarity and cocircularity maps. Each map was created based on the best-fitting parameters obtained for the line stimuli (κ_{sim} = 1.87, σ_{sim} = 3.80; κ_{cocirc} = 4.70, σ_{cocirc} = 3.30). The white horizontal bar at the center represents a reference item. The bowtie-shaped elements in the surround represent the strength of grouping by similarity or cocircularity at the corresponding item locations, which is modulated by the orientation deviation of the second item, as well as its distance from the reference item. The perfectly similar (i.e., parallel) or cocircular orientation for the second item is indicated by a thin yellow line. The progressive weakening of similarity or cocircularity as a function of orientation deviation and distance is indicated by the transition from yellow to red to black.

The last component of the model is *radiality/circularity.* Unlike the first two parameterized grouping mechanisms, which are based on local relationships among pairs of items, radiality/circularity considers the configuration of *all items* in a display simultaneously. A radial orientation can be defined at any position in a display relative to the central fixation point, and is specified by the angular position of an item arranged on the virtual circle. When all items in a display are radially oriented, they comprise a globally radial pattern (shown as cyan spokes in **Figure 25C**). In contrast, when all items are orthogonal to the radial axis (i.e., tangent to the virtual circle), they comprise a globally circular pattern (shown as a magenta circle). As the items deviate from the radial axis, radiality decreases and circularity increases. Thus, radiality and circularity are two opposing patterns.

The model assumes that any arrangement of randomly oriented items with their centers positioned on a virtual circle lies on a continuum of radiality and circularity, with the strongest patterns occurring at the two extremes and the weakest patterns occurring at intermediate levels of *radiality/circularity*. This is because a display with an even mix of radial and tangential orientations would not be expected to elicit a strong sense of pattern coherence. The function specifying this relationship is depicted in **Figure 25C**, which is characterized as a mixture of two von Mises functions centered at zero (i.e., radial orientation) and 90˚ (i.e., tangential orientation) with different concentration parameters (κ_{rad} and κ_{circ}). The weight parameter w_{rad} determines the peak strength of a radial pattern relative to that of a circular pattern. The input for this function is the average of the absolute deviations ($\Delta\theta$) of all six items from the radial axis, such that the mean absolute deviations of 0˚ and 90˚ indicate the strongest radiality and circularity[3], respectively, and the mean deviations around 45˚ (e.g., three 0˚s and three 90˚s) indicates the weakest pattern strength.

Radiality/circularity also partly overlaps with the other two components, but it requires very specific combinations of local similarity or circularities. For instance, a globally circular pattern consists entirely of cocircular items. Also, a globally radial pattern can arise from a set of collinear item pairs, each forming a spoke of the radial axis. On the other hand, the strength of global radiality/circularity is generally opposed to that of similarity, as the various stimulus locations along the virtual circle dictate that the corresponding radial/circular orientations be highly heterogeneous. Separately modeling these global properties implies that VWM might receive an extra advantage from such a higher-order configuration, beyond the sum of benefits from local properties.

Across the three components of the model, there were seven free parameters to determine the shape of the grouping functions. These include two sets of orientation and distance sensitivity parameters, for similarity strength (κ_{sim}, σ_{sim}) and cocircularity strength (κ_{cocirc}, σ_{cocirc}), two orientation sensitivity parameters for radiality and circularity strengths

[3] It is extremely rare to obtain these perfect radial or circular patterns just by coincidence. In fact, the averaged absolute deviations from the radial axis for the 96 randomly sampled displays in the present study ranged between 20.0˚ and 76.6˚ (mean = 45.3˚; s.d. = 11.1˚). Thus, highly radiality or circular patterns occurred only on a small subset of the displays.

(κ_{rad}, κ_{circ}), and the relative weight between radiality and circularity strengths (w_{rad}). The *overall pattern strength* of a display was calculated by a weighted sum of the outcomes of the three components. The weights among the three components were specified by two additional free parameters, w_1 and w_2: Similarity = w_1; cocircularity = $(1 - w_1) \cdot w_2$, and radiality/circularity = $1 - (1 - w_1) \cdot w_2$. Thus, a total of nine parameters were estimated based on the data.

A set of parameter values that maximized (negative) correlation between the display's pattern strength and RMSE score were searched, using the genetic algorithms (GA) in MATLAB's Global Optimization Toolbox. The model was fitted separately to the grating and line data.

Modeling results

The model combining the similarity, cocircularity, and radiality/circularity strengths could reliably predict VWM performance for specific displays to a considerable degree, for both grating and line stimuli. Each component of the model was found to be a significant predictor of VWM performance.

The estimated grouping functions for individual components are presented in **Figure 27A** (left two columns). Overall, the results revealed very similar parameter estimates for gratings and lines, far more similar than I had initially anticipated. For the *similarity* component, the best-fitting orientation sensitivity (κ_{sim}) and distance sensitivity (σ_{sim}) parameters were highly consistent across the two stimulus types, as indicated by the overlapping curves for grating (gray) and line (black) stimuli. The display RMSEs showed a significant correlation with the display similarity scores calculated by these functions (right two columns). The correlation was stronger for the line displays ($r = –0.6412$; $p < .001$) compared to the grating displays ($r = –0.4491$; $p < .001$), which could be in part due to the greater frequency of guessing responses that occurred for the grating displays.

