Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–10

DOI: 10.1177/0956797618822798
www.psychologicalscience.org/PS
SAGE

Similarity Grouping as Feature-Based Selection

Dian Yu[1], Xiao Xiao[2], Douglas K. Bemis[3], and Steven L. Franconeri[1]

[1]Department of Psychology, Northwestern University; [2]Feinberg School of Medicine, Northwestern University; and [3]Geometric Intelligence, New York, New York

Abstract

Across the natural world as well as the artificial worlds of maps, diagrams, and data visualizations, feature similarity (e.g., color and shape) links spatially separate areas into sets. Despite a century of study, it is yet unclear what mechanism underlies this gestalt similarity grouping. One recent proposal is that similarity grouping—for example, seeing a red, vertical, or square group—is just global selection of those features. Although parsimonious, this account makes the counterintuitive prediction that similarity grouping is strictly serial: A green group cannot be constructed at the same time as a red group. We tested this prediction with a novel measure—a grouping illusion within number-estimation tasks that should work only if participants simultaneously construct groups—and found the strongest evidence yet in favor of serial feature-based attention (*N*s = 14, 12, and 12 for Experiment 1, Experiment 2, and Experiment 3, respectively).

Keywords

grouping, perceptual organization, visual attention, feature-based selection, number estimation, open data, open materials

Received 9/17/17; Revision accepted 10/3/18

Objects with similar features—colors, orientations, or shapes—are perceived as belonging together even when they are spatially distributed. Although grouping by feature similarity has been studied in gestalt psychology for more than a century (von Ehrenfels, 1890/1988), there is no consensus on its underlying mechanism. One proposal is that similarity groups are constructed through feature selection (Huang & Pashler, 2007; Levinthal & Franconeri, 2011). Feature selection amplifies representations of areas that contain certain colors, orientations, and shapes, even when they are spatially distributed (see Wolfe & Horowitz, 2004).

Although this account provides a parsimonious explanation for similarity grouping, it makes a counterintuitive prediction that similarity groups can be constructed only serially; when one selects for the color red on a display, one can see only a red group. This means that one cannot construct multiple similarity groups in the same perceptual instant. This prediction has been tested by seeking evidence showing that when people are asked to complete tasks that require grouping across a greater number of feature values (red, green, blue, etc.), a greater number of serial selections through feature space is required, which should take more time to complete. In contrast, if feature-similarity grouping occurs in parallel, increasing the number of feature values to group should not affect performance. Indeed, when people are asked to make symmetry judgments on patterns consisting of increasing diversities of colors, shapes, or orientations, their judgment takes longer as the number of feature values increases (Huang & Pashler, 2002). Even common-fate grouping could be explained by this mechanism if motion direction, like orientation, is assumed to be a feature space. Indeed, when participants are asked to search for a particular common-fate group among other common-fate groups (e.g., find a vertically oriented

Corresponding Author:
Dian Yu, Northwestern University, Department of Psychology, 2029 Sheridan Rd., Swift Hall, Evanston, IL 60208
E-mail: dianyu2017@u.northwestern.edu

common-fate group among horizontal ones), search time increases substantially as the number of groups increases (Levinthal & Franconeri, 2011).

Although such evidence is consistent with serial formation of similarity groups, it is open to the critique whether groups are actually created in parallel and whether the response-time increases are due to judgments required of those groups, such as discriminating the shape or size of each group (Trick & Enns, 1997) or other postselection and decision-stage processes (Palmer, 1995; Townsend, 1972). We sought a task that distills grouping closer to its essence—segregation of a visual scene into distinct units (Wagemans et al., 2012)—without overlaying additional requirements to judge the shape or size of those units.

Here, we provide the strongest support yet for the feature-based-selection account by showing evidence consistent with serial grouping in a task that should be immune to these critiques. Because the essence of grouping is that it segregates a visual scene into distinct units for other processes, number estimation can serve as a particularly pure measure of grouping; this is because perceived numerosity is a direct reflection of the number of distinct units (Burr & Ross, 2008). We relied on a strong grouping illusion found in number-estimation tasks, in which displays containing objects grouped by spatial-grouping cues (grouping objects into pairs via connecting lines, common regions, or proximity) are systematically underestimated compared with ungrouped displays (Franconeri, Bemis, & Alvarez, 2009; He, Zhang, Zhou, & Chen, 2009; He, Zhou, Zhou, He, & Chen, 2015).

If similarity grouping relies on serial selection of each feature value over time, it should not be possible to achieve such grouping in a number-estimation task, which requires broad and simultaneous selection of an entire collection. Indeed, although we replicated the underestimation effects for these spatial-grouping cues, we repeatedly found them absent for the strongest similarity-grouping cues (color, shape, and even redundantly grouped colored shapes) across multiple exposure durations and even after a procedure designed to equate the grouping strength of the spatial and similarity cues. These results are consistent with the parsimonious claim that feature-based attention underlies gestalt similarity grouping.

