Research Article

Psychological Science
2017, Vol. 28(1) 12–22

DOI: 10.1177/0956797616671524
journals.sagepub.com/home/pss

The Whole Warps the Sum of Its Parts: Gestalt-Defined-Group Mean Size Biases Memory for Individual Objects

Jennifer E. Corbett
Aysel Sabuncu Brain Research Center, Bilkent University

Abstract

The efficiency of averaging properties of sets without encoding redundant details is analogous to gestalt proposals that perception is parsimoniously organized as a function of recurrent order in the world. This similarity suggests that grouping and averaging are part of a broader set of strategies allowing the visual system to circumvent capacity limitations. To examine how gestalt grouping affects the manner in which information is averaged and remembered, I compared the error in observers' adjustments of remembered sizes of individual circles in two different mean-size sets defined by similarity, proximity, connectedness, or a common region. Overall, errors were more similar within the same gestalt-defined groups than between different gestalt-defined groups, such that the remembered sizes of individual circles were biased toward the mean size of their respective gestalt-defined groups. These results imply that gestalt grouping facilitates perceptual averaging to minimize the error with which individual items are encoded, thereby optimizing the efficiency of visual short-term memory.

Keywords

perceptual averaging, gestalt grouping, summary statistics, visual short-term memory, open data

Received 3/13/16; Revision accepted 9/2/16

How do observers internally represent the external environment? The visual system can explicitly encode only a fraction of what is visible in each glance, and even then, the resultant internal representations are far from faithful to the true state of the external environment. Yet somehow, people get the gist and perceive the world as stable and complete. How does the brain achieve this illusion of perceived order amid continuous sensory chaos?

The understanding of how the brain represents objects has grown exponentially since the monumental discovery of visual neurons tuned to physical stimulus properties (Hubel & Wiesel, 1959). Researchers have classified massive interconnected cortical networks with unique global response patterns for different types of objects (Haxby et al., 2001), decoded details of objects in visual short-term memory (VSTM) from cortical activity (Harrison & Tong, 2009), and even "read out" neural patterns that indicate what object a person sees before the person is explicitly aware of the object (Thorpe, Fize, & Marlot, 1996). Recent landmark advances in computational power have allowed for the design of systems that can recreate the activity of networks of billions of neurons in response to visual stimuli on the basis of general principles abstracted from the visual system's structure and function (e.g., Itti & Koch, 2001; Serre, Oliva, & Poggio, 2007). Such powerful algorithms can take all the information available in an image as input in parallel. However, the brain is not a computer. There is overwhelming evidence that it cannot internally reproduce a one-to-one mapping of the external environment. For example, people are often "blind" to salient changes right in front of them (e.g., Simons & Chabris, 1999). In fact, it is widely accepted that attention is necessary to see change (Rensink, O'Regan, & Clark, 1997), yet people can attend only a few objects at once (Luck & Vogel, 1997). Therefore, vision researchers across science and technology are challenged to bridge the growing gap

Corresponding Author:
Jennifer E. Corbett, Aysel Sabuncu Brain Research Center, Room SC-205, Bilkent University, Ankara Çankaya 06800, Turkey
E-mail: jennifer.e.corbett@gmail.com

between overwhelming progress in understanding how humans internally represent objects and the underdeveloped knowledge of how the brain uses the remaining majority of information that is not explicitly encoded.

Instead of ignoring, suppressing, or discarding the mass of information that escapes focused attention, a number of recent findings converge to suggest that the visual system circumvents capacity limitations, in part by relying on a default set of heuristics that are based on regularities in the external environment. These heuristics guide the formation of initial perceptual chunks that pragmatically constrain further processing. Toward advancing our understanding of how the limited-capacity visual system allows for such amazing perceptual capabilities, the present investigation examined how several such constraints determine what observers remember.

An ever-growing body of research suggests that *perceptual averaging*, an efficient means of statistically compressing redundant information in sets of objects, is a promising strategy for coping with the visual system's limited representational capacity (Alvarez, 2011; Ariely, 2001). For example, average properties are represented even when individual items are not (e.g., size; Corbett & Oriet, 2011) and are encoded as basic visual dimensions (Corbett, Wurnitsch, Schwartz, & Whitney, 2012); representations of average properties underlie the ability to maintain stable perception while interacting with the external environment (Corbett & Melcher, 2014a, 2014b; Corbett & Song, 2014). The efficiency of summarizing average properties of sets of similar objects without encoding redundant details bears many similarities to previous gestalt proposals that perception is parsimoniously organized as a function of recurrent order in the physical world (Wertheimer, 1925/1938). There is mounting evidence that grouping and averaging are part of a broader set of strategies the human visual system has developed to alleviate representational capacity limitations. For example, recent findings suggest that gestalt grouping facilitates perceptual averaging (Im & Chong, 2014) and reduces the neural resources needed to maintain sets of items in VSTM (Peterson, Gözenman, Arciniega, & Berryhill, 2015; Xu & Chun, 2007). These findings raise the intriguing and testable hypothesis that the visual system efficiently represents environmental regularities by using gestalt heuristics to group and statistically compress information.