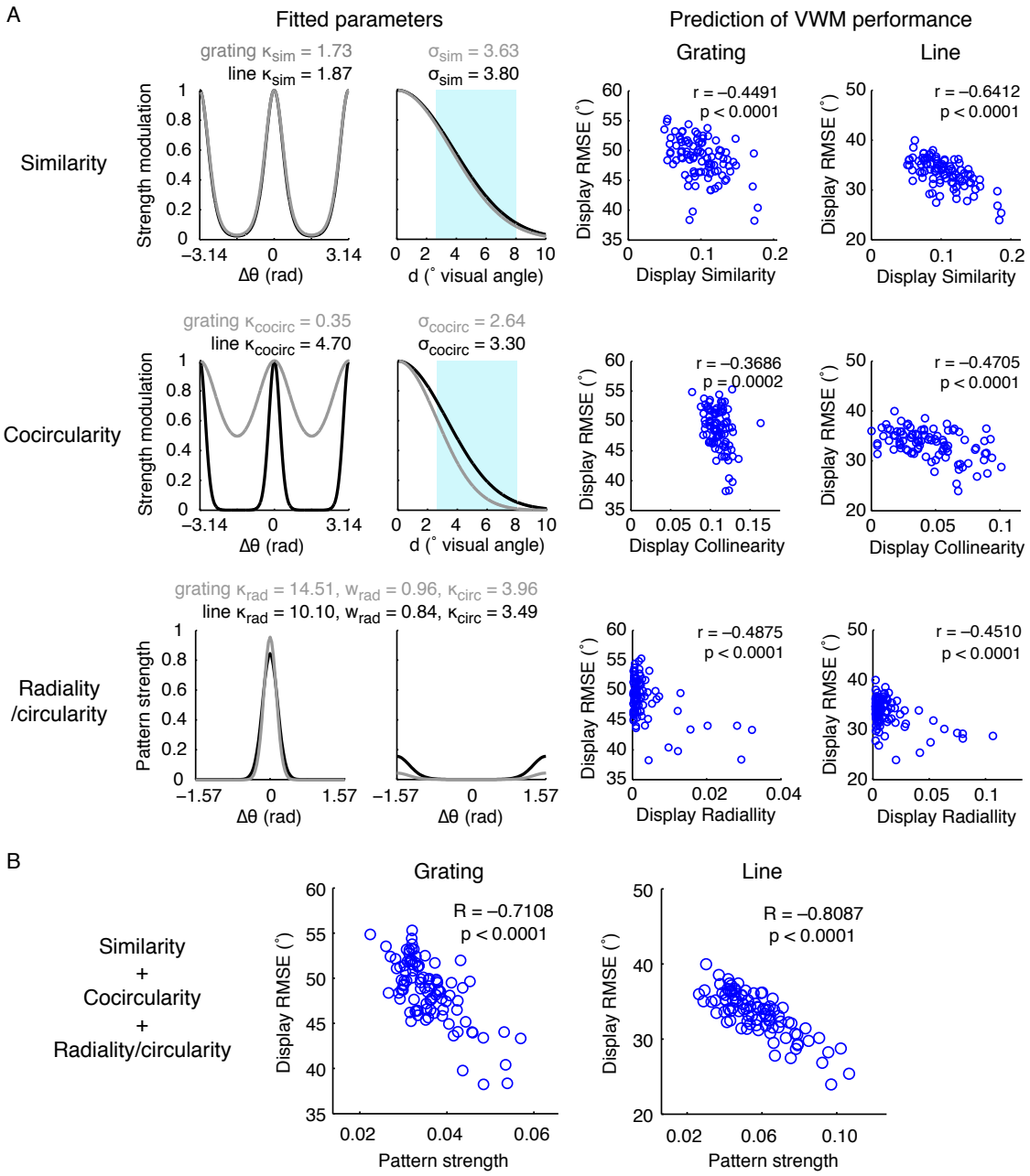

Figure 27. The results of the Gestalt-inspired model. (A) The similarity, cocircularity, and radiality/circularity grouping functions (left two columns), based on the best-fitting parameters for the grating (gray) and line (black) data. These parameters were fitted simultaneously to the data from each stimulus type. The shaded region in the distance sensitivity function (second column) for similarity and cocircularity indicates the range of possible distance between items in the current stimulus set. The correlation between the strength of each display property and the display RMSE score is shown on the right two columns, separately for the grating and line data. (B) The model's prediction of display RSMEs based on the combined strength of display similarity, cocircularity, and radiality/circularity.

For the *cocircularity* component, orientation sensitivity was somewhat sharper (i.e., a higher κ_{cocirc}) for the line displays than for the grating displays[4], while the cocircularity strength for the grating displays tended to decline more sharply with increasing distance between the items (i.e., a smaller σ_{cocirc}) compared to that for the line displays. Although the significance of these differences is not yet clear, these results suggest that the visual system might be more sensitive to the cocircular structure defined by line items and capable of detecting cocircular relationships between relatively more distant items. Again, the correlation between the display collinearity and RMSE was highly significant for both grating ($r = –0.3686$, $p < .001$) and line data ($r = –0.4705$, $p < .001$), with a stronger relationship observed for the line data (**Figure 27A**; middle row, right two columns).

The *radiality/circularity* component revealed a pronounced sensitivity for globally radial patterns, and a much weaker sensitivity for globally circular patterns, for both stimulus types. This can be appreciated by comparing the heights of the two von Mises functions representing the radiality (centered at 0˚) and circularity components (centered at 90˚). (However, it should be noted that circular displays also lead to high cocircularity, so a component of this variance will be accounted for based on this grouping measure as well.) The asymmetry between the sensitivities to radiality and circularity was greater for the grating displays than for line displays, as demonstrated by the sharp peak of the pattern strength around 0˚ and the nearly flat modulation around 90˚. Radiality seems to play a particularly important role in VWM for gratings, as this component accounted for the largest variability across grating displays among the three components. Moreover, radiality/circularity was the only component that provided a better prediction of the RMSEs for the grating displays ($r = –0.4875$, $p < .001$) than it did for the line displays ($r = –0.4510$, $p < .001$).

For the overall pattern strength, similarity, cocircularity, and radiality/circularity

[4] The orientation sensitivity function for cocircularity might appear as if the overall cocircularity strength is greater for gratings than for lines. However, the range of the absolute function values is not so relevant as the *relative* function values and the *shape* of the gradient, in terms of the model's ability to predict the variable RMSE scores across display. This is because a uniform elevation of the absolute function values across different orientation deviations would simply add a constant to the pattern strength for *all* displays, without accounting for any additional variability in RMSE across displays.

strengths were combined with weights of 0.14, 0.18, and 0.68 for grating displays, and 0.42, 0.17, and 0.41 for line displays. As shown in **Figure 27B**, the prediction of display RMSE was considerably improved by combining these three components (R = –0.7108 for grating; R = –0.8087 for line). R^2 values indicated that the model could account for 50.5% of the variance in RMSEs across grating displays, and 65.4% of the variance in RMSE across line displays.

To provide a better indication of the shared perceptual organization mechanisms underlying VWM for grating and line displays, I examined how well the model's prediction of display-specific VWM performance generalized across stimulus types. I also sought to examine how the stimulus format influences the modulation of VWM performance by pattern strength. As shown in **Figure 28**, the same pattern strengths (z-scored for comparison across stimulus types) in terms of similarity, cocircularity, and radiality/circularity of display elements, could reliably predict the VWM performance for specific configurations of both grating (gray circles) and line (black circles) stimuli, even when the model was fitted to one stimulus type and tested on the other. VWM performance for both stimulus types exhibited similar patterns of modulation by the combined display pattern strength (**Figure 28**, bottom row), with the RMSEs for grating displays showing a nearly additive increase relative to those for line displays, across all levels of pattern strength.

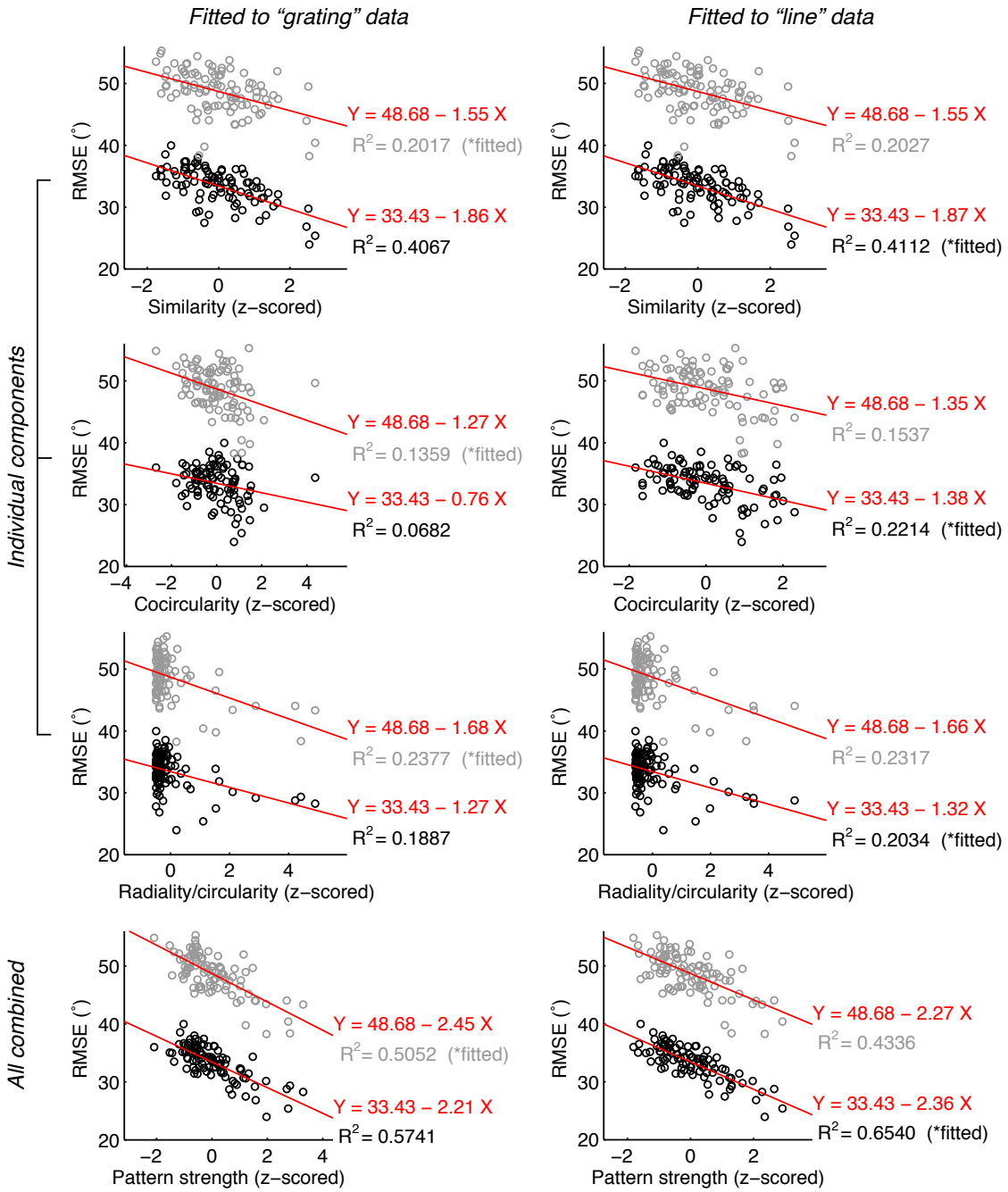

Figure 28. Generalization of the model's prediction across stimulus types. On each panel, VWM performance for grating displays (gray circles) and line displays (black circles) in terms of RMSE is plotted against the z-scored pattern strengths of the displays, along with the best-fitting regression line for each stimulus type (red). The pattern strengths on the left and right columns are based on the models fitted to the grating and line data, respectively. The generalization of prediction based on the strengths of individual components fitted together is shown on the top three rows. The generalization based on the combined model is shown on the bottom row.