Experiment 1 replicated the underestimation effect with three spatial-grouping cues—connectivity, common region, and proximity—but revealed no effect for three feature-similarity-grouping cues—color, shape, and orientation. Experiment 2 replicated the underestimation effect for common region (the relatively weakest of the spatial-grouping cues) but found no effect for the combination of two similarity cues (color and shape), despite the fact that combined cues produce stronger grouping than individual cues on their own (Nothelfer, Gleicher, & Franconeri, 2017). These results remained robust even after we doubled the presentation durations of both displays to ensure that feature-similarity groups could have sufficient time to form. In Experiment 3, we again replicated these effects after using an independent task to match the perceptual-grouping strength of proximity grouping and redundant color-shape similarity grouping.

Experiment 1

Method

Participants. Fourteen Northwestern University undergraduates and members of the Evanston, Illinois, community participated in exchange for $5 or course credit. The number of participants was specified a priori on the basis of previous pilot experiments and results from a previous study using a similar experimental paradigm (Franconeri et al., 2009). All participants were required to have normal or corrected-to-normal visual acuity and normal color vision.

Apparatus. We used MATLAB (The MathWorks, Natick, MA) with the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) on an Apple Mac Mini running the OS X operating system. Displays were 19-in. LED monitors with a resolution of 1,280 × 1,024 pixels. Participants' head movements were not restrained, but at an average viewing distance of 60 cm, the screen area subtended approximately 35.2° × 26.9° of visual angle.

Design. We measured number-estimation performance for displays that were grouped by one of six cues. There were three spatial-grouping cues (proximity, connectedness, and common region) and three similarity-grouping cues (color, shape, and orientation). In the grouped condition, participants compared the number of objects on a grouped display with the number of objects on an ungrouped display (see example stimuli in Fig. 1). Grouped displays always contained 24 objects, and ungrouped displays contained a variable number of objects. We used a staircase procedure to find the perceptually equivalent number of objects as in the grouped display. Each time participants indicated that the grouped display contained more items, we increased the number of items on the ungrouped display by 1. Each time participants indicated that the grouped display contained less items, we decreased the number of items on the ungrouped display by 1. Each staircase ended after 20 reversals. In the ungrouped (control) condition, participants compared