The idea that information about sets of objects can be remembered en masse stands in sharp contrast to traditional slot models of memory (e.g., Cowan, 2001; Pashler, 1988), which posit a finite VSTM capacity of about four items. However, a number of recent findings contest this strict-capacity view and suggest that observers rely on higher-order structure to remember summary representations of sets of multiple objects (e.g., Brady & Tenenbaum, 2013). Therefore, the present investigation of whether gestalt grouping facilitates the manner in which observers rely on statistical representations will help to further elucidate the nature of VSTM capacity.

Measuring the error in observers' recollections of individual object properties allows for an assessment of the format of structured memory representations. For example, Lew and Vul (2015) reported that observers' memories of the locations of objects arranged in different spatial clusters were biased toward the respective clusters' centers. Participants studied displays of objects arranged in different spatial groupings for several seconds. When they were shown a subsequent display with the objects at the bottom of the screen and asked to place the objects in their corresponding locations from the previous study display, the magnitude of the observers' recall errors for objects within the same clusters were more similar than would be expected if the objects' locations were independently coded. Brady and Alvarez (2011) reported a similar bias for the remembered sizes of individual circles. Observers were shown displays of red, blue, and green circles of various sizes, followed by a single black test circle. When observers were asked to adjust the size of the test circle to match the size of the circle that had occupied the same location in the previous display, they tended to make larger adjustments when other circles of the same color (e.g., blue) were large than when other circles of the same color (e.g., red) were small.

The present investigation used a hybrid of Lew and Vul's (2015) error-similarity paradigm and Brady and Alvarez's (2011) adjustment task to quantify how gestalt grouping affects the manner in which information is averaged and remembered. Specifically, I measured the error in observers' adjustments of the remembered sizes of individual circles in two different mean-size sets that were presented for two different durations (500 ms or 5 s) and defined by four different gestalt principles of grouping (similarity, proximity, connectedness, and common region). To the extent that grouping facilitates averaging, the magnitudes of participants' errors would be expected to be more correlated for circles in the same gestalt-defined sets, and participants should recall individual circles' sizes with bias toward respective set means. To the extent that grouping and averaging increase VSTM efficiency, this mean-size bias would be expected to minimize the overall error in individual estimates.

Method

Participants

Thirty-three students from Bilkent University (mean age = 21.61 years, age range = 19–29; 22 female), all with normal or corrected-to-normal vision, voluntarily participated in the main experiment either for course credit or for money. A minimum sample size of 30 participants was chosen a priori on the basis of the sample sizes (ranging from 16 to 35 participants) used in the previous studies by Lew and Vul (2015) and Brady and Alvarez (2011). All procedures and protocols were in accordance with the guidelines of Bilkent University's ethical review board.

Task

On each trial, participants viewed a study display of 16 circles of different sizes followed by a test display of six circles of the same size. Their task was to use the computer mouse to select a circle in the test display. They then used the "↑" and "↓" keys on the computer keyboard to increase or decrease the size of the test circle until it matched the remembered size of the circle in the corresponding location in the study display. They performed this task for each of the six text circles, one after another.

Apparatus

A Dell PC presented the stimuli against a gray (midway between black and white) background on a 41- × 23-cm Samsung LCD monitor with a 60-Hz refresh rate and a resolution of 1,366 × 768 pixels. Participants were seated with their heads centered approximately 57 cm from the middle of the screen, such that 1° of visual angle corresponded to approximately 34 pixels. MATLAB (Version 2015a; The MathWorks, Natick, MA) and Psychophysics Toolbox (Version 3; Brainard, 1997; Pelli, 1997) controlled all the presentation, timing, and response functions.

Stimuli

Study displays. Each study display was composed of two sets of 8 circles of various sizes, for a total of 16 circles per study display. The small-mean-size set contained 8 circles ranging in diameter from 0.5° to 1.2° in 0.1° steps (mean size = 0.85° of visual angle); the large-mean-size set contained 8 circles ranging in diameter from 0.9° to 1.6° in 0.1° steps (mean size = 1.25° of visual angle). Note that because these ranges overlapped, 4 of the circles in each set had a counterpart of the same size in the other set. The 16 circles were positioned pseudorandomly inside an imaginary 4 × 4 square grid, subtending approximately 12° of visual angle, in the center of the screen. The positions of the individual circles in each set were randomized on every trial within the gestalt grouping constraints (see the next four sections), such that no location consistently contained a circle that was larger or smaller than any other circle in the set; only the mean sizes of the sets and the individual sizes of the circles in each set remained constant over the course of the experiment. On each trial, the small-mean-size and large-mean-sets were grouped according to one of four gestalt heuristics: similarity, proximity, connectedness, or common region (Fig. 1):