I conducted a series of simulation analyses to characterize the observed relationship between pattern strength and VWM performance for grating and line displays. Delayed estimation responses for line displays were generated using a set of mixture model parameters that produced the observed range of performance variability in terms of RMSE scores (see **Figure 29A**, black solid line with squares). I assumed that VWM capacity (K) for line displays linearly increases from 3 to 6 items (i.e., P_{memory} increases from 0.50 to 1.00 with set size 6) as the display pattern strength increases, with a modest trade-off in memory precision (SD), which linearly increases from 17.1˚ to 24.9˚. Based on these parameters, error data were generated for 10 levels of display pattern strength, 600 trials each. The RMSE was calculated for these error data. This was repeated for 50 times at each level of pattern strength, resulting in 50 display RMSE scores at each level of pattern strength. On the rightmost column of **Figure 29A**, the mean RMSE score across 50 stimulated error distributions is shown as a function of display pattern strength (black squares).

Next, I generated error data for grating displays assuming three different scenarios regarding the relationship between grating and line performance. The first scenario assumes that VWM performance for grating displays is not modulated by display pattern strength, such that both K and SD remain constant (K = 2.25, SD = 21˚) across all displays (**Figure 29A**, gray dashed line with triangles). The second scenario assumes that P_{memory} is reduced by 50% for grating displays relative to that for line displays, resulting in halved memory capacity for grating displays across all levels of pattern strength (K = 1.5~3.0 items; **Figure 29A**, gray solid line with circles). For simplicity, memory precision for grating displays is assumed to be identical to that for line displays, showing the same decline as a function of increasing pattern strength. The last scenario assumes that P_{memory} is reduced by a constant amount (–0.375) for grating displays relative to that for line displays, resulting in a reduction of memory capacity by 2.25 items across all displays (K = 0.75~3.75 items; **Figure 29A**, gray solid line with x's). As can be seen on the rightmost panel in **Figure 29A**, this last scenario leads to a sharper modulation of VWM performance by display pattern strength in terms of RMSE scores (i.e., the overall amount of stored information), compared to the first two scenarios. The actual modulation of VWM performance by display pattern strength (shown in **Figure 28**, bottom panels) appears more

consistent with the second scenario, which assumes a proportional reduction of memory capacity for grating displays relative to line displays.

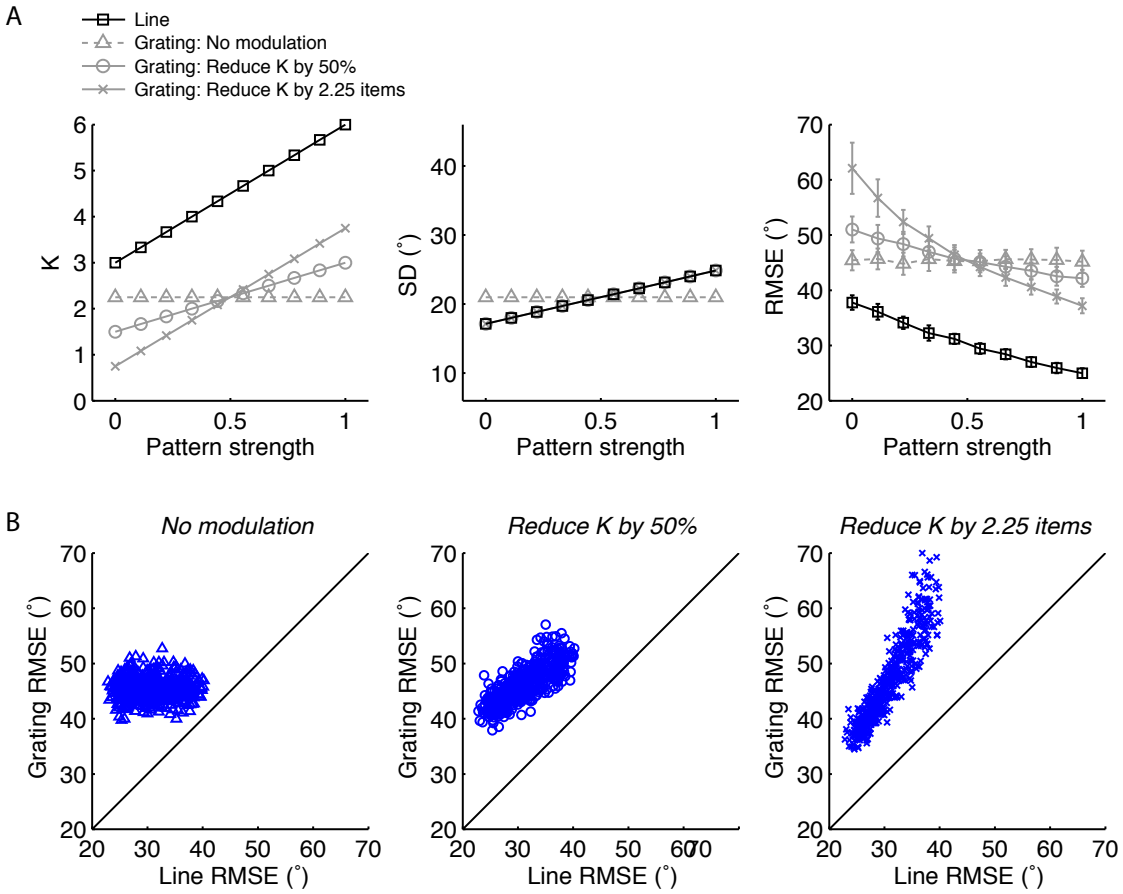

Figure 29. Simulations of the effects of display pattern strength on VWM performance for gratings and lines. (A) The left two panels show memory capacity (K) and precision (SD) parameters used to generate delayed estimation responses (600 trials per display) for grating and line displays, as a function of display pattern strength. The rightmost panel shows the mean display RMSE score across 50 simulated response distributions (600 trials per display), as a function of display pattern strength. The error bars represent ±1 SD of RMSE scores across 50 simulations at each level of display pattern strength. (B) The relationship between RMSE scores for simulated grating and line data, assuming three different scenarios regarding the impact of stimulus format on the sensitivity to perceptual grouping cues.

Figure 29B shows the relationship between RSME scores for the simulated grating and line data, based on each of these three scenarios. The relationship observed in the

actual data (see the bottom left corner of **Figure 21**) appears quite consistent with the second scenario (**Figure 29B**, middle panel), suggesting that VWM is sensitive to the perceptual grouping cues in both grating and line displays, with the exception that the probability of successful memory maintenance is much lower for grating displays. The empirical and modeling results found here, as well as the simulation results, reveal an important yet non-intuitive finding, namely that common mechanisms of perceptual grouping can operate on both line and grating stimuli, yet the impact of grouping on effective memory capacity can still vastly differ across the two stimulus types.

How is this possible? Consider the results from Chapter 3, in which estimates of the probability of grouping were far greater for lines (0.63 to 0.71) than for gratings (0.28 to 0.31), leading to a substantial difference in effective capacity at the largest set size. Although estimated parameters for each grouping mechanism are very similar for lines and gratings, leading to similar orientation and distance tuning functions, the absolute magnitude in which these grouping interactions scale to the effective capacity is clearly much weaker for gratings than for lines. In this context, the absolute probability of grouping, or effective response *gain* of these grouping mechanisms, can differ across gratings and lines while the shape of their tuning functions remain invariant to changes in stimulus form. Such invariance of tuning is well documented in the vision literature, such as the stability of orientation tuning across changes in stimulus contrast. With respect to this study, the current findings suggest that perceptual grouping mechanisms may operate with quite stable or invariant tuning properties across stimulus forms, though this remains a question that should be studied in greater detail in future work.