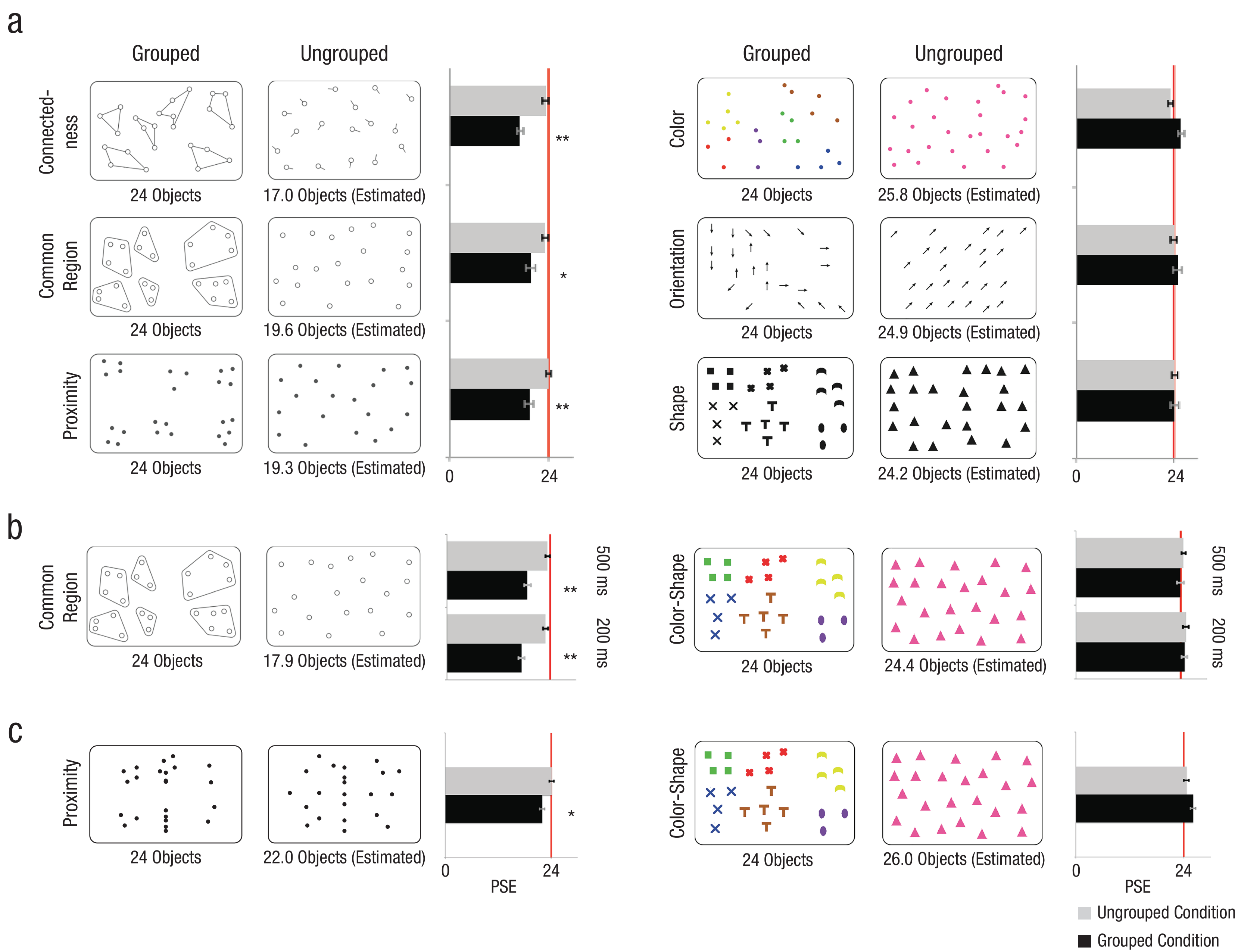

Fig. 1. Experimental displays and results from (a) Experiment 1 ($N = 14$), (b) Experiment 2 ($N = 12$), and (c) Experiment 3 ($N = 12$). Grouped displays were created using three spatial-grouping cues (proximity, connectedness, and common region) and three similarity-grouping cues (color, shape, and orientation). Experiment 1 contained all six grouped displays, Experiment 2 contained two grouped displays (common region and a combination of color and shape), and Experiment 3 contained two grouped displays (proximity and the color-shape combination). Each grouped display was matched with an ungrouped display. The graphs show the point of subjective equality (PSE) in the grouped and ungrouped conditions. Asterisks indicate significant differences between PSEs and the accurate estimate (24) as indicated by the red line (*$p < .05$, **$p < .01$). Error bars show within-subject standard errors.

the number of objects on two ungrouped displays. One of the ungrouped displays always contained 24 objects, whereas the other contained a variable number. We used the same staircase procedure as described above to find the perceptually equivalent number of objects as in the ungrouped display. This ungrouped (control) condition was designed to find participants' baseline number-estimation ability. We compared the perceptually equivalent number for the grouped display with this baseline as well as with the true value (24) to determine whether participants showed a number-estimation bias for the grouped displays.

Stimuli. We created six types of stimuli demonstrating six types of grouping conditions as well as their corresponding ungrouped control stimuli (for an illustration, see Fig. 1a). The grouped displays always contained 24 items, and the initial number of items on the ungrouped displays ranged from 14 to 22 for half of the participants and from 26 to 34 for the other half of the participants. All items were drawn within a rectangular area of 26.8° × 16.0° against a gray background (RGB value: 80, 80, 80).

Proximity-grouped displays contained six clusters of circles (each 1.0° in diameter) arranged in a 3 × 2 matrix, with 3 to 5 circles in each cluster. Corresponding control displays consisted of circles that were semirandomly located within the same area. (A detailed description of all displays is available in the Supplemental Material available online.) Connectedness-grouped displays contained outline circles semirandomly located within the same area. Twenty-four circles were connected to form six nonoverlapping polygons, with 3 to 5 circles serving as vertices of each polygon. Corresponding control displays contained circles arranged in the same layout but without connections. Common-region-grouped displays contained outline circles arranged in the same manner as the connectedness-grouped displays, with 24 circles enclosed by six nonoverlapping bubbles and 3 to 5 circles as internal vertices of each bubble. The corresponding control displays contained circles arranged in the same way but without bubble enclosures.

In the similarity-grouping conditions, item positions for both grouped and control displays were set in the same manner as the control displays for proximity grouping. Each display contained six spatially localized groups of items: circles (1° in diameter) for color grouping, shapes (1.4° × 1.4°; triangle, clover, oval, square, T shape, and X shape) for shape grouping, and arrows (1.3° × 1.7°; up, down, left, right, and the four diagonals between them) for orientation grouping. The set of colors consisted of red (RGB value: 190, 0, 0), green (RGB value: 49, 126, 0), blue (RGB value: 0, 97, 168), orange (RGB value: 165, 84, 0), purple (RGB value: 130, 0, 169), brown (RGB value: 124, 111, 0), and cobalt (RGB value: 0, 48, 145). On grouped displays, neighboring items were of the same color, shape, or orientation, creating six unique spatially localized color, shape, or orientation groups with three to five items forming each group. Ungrouped control displays contained a homogeneous set of objects of a feature value not present in the grouped display.