- *Similarity.* When the circles were grouped by similarity, the sets were arranged by alternating the light-gray and dark-gray color of neighboring circles. The circles in the top leftmost and bottom rightmost positions were always the same color (i.e., in the same set). The small-mean-size set was randomly assigned either the dark-gray or the light-gray color (corresponding to 75% and 25% of the monitor's white-to-black contrast range, respectively) on each trial, and the large-mean-size set was assigned the opposite color. In the example shown in Figure 1 (top row, left-most panel), the small-mean-size set is presented in dark gray.
- *Proximity.* When the circles were grouped according to proximity, the sets were arranged in either rows or columns. The small-mean-size set was randomly assigned to the top or bottom of the screen (for rows) or to the left or right side of the screen (for columns), and the large-mean-size set was assigned to the opposite location. Each set was shifted away from the opposite set, so that the distance between the innermost rows or columns (from circle center to circle center) was 3.8°. The centers of the circles in each set were separated from each other by 3° horizontally and 3° vertically. All of the circles were randomly assigned one color (black or white) on each trial. In the example shown in Figure 1 (top row, second from left), the circles are grouped in columns, and the small-mean-size set is presented on the left.
- *Connectedness.* When the sets were grouped by the gestalt heuristic of connectedness, a 0.05° wide line was extended from circle to circle to form two roughly rectangular arrangements on either the top and bottom (rows) or the left and right sides (columns) of the screen. The small-mean-size set was randomly assigned to the top or bottom of the screen (for rows) or to the left or right side of the screen (for columns), and the large-mean-size set was assigned to the opposite location. All of the circles and the lines connecting them were randomly assigned one color (black or white) on each trial. In the example shown in Figure 1 (top row, second from right), the circles are grouped in rows, and the small-mean-size set is presented on the bottom.
- *Common region.* When the circles were grouped according to the gestalt heuristic of common

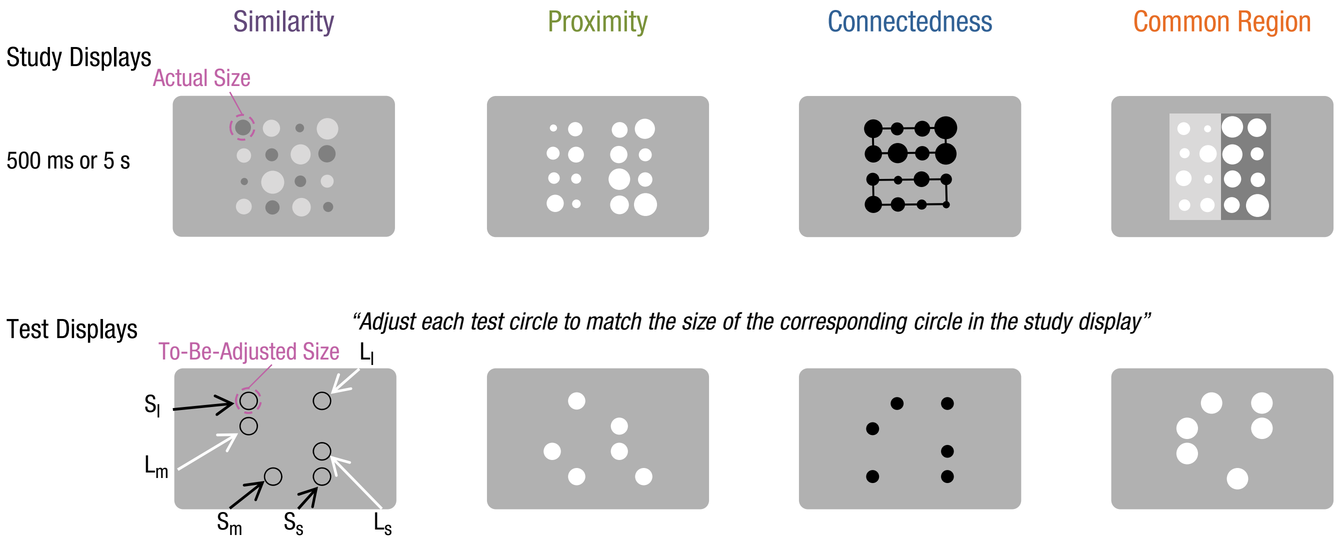

Fig. 1. Experimental stimuli and procedure. Study displays (top row) contained two sets of eight circles of various sizes. One set had a small mean circle size, and the other had a large mean circle size. The sets appeared in one of four gestalt groupings (similarity, proximity, connectedness, and common region) and were displayed for 500 ms or 5 s. For details, see the Study Displays section. After viewing the study displays, the participants viewed test displays (bottom row) in which they adjusted six test circles of the same size (three from each gestalt-defined group) to match their memory of the sizes of the circles in those locations in the preceding study displays. For details, see the Test Displays section. In the leftmost panel of the bottom row, capital letters *S* and *L* refer to whether a test circle belongs to the small- or large-mean-size set; subscript lower-case letters *s, m*, and *l* refer to the relative small, medium, and large sizes of the circles within each set when they appeared in the preceding study display. The dashed circle in the top left study display indicates the actual size of an individual test circle, and the dashed circle in the bottom left test display indicates the corresponding circle to be adjusted by the participant. The magnitude of error for the individual circle was calculated as the circle’s adjusted size minus its actual size.

region, the sets were arranged in either rows or columns. The small-mean-size set was randomly assigned to the top or bottom of the screen (for rows) or to the left or right side of the screen (for columns), and the large-mean-size set was assigned to the opposite location. A dark-gray rectangle (56% of the monitor’s white-to-black contrast) was randomly assigned to one of the sets of circles, and a light-gray rectangle (44% of the monitor’s white-to-black contrast) was assigned to the other set. Each rectangle fully enclosed the circles of the set to which it was assigned. The imaginary line at which the two rectangles met was centered on the screen, either vertically (for rows) or horizontally (for columns). Each rectangle was 12° wide × 6° tall (for rows) or 6° wide × 12° tall (for columns). All of the circles were randomly assigned one color (black or white) on each trial. In the example shown in Figure 1 (top row, rightmost panel), the circles are grouped in columns, and the light-gray rectangle on the left encloses the small-mean-size set.