Discussion

In the present study, I examined the role of perceptual organization in the efficient storage of orientation information in VWM. By obtaining a large data set measuring delayed estimation performance for 96 random arrangements of oriented gratings and lines, I found that specific stimulus configuration had a major impact on how accurately a given display could be remembered, for both stimulus types. The magnitude of estimation errors varied substantially across displays, in a manner that was highly reliable across participants, as well as across the two stimulus types. In order to determine how much of

this VWM performance variability can be explained by the variability in perceptual organization across displays, I constructed a model of display-specific VWM performance, informed by well-known Gestalt grouping principles. By quantifying the strengths of similarity, cocircularity, and radiality/circularity structures that the observers might extract from each stimulus configuration, the model could successfully predict VWM performance for specific displays, accounting for 50.5% and 65.4% of the variance in error magnitude for grating and line displays, respectively.

VWM performance for both stimulus types showed largely comparable sensitivities to each of the grouping cues included in the model. While the model could more reliably predict the VWM performance for line stimuli than that for grating stimuli, both stimulus types were similar in terms of the magnitude of modulation by the display pattern strength. Simulation analyses indicated that the observed relationship between VWM performance for grating and line displays, as well as their modulation by display pattern strength, could be reasonably well characterized by the joint influences of perceptual grouping and stimulus form on the number of items encoded in a particular display.

These results suggest that perceptual organization plays an important role for VWM for *both* grating and line orientation, and the enhanced capacity for storing line orientations cannot be attributed to a more effective operation of perceptual organization processes on the line stimuli. It appears that the visual system can extract regular spatial structures present in both grating and line stimuli, providing higher-order perceptual units that can be stored in VWM efficiently. These perceptual units seem to be utilized by VWM with different levels of efficiency depending on the stimulus format, such that the storage efficiency for line stimuli is enhanced over that for grating stimuli by a similar proportion (i.e., a twofold increase in the number of stored orientations) across *all* levels of perceptual organization.

The finding that perceptual grouping cues, such as parallelism and collinearity, benefit VWM performance not just for line orientation, but also for grating orientation is not surprising given the prevalent use of both stimulus formats in research on the role of the *good continuation* principle for contour integration. In a classic study, Field and colleagues (1993) tested the observer's ability to detect the presence of an elongated path, formed from oriented Gabor patches embedded in an array of randomly oriented distractor

elements. They found that observers could easily detect the contours where the orientations of the Gabors were aligned smoothly along the path, but not the contours where the orientations were randomly varied, demonstrating the effective grouping of Gabor elements by the principle of good continuation (see also Ledgeway, Hess, & Geisler, 2005; Machilsen, Pauwels, & Wagemans, 2009). Other studies have instead employed oriented line segments for the contour detection task (Geisler et al., 2001; Tversky, Geisler, & Perry, 2004), or both line and Gabor stimuli across different species (Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003). Oriented lines seem to be the preferred choice for studying the neural mechanisms of contour grouping (e.g., Kapadia, Ito, Gilbert, & Westheimer, 1995; Wannig, Stanisor, & Roelfsema, 2011). Thus, both stimulus formats seem to be used interchangeably to tap into the fundamental perceptual processes of the visual system. However, it would be of interest for future studies to explore how these two stimulus formats might differ in terms of sensitivity to various grouping cues, including higher-order properties, such as mirror symmetry or more complex shapes.

In the Gestalt-inspired model of VWM performance, I integrated proximity with each of the similarity and cocircularity components rather than modeling it as a separate component, such that any performance benefits arising from two nearby items would depend on the specific alignment of their orientations. In other words, two similar or cocircular orientations would be perceived as a coherent unit when they are close together, but as independent units when they are distant. Integration of proximity with similarity or cocircularity may be adaptive given the ecological statistics of proximity cues in natural scenes (Brunswick & Kamiya, 1953; Geisler et al., 2001), in which local features tend to be correlated, but less so at larger distance separations.

The modeling results confirmed that the benefits of similarity and cocircularity on VWM performance interacted with proximity, as indicated by the sharp decline of similarity and cocircularity strengths (σ_{sim} = 3.63-3.80˚; σ_{cocirc} = 2.64-3.30˚) across the possible range of distance between items (d = 2.62-8.00˚). When the distance sensitivity function was removed from both similarity and cocircularity components, the model's overall prediction (R^2) of display-specific VWM performance dropped from 50.5% to 40.2% for gratings, and from 65.4% to 52.5% for lines, providing further support for this conclusion.

The present finding is also consistent with previous reported effects of proximity and similarity in VWM for orientation and color. In Orhan and Jacobs' (2013) study, memory representations of orientations for two Gabor items showed a strong positive correlation when the actual stimulus orientations were similar, indicative of similarity-based grouping, but this dependency between stored representations gradually decreased as the distance between the two stimuli increased. In a color change detection task, a significant benefit of color repetition on the sample array was found only when the two same-colored items were next to each other (Peterson & Berryhill, 2013), providing converging evidence that similarity-based grouping is constrained by spatial proximity.

In the present model, similarity and cocircularity were defined by local grouping functions, in which the display-level strength was essentially a *summed* strength of all pairwise relationships. On the other hand, radiality/circularity was defined by a global grouping function, in which the display-level strength was determined by the configuration of the entire display, which could not be fully captured by simply summing across the strengths of individual items or pairs of items. The radiality/circularity component was initially motivated by the observation that items whose orientations were aligned with the radial axis of the display or tangential to the virtual circle tended to be reported more accurately than the rest of the items. This radial/tangential benefit was characterized by a heightened frequency of responses centered on these *location*-dependent orientation values, which occurred largely independently of the true *orientation* of the probed item. Such biases have been previously observed in delayed estimation tasks using a circular stimulus layout (e.g., Park, Rademaker, & Tong, VSS 2014), but only under high memory load that exceeded the typical estimates of working memory capacity. In the current study, the radial/tangential bias was more pronounced in the response distribution for grating stimuli than that for line stimuli, which raises the possibility that the observed effects of radiality/circularity on the display performance may at least in part reflect a guessing strategy based on the location of the probed item.

Given that individual items with radial or tangential orientations can benefit from such response biases, one might expect to find an additive effect of item-level radiality/circularity strengths on the overall accuracy of a display. However, I found little correlation between the sum of individual items' radiality/circularity scores and the display

RMSEs, which is hard to reconcile with the strong effects of individual items' radiality/circularity scores on their accuracy. A substantial modulation of the display RMSE by radiality/circularity could be revealed only after applying a constraint that items on a display should be either *all radial* or *all circular* to achieve a high display-level score. One way to reconcile the inconsistent effects of item-level and display-level radiality is to assume that VWM benefits from a globally radial or circular pattern, which leads to reduced random guessing in relatively few displays, and that the incoherent configurations produced by a random mix of radial and tangential items, which are more common, leads to increased guessing in those displays. The cost of such poor configurations might effectively cancel out the benefit of radial/tangential guessing strategy occurring in these displays, resulting in performance indistinguishable from the displays with intermediate orientations. Thus, the observed benefit for individual radial/tangential items might be in part due to the configural benefit in remembering these items, in addition to the pure guessing strategy.

The present study focused on demonstrating and explaining the variability in VWM performance across *displays* rather than the variability across individual *items* within a display, both of which were highly systematic and reliable. The primary reason was that display-level performance was a more reliable indicator of how efficiently a given display could be remembered on average, regardless of potential biases in memory resource allocation across the items. On the other hand, item-level performance could fluctuate within a display due to a prioritized encoding of certain items into VWM, reflecting a trade-off among items, which does not necessarily indicate an increased availability of memory resources due to grouping. However the item-level similarity and cocircularity scores, obtained by averaging across five pairwise relationships per item, were generally correlated with the individual item's accuracy, suggesting that the local perceptual grouping also benefited memory for individual items. It remains to be determined whether these benefits reflect increased perceptual precision for these items (i.e., information that can be stored per unit VWM resource or attention) or prioritized allocation of limited VWM resources (or attention) to items with salient grouping cues. It is also important to determine whether different grouping cues, such as similarity and cocircularity, might be optimally combined or compete within a display, by examining the systematical biases in

the reported orientations, as these grouping cues would induce different types of biases (e.g., attraction toward nearby similar items to make them more parallel vs. repulsion in the opposite direction to make them more cocircular).

The strong effects of perceptual organization on VWM performance for both grating and line displays suggest that VWM advantage for line stimuli cannot be explained by more effective perceptual grouping for these stimuli. Rather, perceptual organization critically modulates the storage efficiency for both stimulus formats, with line stimuli receiving relative benefits in efficiency over grating stimuli across all levels of perceptual organization. One possibility for this joint influence of perceptual organization and stimulus form on VWM performance is that attention could more readily spread among the perceptual groups formed by lines than those formed by gratings. The spread of attention might be interrupted by the round boundary perceived in each grating, making it difficult to simultaneously attend multiple perceptual groups, or a higher-order group, formed by grating stimuli.

The present findings have important implications for understanding the architecture of VWM, as well as the nature of perceptual grouping among oriented elements. Perceptual grouping is important not just for correctly delineating the units of perceptual experiences into parts, objects, and surfaces, but also for understanding how an arbitrary configuration of oriented elements is actively held in mind and how such information can represented more efficiently in the mental workspace.

Chapter 5

General Discussion

Summary of findings

In this dissertation, I explored the role of stimulus form in determining how much orientation information can be encoded in VWM. I presented a series of behavioral findings demonstrating a substantial increase in the capacity of VWM for storing line orientation compared to grating orientation, mediated by visual grouping of multiple line orientations. I proposed two novel modeling methods, which formally characterized the visual grouping mechanisms that allow for a highly efficient storage of orientation information in VWM.

In Chapter 2, I used a delayed estimation paradigm and the mixture model approach to estimate the number of orientations stored in VWM and their representational precision. I found that memory precision was comparable for grating and line orientation, but that estimated capacity was about twice as large for lines as compared to gratings. The capacity advantage for lines was severely disrupted by sequential presentation of the stimuli, while the capacity for gratings was unaffected by sequential presentation, suggesting that visual grouping of simultaneously presented items is crucial for the capacity advantage for lines. Performance at a change detection task also benefitted from presenting an entire array of lines at test, rather than just a single test item, implying a benefit of relational coding.

In Chapter 3, I examined how visual grouping can lead to a greater proportional increase in the effective capacity of VWM when there are more items to be remembered in the memory array. I found that increasing the number of items in the memory array produced a much shallower decline in VWM performance for line arrays than for grating arrays, such that the total amount of stored information for line arrays, considering both the number and precision of stored orientations, continually increased with increasing set size. This difference between gratings and lines, in terms of their sensitivity to the increased memory load, could not be readily explained by the existing discrete-slots model, in which the amount of available memory resources is set at a constant value across set sizes, as well as across stimulus formats. However, by incorporating a probabilistic grouping process into the discrete-slots model to allow multiple items to be compressed into a coherent

memory unit, I could successfully account for the observed changes in information capacity across set sizes and across stimulus types while keeping the number of working memory slots constant within an individual.

In Chapter 4, I investigated the role of perceptual organization mechanisms in improving the efficiency of VWM for storing orientation information. I found substantial variability in the magnitude of delayed estimation errors across randomly generated arrays of oriented elements, which was highly consistent across participants and across stimulus types. To account for this display-specific effect on VWM performance, I developed a Gestalt-inspired model with multiple perceptual grouping mechanisms to detect the presence of different types of structure in these displays. This involved specifying mechanisms to encode local similarity and cocircularity relationships between pairs of items, as well as the global radiality/circularity pattern across the displays. This model could account for 50.5% of the variance in VWM performance across grating displays, and 65.4% of the performance variance across line displays. These results demonstrated that the observers can exploit perceptual grouping cues even from the arbitrary arrangements of orientations, allowing for a highly efficient storage of orientation information in VWM.

The relationship between visual grouping and summary statistics

Studies of visual perception have shown that observers can quickly and accurately compute statistical information about multiple objects in a scene, such as the average orientation (Alvarez & Oliva, 2009; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), location (Alvarez & Oliva, 2008), and size (Ariely, 2001; Chong & Treisman, 2003) of multiple elements in a briefly presented display. Of potential relevance, a recent study found that the orientations conveyed by line stimuli can be more efficiently averaged than grating orientations (Choo, Levinthal, & Franconeri, 2012). One might wonder if VWM capacity advantage for line orientation is due to the fact that such summary information is more readily accessible from line stimuli than from grating stimuli.

However, the ability to hold multiple pieces of information in VWM and the ability to extract summary statistics from a display represent two distinct cognitive abilities. A typical VWM task requires the *individuation* of each object in the display, by retaining precise information about the feature and spatial location of each object, whereas the

averaging task requires *abstraction*, discarding information about individual items so that attentional and working memory resources can be freed up to focus on encoding and maintaining a more accurate representation of ensemble statistics (e.g., Feigenson, 2008; Halberda, Sires, & Feigenson, 2006). I argue that the representation of the mean feature value, apart from the individual features, can only be useful as a secondary buffer in a delayed estimation task, as it may allow the observer to make an informed guess about the probed item when that item is not individually represented in working memory. If information can be redundantly coded at multiple levels of abstraction (Brady & Alvarez, 2011), as an item and as an ensemble, then it would always be adaptive to use all of the independent sources of information to improve the overall accuracy of estimation across trials.

Although the summary statistics can be useful if the feature values in a display tend to be highly correlated, allowing VWM to take advantage of the learned statistical regularities (e.g., Brady, Konkle, & Alvarez, 2009), such summary information is far too coarse to meaningfully enhance performance in a delayed estimation task with many randomly generated stimuli. A reliance on the mean feature value can lead to an accurate response if, by chance, most items in the sample array have highly similar feature values, which would serve to reduce the frequency of extreme errors (i.e., random responses). However, such benefits would occur on a only small proportion of trials in our task, since the feature values of each of the six items were highly variable and independently generated. As the set size increases, the information shared between each item and the mean becomes smaller, reducing the benefit of this additional source of information. Therefore, this global summary representation is unlikely to produce the pronounced increase in VWM capacity for line orientation.

While the global mean and variance statistics could not have played a major role in enhancing performance for line stimuli over grating stimuli in general, the gist representation can account for the whole-array probe benefit in change detection for lines compared to the single-item probe condition (Chapter 2, Experiment 4). Previous studies have suggested that observers are better at detecting changes that alter the global statistics of the display, such as the mean or variability of colors (Brady & Tenenbaum, 2013; see also Brady & Alvarez, 2015b). According to Brady and colleagues, such summary

information can boost change detection performance even in the absence of information about the feature values of individual items, and can thereby *inflate* the estimated capacity of working memory. In my change detection tasks, the observers might have noticed a change in the mean line orientation across displays or an increase/decrease in the variability of the orientations contained in the whole-array probe, even though they may not have been aware of exactly which item had undergone a change. Such gist information would increase the probability of correct "hit" responses for the whole-array probe, but cannot account for the differences between line and grating performance in the single-item probe condition for which item-specific information is required.

This summary statistics account is different from my visual grouping account, according to which grouped line items held in VWM are disrupted due to the presentation of a single-item probe. The visual grouping account suggests that the whole-array probe provides a more accurate estimate of the number of line items initially encoded in VWM, whereas the single-item probe *underestimates* this capacity. Phenomenally, the location of the changed line item on the whole-array probe would pop up, without having to rely on gist to make a guess. It remains an open question which of these two accounts correctly characterizes the level of awareness of individual items on the whole-array change detection. One way to test these accounts would be to compare the performance between change detection and change localization for line stimuli, with the whole-array probe. While the whole-array probe benefit in change detection for line stimuli can be explained either by the summary representation or by the grouped item-representation, the latter seems to provide a more parsimonious account of both the large capacity advantage for line stimuli and the sensitivity to the probe stimulus in change detection for lines.