Test displays. Each test display was composed of six circles of the same size; on each trial, the circles’ size was randomly selected to correspond to the size of one of the 16 study circles. The locations of the six circles were randomly selected on each trial from among the 16 locations in the previous study display, three from each gestalt-defined set (Fig. 1). Each test circle was centered relative to the corresponding circle in the previous study display. For circles grouped by proximity, connectedness, and common region, test circles were the same color as in the previous study display. For circles grouped by similarity, test circles were presented as 2-pixel wide black outlines filled with the same medium-gray color as the global screen background.[1] The dark- and light-gray colors of the two groups of circles in the similarity conditions, the distance between the two groups in the proximity condition, the width of the lines connecting circles within each group in the connectedness condition, and the dark- and light-gray colors of the rectangles surrounding each group of circles in the common-region condition were selected on the basis of the results of earlier pilot studies that had tested a range of gestalt strengths.

Procedure

Participants initiated each trial by pressing the space bar. Immediately afterward, the study display of 16 circles was presented for either 500 ms or 5 s, followed by the test display of 6 circles, which remained on the screen until participants had finished adjusting all the test circles. A 0.4° red circle was presented simultaneously in a random position along with the six circles in the test display on each trial. Participants were instructed to use the mouse to move this red circle over the test circle they wished to adjust. When participants clicked the red mouse circle on any location inside a particular circle, the red circle disappeared and a 0.1° green circle appeared in the center of the given circle to indicate that it had been selected for adjustment. Participants then used the "↑" and "↓" keys on the computer keyboard to increase or decrease the size of the selected circle in steps of 2 pixels. When they had adjusted a given circle to match the remembered size of the corresponding circle in the previous study display, participants pressed the "Enter" key to continue to the next circle. The red circle reappeared at a random location within the display, and the participants selected a new circle to adjust.

Participants were informed that each circle could be adjusted only once and that they had to adjust every test circle before they would be able to proceed to the next trial. They were also instructed to respond as quickly and accurately as possible on each trial. Unbeknownst to participants, they were required to adjust each circle with a minimum of three presses of the arrow keys; this requirement was intended to guard against a tendency, observed in an earlier pilot study, for some participants to select individual circles and press enter without performing the required size adjustments. If participants did not adjust a given test circle to satisfy this criterion, they saw a message reading "Please adjust the circle more carefully" in the center of the display until the participant pressed one of the two arrow keys to continue adjusting the selected circle. If participants received this warning 10 times in a given block, the experimental session was terminated and their data were excluded from analysis.

Observers completed one practice block of 5 trials (excluded from further analysis), followed by four blocks of 32 trials each in the main experiment. Each block in the main experiment contained two repetitions of the 16 possible combinations of the four gestalt conditions (similarity, proximity, connectedness, and common region), the two duration conditions (500 ms and 5 s), and the two possible arrangements (rows or columns in the proximity, connectedness, and common-region conditions, or starting with a dark-gray circle or a light-gray circle in the top leftmost position in the similarity condition), presented in pseudorandom order, for a total of 128 trials per participant. Trials were collapsed across the two types of arrangements, for a total of 16 trials per point in each of the eight possible combinations of the gestalt and duration conditions of interest in three main analyses: error similarity, mean-size bias, and VSTM efficiency.

Participants were shown written illustrated instructions on the computer screen in either English or Turkish at the beginning of the practice block and each experimental block, and all text displayed over the course of the experiment was also presented in the participant's language of choice (Turkish or English). The experimenter ensured that each participant fully understood the task before beginning the main experimental blocks. Six participants (not included in the earlier description of participants) were not able to complete the main experiment because they were warned 10 times in a single block to carefully adjust the test circles (these participants were stopped after their third attempt at the practice block). For participants who completed the experiment, the entire session lasted between 1 and 1.5 hours. All of these participants were given a short (5–10 min) compulsory break after completing the second experimental block.

Results

Error similarities

To compare the precision of the size adjustments within and between the two gestalt-defined mean-size groups in each display, I calculated participants' adjustment errors for each of the six test circles on each trial relative to the actual sizes of the corresponding circles in the preceding study display (adjusted size minus actual size). Specifically, for data from each trial, within each of the two gestalt-defined groups of circles, the actual sizes of the three test circles were ordered from smallest to largest (Fig. 1, bottom row, leftmost panel), and the corresponding adjustment errors were calculated. For the eight combinations of the gestalt and duration conditions, each subject's average error similarities—that is, the average sample Pearson correlations (*r*) between errors—for circles within the same gestalt-defined small- or large-mean-size groups and between different gestalt-defined groups[2] were calculated from the individual trial errors (*i*) for a given pair of the six test sizes (*x,y*) using the following formula:

$$\text{error similarity} = r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}},$$

where $\bar{x}$ and $\bar{y}$ are the respective sample mean errors for a given pair of test-circle sizes (e.g., $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$).

Error similarity was greater for circles within the same gestalt-defined groups than for circles within different gestalt-defined groups. A 2 (gestalt-defined group: same, different) × 4 (gestalt condition: similarity, proximity, connectedness, common region) × 2 (duration condition:

500 ms, 5 s) repeated measures within-subjects omnibus analysis of variance (ANOVA) on error similarities provided initial support for this conclusion, revealing significant main effects of gestalt-defined group and gestalt condition (both *p*s < .001). It was surprising that the main effect of duration condition (500 ms, 5 s) was not significant, suggesting that size was encoded relative to gestalt-defined group regardless of whether participants viewed displays very briefly (i.e., for half a second) or much longer (i.e., for 5 s). No significant two- or three-way interactions were observed. For the full results of this omnibus ANOVA, see Table S1a in the Supplemental Material available online.

The main series of eight planned comparisons ($\alpha = .05$; $p = .03$, adjusted for false-discovery rate, or FDR) examined participants' average error similarities within the same gestalt-defined groups relative to between different gestalt-defined groups. The results revealed significantly greater correlations within the same gestalt-defined groups than across different gestalt-defined groups for each combination of the gestalt and duration conditions, except for the 500-ms connectedness condition (Fig. 2a; for the full results of the eight planned comparisons, see Table S1b in the Supplemental Material). Importantly, follow-up one-sample *t* tests ($\alpha = .05$; $p = .02$, FDR-adjusted) confirmed that participants' error similarities for circles within the same groups were significantly correlated (i.e., the coefficient was greater than 0), supporting the proposal that individual circles' sizes were recalled relative to their respective gestalt-defined groups in all combinations of the gestalt and duration conditions. However, this was the case for only half of the corresponding error similarities between circles in different groups, such that there were no significant differences for the similarity or connectedness conditions at either duration (for the full results of both sets of one-sample *t* tests, see Table S1c in the Supplemental Material).

Mean-size bias

As outlined in the introduction, Brady and Alvarez (2011) found that the same circle was remembered as smaller or larger depending on whether other circles in the same color in the preceding display were small or large, respectively. To directly examine whether such bias toward the gestalt-defined-group mean size was an underlying influence on observers' recollections of individual objects' sizes, I compared their adjustments of physically identical test circles when presented in the small-mean-size gestalt-defined group and when presented in the large-mean-size gestalt-defined group. Recall that four of the circles in each mean-size set had a counterpart of the same size (i.e., 0.9°, 1.0°, 1.1°, and 1.2°) in the other set. Therefore, although the circles tested from each small- and large-mean-size gestalt-defined group were randomly determined on every trial, each of these four sizes had an equal probability (~12.5% of trials) of appearing as a test circle in either group in each combination of the gestalt and duration conditions.

An initial 2 (gestalt-defined group: small mean size, large mean size) × 4 (gestalt condition: similarity, proximity, connectedness, common region) × 2 (duration condition: 500 ms, 5 s) repeated measures within-subjects omnibus ANOVA was conducted to investigate whether participants recalled identically sized test circles as a function of their membership in a small-mean-size or large-mean-size gestalt-defined group. This analysis revealed a significant main effect of mean size of the gestalt-defined group, a significant main effect of gestalt condition, a significant interaction between mean size of the gestalt-defined group and gestalt condition, and a significant three-way interaction (all *p*s < .05; for full ANOVA results, see Table S2a in the Supplemental Material).

The main series of eight planned comparisons ($\alpha = .05$; $p = .002$, FDR-adjusted) confirmed significant differences between the adjusted sizes of identical test circles for the 500-ms proximity condition and for the 5-s proximity, connectedness, and common-region conditions. Participants' adjustments of identically sized test circles were larger when the test circles were presented as part of large-mean-size gestalt-defined groups than when they were presented within small-mean-size gestalt-defined groups (Fig. 2b; Table S2b in the Supplemental Material available online presents the full results of the eight planned comparisons).

VSTM efficiency

Comparing the magnitudes of errors in participants' adjustments of the individual circles relative to the gestalt-defined-group mean size with their errors relative to the corresponding actual sizes in the study display allowed for an explicit test of whether gestalt-defined group mean-size bias minimized the magnitude of error with which individual circles were encoded. If mean-size bias increases VSTM efficiency, then adjustments of individual test circles should show lower root-mean-square errors (RMSE) when calculated from the mean size of the entire gestalt-defined group (adjusted size minus group-mean size) than when calculated using the actual sizes corresponding to the individual test circles (adjusted size minus actual size).