The relationship between visual grouping and ensemble representations

Ensemble representation is used in the literature as an umbrella term to refer to a variety of concise visual representations, from the simple statistical summary of a scene (e.g., the mean stimulus location in Alvarez & Oliva, 2008) to a higher-order spatial layout provided by the clustering of similar features into distinct regions or textures, to so-called *spatial* ensembles (Alvarez & Oliva, 2009; Brady & Alvarez, 2015a; Brady & Tenenbaum, 2013). The latter type of ensemble representation is less gist-like and more configural in

nature, essentially determined by similarity and proximity grouping principles. For example, a display can be summarized as two distinct regions of warm and cold colors (Brady & Tenenbaum, 2013) or two distinct textures of Chinese characters and 3D cubes (Brady & Alvarez, 2015b). Compared to the global mean or the variance statistics, spatial ensemble representations are believed to preserve more specific featural and spatial information, thereby providing more useful information for correct estimation of individual items. Thus, there is a good deal of conceptual overlap between the spatial ensemble representation proposed by Brady and colleagues and the perceptual groups defined by similarity and proximity (Chapter 4).

However, such similarity-based ensemble representations cannot fully capture the rich spatial structure present in arrays of oriented elements, comprised of higher-order configural information, which can be detected by perceptual grouping mechanisms sensitive to cocircularity, radiality, and symmetry. While grouping based on *texture* requires high feature similarity or parallelism, grouping based on *contour* (i.e., by a smooth variation of feature values along a path) or *shape* provides a means to efficiently represent elements that are not necessarily similar to one another. For example, even a heterogeneous set of orientations can be represented by a highly efficient code if all elements lie tangent to a common circle or are aligned with the radial axis of the array. In my Gestalt-inspired model of VWM performance for orientation, these configural properties are processed by grouping mechanisms sensitive to the precise geometric relationships among the items, and VWM's sensitivity to each property is revealed concretely in the form of local orientation-tuned grouping functions. Thus, the present work goes well beyond previous work on ensemble representation that relies on the global mean or local feature similarities, and provides concrete evidence that VWM can exploit even richer configural properties in the display.

Implications for current models of VWM capacity

Recent research in VWM has been driven by two broad categories of research questions. One domain concerns characterizing the capacity limitations of VWM and understanding why such limitations arise, especially through the development and testing of different working memory models. Another major domain of inquiry concerns

understanding what information is represented in VWM and how it is organized. Even though these two areas are potentially relevant to one another, they have been studied separately for the most part. The critical difference between the two approaches lies in what constitutes a representational unit in VWM.

Most theories of VWM share the fundamental assumption that individual visual objects, or items, are represented by independent memory units. A discrete-slots model of VWM proposes that VWM can hold no more than 3 or 4 discrete objects at a time (Luck & Vogel, 2013, for review), while the continuous-resource view proposes that there is no upper limit on the number of objects that can be held, but that the distribution of finite memory resources across all of the objects in the memory array leads to a reduced quality of stored representations (Ma, Husain, & Bays, 2014, for review). In yet another account of VWM, the capacity of VWM is specified by both the visual information load of each object and by the number of objects, such that the working memory capacity tends to be greater for simple objects and tends to be reduced for more complex objects (Alvarez & Cavanagh, 2004). A common theme across all of these theories of VWM is the assumption that individual objects are stored as independent memory units, which allows for a quantification of the amount of stored information in terms of the number and/or precision of memory for each stored item.

The present work challenges the fundamental assumption of item-based memory units, and suggests that a variable number of items can form a coherent memory unit depending on the strength of visual grouping. The modeling work I presented in Chapter 3 (*the slots-plus-averaging model with grouping*) demonstrates that the large difference in the VWM capacity estimates between grating and line stimuli can be explained by more efficient compression of the line displays into fewer memory units, leading to an apparent lack of a fixed item-limit for line stimuli. Moreover, probabilistic visual grouping between pairs of items can account for the relatively shallow decline in VWM performance for line stimuli as a function of increasing set size, which the existing discrete-slots model fails to capture. By incorporating probabilistic grouping, the new discrete-slots model allowed for some fluctuations in memory precision across trials due to stochastic variability in the effective memory load (i.e., number of groups), and hence, a variable number of slots assigned to each group. Trial-to-trial variability in memory precision and the lack of fixed

item-limits are two key aspects of VWM performance that are readily captured by assuming a continuous and variable resource, as was proposed in the variable precision (VP) model (van den Berg et al., 2012). The present modeling work suggests that such aspects of VWM performance can be also explained by the variability in the allocation of discrete slots to perceptual groups that are detected in individual displays. The model I have developed provides a compelling account for why the precision of working memory performance will vary across trials and especially across displays, in a manner that is not addressed by variable precision model that assumes a stochastic distribution of memory resources across items and trials.

Other researchers have emphasized going beyond an item-based approach of studying capacity limits, and focusing on the encoding of higher-order information in VWM, to examine how structured representations influence VWM performance (Brady, Konkle, & Alvarez, 2011, for review). These studies have contributed to uncovering various contextual effects in VWM performance even in tasks using simple arbitrary stimuli, such as evidence of a reliance on summary statistics as described above. This work has provided important evidence to challenge the conventional notion that individual items in a display are represented entirely independently of one another in VWM. However, the structured representations characterized in these studies are assumed to operate independently of the severe limits of working memory capacity, and issues of how the formation or the utility of such representations might interact with the individual's working memory capacity remains unresolved. In the present work, I focused on the extent to which the item-level information capacity of VWM can be increased by such structured representation. In my SA model that incorporated grouping, the representations of individual items are integrated into a higher-order visual group, while preserving the item-specific feature and location information. These visual groups serve as effective memory units (i.e., chunks), such that the storage of groups of different sizes is subject to the same capacity limit of working memory. Thus, the present research is a novel attempt to bridge these two lines of research – one that focuses on the capacity limit, and one that focuses on the contents and higher-order structures – by integrating the effects of visual grouping with the limited capacity of working memory, a fundamental characteristic of this cognitive function.

Comparison with other models of the organization of VWM

Recently, a handful of models have been proposed with the goal of accounting for systematic variability in VWM performance based on contextual stimulus factors (Brady & Alvarez, 2015a; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013). In Orhan and Jacobs' (2013) probabilistic clustering theory (PCT), VWM uses organizational processes that assign clusters or partitions for a set of items in an array, based on the clustering of noisy estimates of feature values (or locations, for spatial memory). This clustering process is implemented via a Dirichlet process mixture model, which automatically infers the probability distributions over all possible partitions of the items. This allows an item to be represented at multiple granularities simultaneously. For example, an item can belong to a cluster of its own (item-level) or a cluster that contains all of the items (display-level) or any levels in between. As a result, a representation of an item depends on the feature values of all of the other items on the displays, as well as its own. The probabilistic clustering theory predicts biases in estimates of items toward the mean of the cluster, as well as correlation between the estimates of items that are assigned to the same cluster, because the presumed observer treats the items within a cluster as arising from the same object. Both predictions were confirmed by the behavioral data. In contrast to the item-limit model, the probabilistic clustering model does not impose a fixed capacity limit on the encoding of items, nor on the maximum number of clusters that items can be grouped into. All items are automatically assigned to multiple levels of clusters, such that no item is completely lost, and the variable partitions of a display lead to the variable precision of clusters and memories.

While the probabilistic clustering model and my Gestalt-inspired model (Chapter 4) are distinct in terms of their underlying architectures, as well as the aspects of VWM performance they seek to explain, there are some similarities between them. In the probabilistic clustering model, clustering entails a loss of individuating information about each item, thereby introducing a bias toward the mean of similar items. At the same time, clustering of similar items enhances the precision of the cluster, which implies that an optimal clustering of similar items can lead to a reduction of the expected error of the estimates. My Gestalt-inspired model revealed that strong similarity grouping cues in a display led to a reduced overall RMSE for the items in the display. This could reflect an

optimal trade-off between the biases in the item estimates toward the group mean and the enhanced precision of memory for the grouped items. In addition, the probabilistic clustering model integrates over all possible partitions of the display, such that any given item's representation is dependent on all other items. In my model, display-level similarity is calculated by averaging across all pairwise similarity of items in a display. Both feature-based clustering and similarity grouping is consistent with the visual system's parallel detection of feature similarity across all image elements.