An initial 2 (error type: adjusted size minus actual size, adjusted size minus group-mean size) × 4 (gestalt condition: similarity, proximity, connectedness, common region) × 2 (duration condition: 500 ms, 5 s) repeated measures within-subjects ANOVA on the two types of

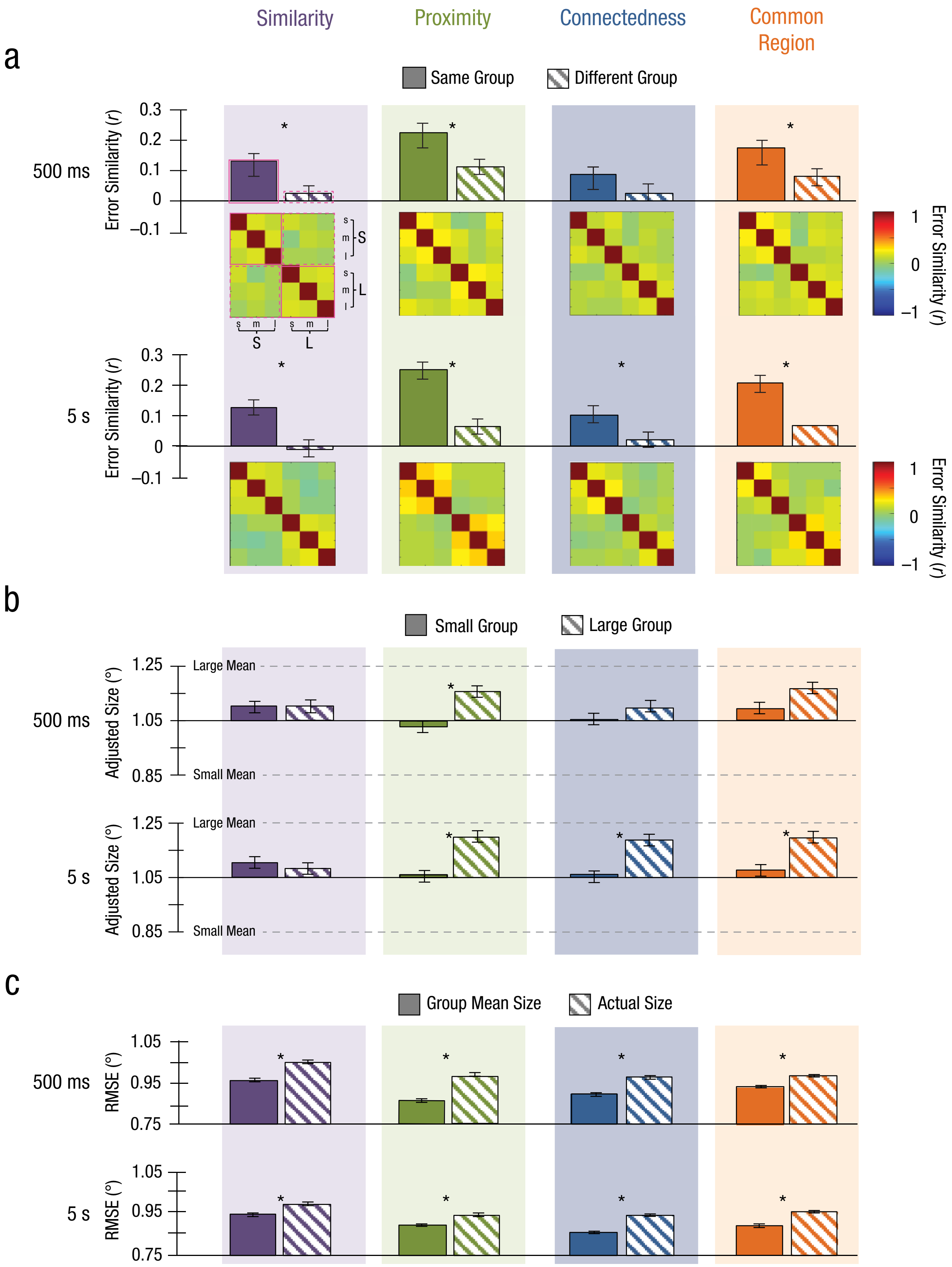

Fig. 2. Results of the three main analyses for each combination of the gestalt and duration conditions (n = 33). The bar graphs in (a) show the mean error similarity (i.e., the mean correlation between adjustment errors) within the same gestalt-defined groups and

(continued on next page)

Fig. 2. *(continued)*

between different gestalt-defined groups. The heat maps under the *x*-axes show the mean error similarities for each combination of small, medium, and large test-circle sizes (*s, m*, and *l*) both within the same gestalt-defined mean-size groups (large or small) and between different gestalt-defined mean-size groups. The outlines in the top left graph and heat map highlight which results are for the same gestalt-defined groups (solid outlines) and which are for the different gestalt-defined groups (dashed outlines). The larger *S*s and *L*s on the edge of the heat maps indicate the gestalt-defined small-mean-size group and the gestalt-defined large-mean-size group, respectively. The graphs in (b) show the mean adjusted size of identically sized test circles when they were presented in small-mean-size groups and when they were presented in large-mean-size groups. The dashed lines indicate the mean sizes of the small (0.85°) and large (1.25°) gestalt-defined mean-size sets. The graphs in (c) show mean root-mean-square error (RMSE) for the adjusted size of test circles relative to the mean size of the circles in the corresponding gestalt-defined group in the preceding display (adjusted size minus group mean size). They also show RMSE for the adjusted size of test circles relative to the actual sizes of the circles in the preceding display (adjusted size minus actual size). Asterisks indicate false-discovery-rate-adjusted significant planned comparisons ($\alpha = .05$). Error bars indicate 95% within-subjects confidence intervals for the corresponding three-way interactions (Loftus & Masson, 1994).

RMSEs for participants' adjustments of the six circles on each trial revealed a significant main effect of error type, a significant main effect of gestalt condition, a significant main effect of duration condition, a significant interaction between error type and duration condition, and a significant three-way interaction (all *p*s $< .05$; for the full ANOVA results, see Table S3a in the Supplemental Material).

The main series of planned comparisons ($\alpha = .05$; $p = .001$, FDR-adjusted) was between participants' RMSEs calculated relative to the test circles' actual sizes and RMSEs calculated relative to the test circles' gestalt-defined-group mean size. Results confirmed that RMSEs calculated relative to gestalt-defined-group mean size were significantly lower than RMSEs calculated relative to actual size for each of the eight combinations of the gestalt and duration conditions (Fig. 2c; for the full results of the eight planned comparisons, see Table S3b in the Supplemental Material). These findings suggest that the mean size of the gestalt-defined group efficiently biased participants' memories of the individual circles' sizes to minimize the error with which individual circles were encoded.

Discussion

The present results converge in support of the proposal that gestalt grouping facilitates perceptual averaging to optimize VSTM efficiency. Error similarity was greater for circles within the same gestalt-defined groups than for circles within different gestalt-defined groups. In line with previous findings (Brady & Alvarez, 2011), adjustments of physically identical test circles were smaller for circles presented in the gestalt-defined small-mean-size sets than for circles presented in large-mean-size sets for the 500-ms proximity condition and the 5-s proximity, connectedness, and common-region conditions. Furthermore, the RMSE for adjustments relative to gestalt-defined-group mean size was significantly lower than the RMSE for adjustments relative to the actual sizes of the individual circles for each of the eight combinations of the gestalt and duration conditions; this finding supports the proposal that mean-size bias reduced the error with which participants remembered the sizes of individual test circles. Recall that half of the circles in each gestalt-defined mean-size set had a counterpart of the same size in the other set. Therefore, participants could not rely solely on the two different mean sizes in each display; instead, their memories of individual sizes were warped by the mean size of the corresponding gestalt-defined groups.

The present effects of gestalt condition and duration condition can be interpreted within the context of a number of previous findings. For example, consistent effects of proximity on participants' error similarities and adjustments of identically sized test circles at both durations accord with various reports of proximity's dominant, even mandatory, influences (e.g., Han, 2004; Peterson & Berryhill, 2013; Rock & Palmer, 1990; Xu, 2006). Significant effects of gestalt-defined-group mean size on participants' adjustments of identically sized test circles in the connectedness and common-region conditions at 5 s, but not at 500 ms, are consistent with recent reports that event-related-potential indices of VSTM benefits resulting from grouping by connectedness (Peterson et al., 2015) and common region (Montoro et al., 2015) emerge during later stages of processing. The lack of significant effects of gestalt-defined-group mean size in the similarity condition at either duration, despite significantly correlated and more similar errors within the same similarity-defined groups relative to between different similarity-defined groups at both durations, echoes discrepant findings regarding the influence of similarity in perception and VSTM. Similarity is unique in that it is not necessarily spatially constrained: It can involve items distributed over an entire display. Several results from change-detection studies have led to the proposal that grouping by similarity requires proximity (e.g., Jiang, Chun, & Olson, 2004; Peterson & Berryhill, 2013). However, other results pointing to an influence of similarity only during later, downstream stages of processing driven by top-down feedback (e.g., Han, 2004; Han, Jiang, Mao, Humphreys, & Gu, 2005; cf. Kubovy & van den Berg,

2008) suggest that grouping by similarity may require more feature and space binding and therefore more attention than other more spatially constrained heuristics. Such proposals are in line with findings (Brady & Alvarez, 2011) that observers did not group items by similarity unless it was task relevant when they had to ignore a third set of items while preparing to recall the size of a single item that could be presented in one of two attended sets of three items each. Overall, the current findings provide evidence not only that gestalt grouping facilitates perceptual averaging, but also that the different heuristics tested in the present investigation may have unique, predictive effects on the manner in which information is economically encoded in VSTM. Future research parametrically varying and thresholding the different strengths of gestalt grouping cues can help to uncover their characteristic effects on the manner in which information is encoded and recalled.

The present results cannot be explained by fixed-capacity models that assume objects are encoded independently in VSTM (e.g., Bays & Husain, 2008; Cowan, 2001; Pashler, 1988); models in which the statistical structure of information is exploited are more likely to be accurate. For example, the present findings extend proposals that the visual system uses statistical regularities to compress covariant information into more efficient *chunks* (e.g., Brady, Konkle, & Alvarez, 2009) by suggesting gestalt-grouping principles as a plausible basis for what constitutes a chunk. The current results can also be interpreted within the context of (a) probabilistic hierarchical-encoding models in which noisy samples of individual objects are remembered as a function of the observer's expectations about other items sharing the same properties (e.g., Brady & Alvarez, 2011) and (b) models that account for the statistical dependencies between individual objects by inferring probability distributions over different clusters of information (e.g., Brady & Alvarez, 2011; Brady & Tenenbaum, 2013; Orhan & Jacobs, 2013) to allow for the representation of visual information at multiple scales simultaneously. It has repeatedly been suggested that statistical descriptions of sets of similar items taken with a small set of explicit object files may underlie humans' remarkable perceptual abilities despite the visual system's limited capacity (e.g., Treisman, 2006). This suggestion is in line with findings (e.g., Hyde & Wood, 2011) that attention may be a key factor in determining whether information is encoded in individual object files (Kahneman & Treisman, 1984) or global statistical approximations.