On the other hand, my Gestalt-inspired model offers novel predictions about the interaction of proximity and feature similarity, and provides well-specified mechanisms to account for other grouping cues, such as cocircularity and radiality. In the probabilistic clustering model, clustering was based on a single task-relevant feature dimension, such that orientation memory was affected by orientation similarity, and location memory was affected by spatial proximity. My model results demonstrate that grouping among similar features depends on their spatial proximity, which yields a prediction that dependencies between the two feature values reported in the same trial should be strongest for two nearby items, and the two reported feature values should become more independent of each other as they are more separated in space. Moreover, such dependencies should also be observed between pairs of highly cocircular orientations, which would show systematic biases toward a more cocircular orientations, rather than their mean orientation.

Brady and Alvarez (2015a) have proposed a similar probabilistic model of structured representations in VWM, based on their characterization of working memory performance for a set of displays containing 3 color patches. As in Orhan and Jacob's probabilistic clustering model, Brady and Alvarez's hierarchical Bayesian model considers all possible clusters (i.e., ensembles) of items in a display, which are integrated in the observer's estimates of individual items. These multiple, interacting levels of representations could explain some of the variability in color estimation performance across displays and items. One modification from the probabilistic clustering model was the introduction of a *guessing* component to account for memory failure in delayed estimation. They assumed that the clusters containing more items are less likely to be corrupted, such that clustering of more items allows more items to be stored in VWM. The mechanism behind the effect of clustering on guess rate remains unexplained, but this

essentially allows the clusters to increase VWM capacity at the cost of introducing biases.

My Gestalt-inspired model was applied to the overall error scores (RMSEs) of the displays rather than the estimates of the number of remembered items (K) from a display and their precision (SD), because there was a modest trade-off between these two estimates across displays. A post hoc analysis indicated that higher model estimates of pattern strength were associated with higher K estimates for that display, whereas there was no clear relationship between pattern strength and display SD estimates. Thus, the reduced RMSE scores for high-similarity displays could be due to a large benefit in reduced guess rate that outweighs the modest cost of introducing biases in the individual items.

All of the models described above, including my Gestalt-inspired model, can successfully characterize some of the systematic variability in VWM performance observed across displays and items. What remains poorly understood, however, is the relationship between one's working memory capacity and the utility of such grouping processes in enhancing VWM performance. In Orhan and Jacob's probabilistic clustering model and Brady and Alvarez's hierarchical Bayesian model, the clustering process operates on the noisy perceptual samples that depend on both sensory and memory noise, but the clustering process itself is controlled by a separate set of parameters in the Dirichlet process. This implies that the clustering process and limited working memory capacity comprise two separate components of VWM. The precision of the inferred clusters, which is a critical factor determining the observer's performance in estimating of individual items, is defined in an obscured way, independent of the observer's sensory/memory encoding precision. Thus, it is unclear how much one can benefit from the use of such clusters in terms of overall VWM accuracy, as clustering of two or more non-identical items would always introduce biases in the estimates of the item.

The optimal clustering strategy may depend on the level of memory noise, which in turn likely depends on one's sensory precision and working memory resources. If the perceptual samples are highly noisy (e.g., due to a large set size), the benefit from the reduced variance of the cluster would be large enough to outweigh the bias introduced by the clustering. On the other hand, the trade-off between bias and variance may be suboptimal if the estimates are highly precise. This raises the possibility that individuals with low VWM capacities might be the ones who can benefit the most from clustering,

whereas individuals with high VWM capacities might do better by relying on precise representations of individual items. The relationship between VWM capacity and a person's sensitivity to perceptual grouping cues could be a fruitful area of research where the notion of capacity limit can be integrated with the structured representations.

The probabilistic clustering model and the hierarchical Bayesian model are computational models that can generate response distributions and can be fitted directly to the observers' psychophysical data. On the other hand, my Gestalt-inspired model (Chapter 4) was a statistical model, which sought to describe the relationship between the observed performance and the presumed perceptual grouping mechanisms. This model serves as good starting point for identifying a set of principles that can explain the variability in VWM performance across displays. It will be important for future studies to test this model's predictions on a novel set of stimulus parameters. Future studies could also incorporate insights from the current model and implement a computational model that can take the stimulus variables into account (cf. Pratte, Park, Rademaker, & Tong, 2017) and generate responses based on multiple perceptual grouping processes, as well as the allocation of working memory resources. Such a model could be fitted to an individual participant's delayed estimation data, allowing for separate estimation of the participant's working memory capacity and the efficiency of the perceptual organization process.

In chapter 3, I developed a discrete-slots model with probabilistic grouping in order to characterize the overall benefit of grouping in terms of the capacity and precision of VWM, rather than the effects of stimulus configuration and perceptual organization. As it turned out in my subsequent study, the effects of stimulus form (grating vs. line) revealed the operation of very similar perceptual grouping mechanisms across the stimulus forms. If common perceptual grouping mechanisms operate in both cases, then what can account for the overall advantage in memory capacity for lines over gratings? The enhanced storage efficiency for lines over gratings can be still considered as arising from visual grouping mechanisms, given that this benefit depended on the simultaneous presentation of multiple line items (Chapter 2, Experiment 3) and that the magnitude of the benefit increased as the memory array contained more items to be remembered. These observations suggest some kind of facilitative interactions between multiple line items held in VWM, which allow them to be treated as a functional memory unit by VWM.

Conclusions

A prominent theory of visual working memory proposes that there is a fundamental limit on the number of items that can be held in memory. An underlying assumption is that these capacity limits are invariant to the specific visual properties of the stimulus to be encoded. The present work challenges this longstanding assumption, by demonstrating that the capacity of VWM can be greatly expanded depending on the degree to which stimulus and display factors allow the observer to perceptually group multiple items into higher-order patterns. The dramatic effect of stimulus form on memory capacity requires a reconsideration of what constitutes an item, or a chunk, in the active maintenance of visual information. The information capacity of visual working memory cannot be considered separately from the ways in which visual information can be efficiently compressed by the visual system. The computational models developed in this work reveal how specific perceptual grouping mechanisms can organize oriented elements into larger configurations, thereby alleviating the severe limitations of working memory. A full understanding of the architecture of VWM thus requires characterization of both the nature of the contents being represented and the fundamental limitations in working memory capacity.

BIBLIOGRAPHY

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for A′ and B″ to below-chance performance. *Psychological Bulletin*, *102*(3), 439.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics*, *70*(2), 346–364. https://doi.org/10.3758/PP.70.2.346

Alvarez, G. A., & Oliva, A. (2008). The Representation of Simple Ensemble Visual Features Outside the Focus of Attention. *Psychological Science*, *19*(4), 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

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

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, *2*, 89–195.

Averbach, E., & Coriell, A. S. (1961). Short-Term Memory in Vision. *Bell System Technical Journal*, *40*(1), 309–328. https://doi.org/10.1002/j.1538-7305.1961.tb03987.x

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. *Psychological Science*, *18*(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D. (1986). *Working memory*. Oxford [Oxfordshire] : New York: Clarendon Press ; Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47–89.

Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7. https://doi.org/10.1167/14.4.7

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Becker, M. W., Miller, J. R., & Liu, T. (2012). A severe capacity limit in the consolidation of orientation information into visual short-term memory. *Attention, Perception, & Psychophysics*, *75*(3), 415–425. https://doi.org/10.3758/s13414-012-0410-0

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368.

Bettencourt, K. C., Michalka, S. W., & Somers, D. C. (2011). Shared filtering processes link attentional and visual short-term memory capacity limits. *Journal of Vision*, *11*(10), 22. https://doi.org/10.1167/11.10.22

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. https://doi.org/10.1177/0956797610397956

Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, *15*(15), 6. https://doi.org/10.1167/15.15.6

Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 921–929. https://doi.org/http://dx.doi.org/10.1037/xlm0000075

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.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brunswik, E., & Kamiya, J. (1953). Ecological Cue-Validity of "Proximity" and of Other Gestalt Factors. *The American Journal of Psychology*, *66*(1), 20–32. https://doi.org/10.2307/1417965

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*(4), 674–692. https://doi.org/http://dx.doi.org/10.1037/a0024695

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Choo, H., Levinthal, B. R., & Franconeri, S. L. (2012). Average orientation is more accessible through object boundaries than surface features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 585–588.

Cover, T. M., & Thomas, J. A. (2001). *Elements of information theory*. New York: Wiley.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87-114-185.