The present study is the first to report effects of similarity, proximity, connectedness, and common region on memory representations within the same experiment, using the same paradigm. In addition to extending proposals that gestalt grouping reduces the amount of resources needed to represent information to a wider range of heuristics, the present results point to a plausible source for this reduction: Grouping facilitates perceptual averaging as part of a broader strategy that the visual system relies on to minimize the variance in memory representations for efficient, pragmatic encoding of redundant information. Building on the proposal from Im and Chong (2014) that mean size is a unit of VSTM, the present results suggest that mean size is an emergent property of gestalt grouping, such that individual item properties are represented as functions of the statistical properties of the entire set. This representational warping allows observers to remember more information about multiple objects, such that "the properties of the parts are determined by the intrinsic structural laws of the whole" (Wertheimer, 1925/1938, p. 7).

Action Editor

John Jonides served as action editor for this article.

Author Contributions

J. E. Corbett is the sole author of this article and is responsible for its content.

Acknowledgments

I thank Ezgi Ersen, Fenise Kayhan, Miray Elif Şakar, and Ayşenur Yılmaz for recruiting and testing participants, and Jaap Munneke for his insightful comments on drafts of the manuscript.

Declaration of Conflicting Interests

The author declared that she had no conflicts of interest with respect to her authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/4492y. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Data. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

Notes

1. Note that although the colors of the study and test circles were chosen to minimize interference from afterimages, afterimages were unlikely to persist over the several seconds that

observers required to adjust all six test circles in each display. Furthermore, the positions of the test circles in each group were randomized on every trial.

2. Note that only one gestalt grouping principle was used on each trial. Consequently, the terms "same gestalt-defined groups" and "different gestalt-defined groups" are used to refer to membership in the two groups formed by that single gestalt principle and not to refer to the same or different gestalt conditions (similarity, proximity, connectedness, and common region).

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*, 122–131.

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

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

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

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

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*, 85–109.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Corbett, J. E., & Melcher, D. (2014a). Characterizing ensemble statistics: Mean size is represented across multiple frames of reference. *Attention, Perception, & Psychophysics*, *76*, 746–758.

Corbett, J. E., & Melcher, D. (2014b). Stable statistical representations facilitate visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1915–1925.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*, 289–301.

Corbett, J. E., & Song, J.-H. (2014). Statistical extraction affects visually guided action. *Visual Cognition*, *22*, 881–895.

Corbett, J. E., Wurnitsch, N., Schwartz, A., & Whitney, D. (2012). An aftereffect of adaptation to mean size. *Visual Cognition*, *20*, 211–231.

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

Han, S. (2004). Interactions between proximity and similarity grouping: An event-related brain potential study in humans. *Neuroscience Letters*, *367*, 40–43.

Han, S., Jiang, Y., Mao, L., Humphreys, G. W., & Gu, H. (2005). Attentional modulation of perceptual grouping in human visual cortex: Functional MRI studies. *Human Brain Mapping*, *25*, 424–432.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, *148*, 574–591.

Hyde, D. C., & Wood, J. N. (2011). Spatial attention determines the nature of nonverbal number representation. *Journal of Cognitive Neuroscience*, *23*, 2336–2351.

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

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*, 194–203.

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

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, D. R. Davies, & J. Beatty (Eds.), *Variants of attention* (pp. 29–61). New York, NY: Academic Press.

Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115*, 131–154.

Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, *15*(4), Article 10. doi:10.1167/15.4.10

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Montoro, P. R., Luna, D., Albert, J., Santaniello, G., López-Martín, S., Pozo, M. A., & Hinojosa, J. A. (2015). A temporo-spatial analysis of the neural correlates of extrinsic perceptual grouping in vision. *Neuropsychologia*, *69*, 118–129.

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

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, *20*, 1282–1289.

Peterson, D. J., Gözenman, F., Arciniega, H., & Berryhill, M. E. (2015). Contralateral delay activity tracks the influence of Gestalt grouping principles on active visual working memory representations. *Attention, Perception, & Psychophysics*, *77*, 2270–2283.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Rock, I., & Palmer, S. E. (1990, December). The legacy of Gestalt psychology. *Scientific American*, *263*(6), 84–90.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA*, *104*, 6424–6429.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*, 411–443.

Wertheimer, M. (1938). Gestalt theory. In W. D. Ellis (Ed.), *Source book of Gestalt psychology* (pp. 1–11). New York, NY: Harcourt, Brace. (Original work published 1925)

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

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences, USA*, *104*, 18766–18771.