Cowan, N. (2010). The Magical Mystery Four How Is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*(1), 120–133. https://doi.org/10.1111/j.1467-7687.2009.00864.x

Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. *Psychonomic Bulletin & Review*, *16*(4), 641–647. https://doi.org/10.3758/PBR.16.4.641

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Delvenne, J.-F., & Bruyer, R. (2004). Does visual short‐term memory store bound features? *Visual Cognition*, *11*(1), 1–27. https://doi.org/10.1080/13506280344000167

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*(1), 170–178. https://doi.org/10.3758/s13423-014-0675-5

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, *2*(4), 5–5. https://doi.org/10.1167/2.4.5

Elder, J., & Zucker, S. (1994). A measure of closure. *Vision Research*, *34*(24), 3361–3369.

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*(6), 1127–1133.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, 102–134.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional Memory: Extraordinary feats of memory can be matched or surpassed by people with average memories that have been improved by training. *American Scientist*, *70*(6), 607–615.

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. *Journal of Neuroscience*, *29*(48), 15258–15265. https://doi.org/10.1523/JNEUROSCI.4388-09.2009

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893–905. https://doi.org/10.1016/j.neuron.2015.07.013

Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition*, *107*(1), 1–18. https://doi.org/10.1016/j.cognition.2007.07.006

Feldman, J. (1997). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835–2848. https://doi.org/10.1016/S0042-6989(97)00096-5

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*(2), 173–193. https://doi.org/10.1016/0042-6989(93)90156-Q

Fisher, N. I. (1995). *Statistical analysis of circular data* (Repr., 1. paperback ed). Cambridge: Univ. Press.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12). https://doi.org/10.1167/11.12.3

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27. https://doi.org/10.1167/10.12.27

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229. https://doi.org/10.1038/ncomms2237

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. https://doi.org/10.3758/17.5.673

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711–724. https://doi.org/10.1016/S0042-6989(00)00277-7

Geisler, W. S., & Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review*, *107*(4), 677–708. https://doi.org/10.1037//0033-295X.107.4.677

Gold, J. M., Fuller, R. L., Robinson, B. M., McMahon, R. P., Braun, E. L., & Luck, S. J. (2006). Intact attentional control of working memory encoding in schizophrenia. *Journal of Abnormal Psychology*, *115*(4), 658–673.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: computing formulas. *Psychological Bulletin*, *75*(6), 424.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, *92*(2), 173.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science*, *17*(7), 572–576.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Hess, R., & Field, D. (1999). Integration of contours: new insights. *Trends in Cognitive Sciences*, *3*(12), 480–486. https://doi.org/10.1016/S1364-6613(99)01410-2

Hollingworth, A. (2003). Failures of Retrieval and Comparison Constrain Change Detection in Natural Scenes. *Journal of Experimental Psychology. Human Perception & Performance*, *29*(2), 388.

Hsu, A. S., Griffiths, T. L., & Schreiber, E. (2010). Subjective randomness and natural scene statistics. *Psychonomic Bulletin & Review*, *17*(5), 624–629. https://doi.org/10.3758/PBR.17.5.624

James, W. (1890). *The principles of psychology, Vol I.* New York, NY, US: Henry Holt and Co, Inc.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702. https://doi.org/http://dx.doi.org/10.1037/0278-7393.26.3.683

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., … Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology*, *27*(2), 220–229.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology. General*, *130*(2), 169–183.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*(4), 843–856. https://doi.org/10.1016/0896-6273(95)90175-2

Kinchla, R. A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics*, *2*(6), 219–229. https://doi.org/10.3758/BF03212471

Koffka, K. (1935). Principles of Gestalt Psychology (New York, 1935). *Google Scholar*.

Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., & Logothetis, N. K. (2003). Integration of Local Features into Global Shapes: Monkey and Human fMRI Studies. *Neuron*, *37*(2), 333–346. https://doi.org/10.1016/S0896-6273(02)01174-1

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*(4), 225–234.

Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*(2), 149–164. https://doi.org/10.1016/S0042-6989(02)00402-9

Ledgeway, T., Hess, R. F., & Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: Empirical tests of "association field" models of

contour integration. *Vision Research*, *45*(19), 2511–2522. https://doi.org/10.1016/j.visres.2005.04.002

Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, *17*(3), 356–372. https://doi.org/10.1080/13506280701766313

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*(12), 11–11. https://doi.org/10.1167/9.12.11

Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7–8), 1213–1219. https://doi.org/10.1016/0042-6989(91)90046-8

Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 369–380. https://doi.org/http://dx.doi.org/10.1037/0278-7393.34.2.369

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103.

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

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortexYasushi Miyashita. *Nature*, *331*(6151), 68–70.

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 141–150.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750.

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.

Oyama, T. (1961). Perceptual Grouping as a Function of Proximity. *Perceptual and Motor Skills*, *13*(3), 305–306. https://doi.org/10.2466/pms.1961.13.3.305

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*(1), 29–55. https://doi.org/10.3758/BF03200760

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(8), 823–839. https://doi.org/10.1109/34.31445

Park, Y. E., Rademaker, R., & Tong, F. (2014). Both variations in perceptual sensitivity and decisional response bias contribute to visual working memory performance. *Journal of Vision*, *14*(10), 1375–1375. https://doi.org/10.1167/14.10.1375

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*(6), 1282–1289. https://doi.org/10.3758/s13423-013-0460-x

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290. https://doi.org/10.3758/BF03203943

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, *1*(1–12), 125–126. https://doi.org/10.3758/BF03342823

Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017b). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 6–17. https://doi.org/10.1037/xhp0000302

Rademaker, R. L., Bloem, I. M., De Weerd, P., & Sack, A. T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1650.

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*(13), 21. https://doi.org/10.1167/12.13.21

Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America A*, *2*(4), 619–621. https://doi.org/10.1364/JOSAA.2.000619

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–330. https://doi.org/10.3758/s13423-011-0055-3

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, *15*(1), 215–222. https://doi.org/10.3758/PBR.15.1.215

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Shaw, M. L., & others. (1980). Identifying attentional and decision-making components in information processing. *Attention and Performance VIII*, *8*, 277–295.

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, *98*(4), 1935–1940. https://doi.org/10.1073/pnas.98.4.1935

Simons, D. J., Chabris, C. F., Schnur, T., & Levin, D. T. (2002). Evidence for Preserved Representations in Change Blindness. *Consciousness and Cognition*, *11*(1), 78–97. https://doi.org/10.1006/ccog.2001.0533

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are There Multiple Visual Short-Term Memory Stores? *PLoS ONE*, *3*(2), e1699. https://doi.org/10.1371/journal.pone.0001699

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, 1–22. https://doi.org/10.3758/s13414-016-1108-5

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29.

Tversky, T., Geisler, W. S., & Perry, J. S. (2004). Contour grouping: closure effects are explained by good continuation and proximity. *Vision Research*, *44*(24), 2769–2777. https://doi.org/10.1016/j.visres.2004.06.011

Uemura, T., Sakaguchi, G., Ito, T., Okazawa, K., & Sakai, S. (1975). Experimental diarrhea in cynomolgus monkeys by oral administration with Clostridium perfringens type A viable cells or enterotoxin. *Japanese Journal of Medical Science & Biology*, *28*(3), 165–177.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/http://dx.doi.org/10.1037/a0035234

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114. https://doi.org/http://dx.doi.org/10.1037/0096-1523.27.1.92

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/http://dx.doi.org/10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, *138*(6), 1218–1252. https://doi.org/http://dx.doi.org/10.1037/a0029334

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience*, *14*(10), 1243–1244. https://doi.org/10.1038/nn.2910

Wertheimer, M. (1938c). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–94). London, England: Routledge & Kegan Paul. (Original work published 1923)

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, *10*(1), 80–87. https://doi.org/10.3758/BF03196470

Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, *64*(8), 1260–1280. https://doi.org/10.3758/BF03194770

Xu, Y. (2006). Understanding the object benefit in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, *68*(5), 815–828. https://doi.org/10.3758/BF03193704

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(47), 18766–18771. https://doi.org/10.1073/pnas.0705618104

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

Zhang, W., & Luck, S. J. (2009). Sudden Death and Gradual Decay in Visual Working Memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

Zhang, W., & Luck, S. J. (2011). The Number and Quality of Representations in Working Memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006