Procedure. Each trial started with a 500-ms fixation screen, followed by two successive displays lasting 200 ms each, with a 400-ms blank screen in between. After the second display, the screen went blank for another 400-ms, and then a question screen appeared. This screen instructed participants to press "1" or "2" on the keyboard to indicate which display contained more items—the first or the second. Figure 2 shows the process of a typical trial.

For half of the trials (the grouped condition), one display showed items grouped by one of the six grouping cues, whereas the other display showed its corresponding ungrouped control display. The other half of the trials (the control condition) used two displays of ungrouped objects to generate a baseline estimate of numerical-comparison precision in displays without grouping manipulations. The order of display types was always randomly determined on each trial. Conditions and their associated staircases were randomly interleaved.

Results

The perceptually equivalent number of items in the grouped displays versus ungrouped displays was calculated by taking the average of the last 5 of 20 total staircase reversals (set a priori). We also calculated the perceptually equivalent number of items for the ungrouped displays (the display that varied, in contrast to the display that was set to 24) to determine the baseline performance of number estimation for the six types of displays. Grouping items by connectedness, proximity, and common region robustly biased participants to underestimate their quantity. Points of subjective equality (PSEs) for these conditions are 17.0 (SE = 0.9) for connectedness and 23.2 (SE = 0.5) for its control condition, $t(13) = 6.09$, $p < .001$, Cohen's $d = 1.63$, 95% confidence interval (CI) = [0.80, 2.43]; 19.3 (SE = 1.2) for proximity and 23.9 (SE = 0.7) for its control condition, $t(13) = 4.15$, $p = .001$, Cohen's $d = 1.11$, 95% CI = [0.42, 1.77]; and 19.6 (SE = 0.8) for common-region grouping and 23.1 (SE = 0.6) for its control condition, $t(13) = 2.80$, $p = .015$, Cohen's $d = 0.75$, 95% CI = [0.14, 1.33].

On the other hand, this underestimation effect was not present for any of the similarity-grouping conditions. PSEs for these conditions were 25.8 (SE = 1.2) for

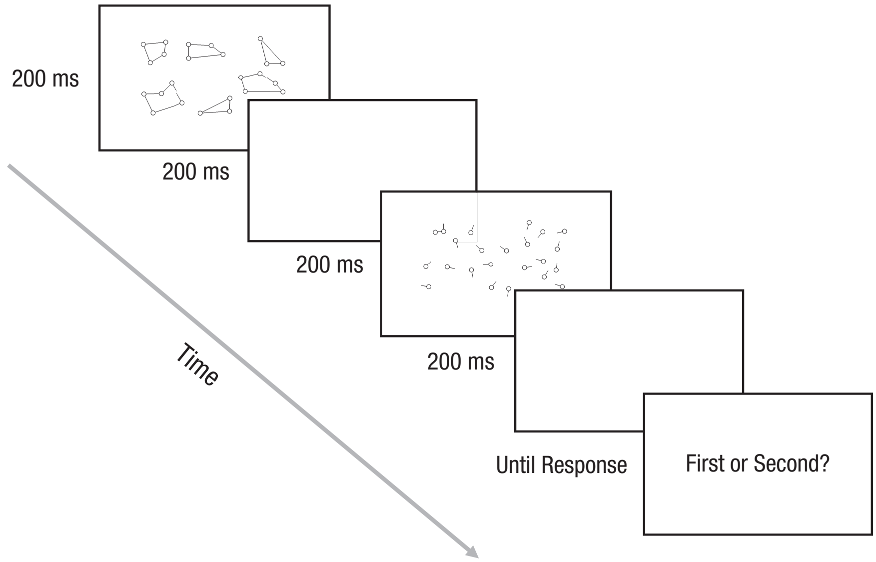

Fig. 2. Display sequence of a typical trial in Experiment 1. The task was to indicate whether the first display or the second display contained more items.

color grouping and 23.2 (SE = 0.6) for its control condition, $t(13)$ = 1.83, p = .091, Cohen's d = 0.49, 95% CI = [−0.05, 1.07]; 24.2 (SE = 0.7) for shape and 24.3 (SE = 0.7) for its control condition, $t(13)$ = .09, p = .927, Cohen's d = 0.02, 95% CI = [−0.50, 0.55]; and 24.9 (SE = 1.2) for orientation grouping and 24.0 (SE = 0.6) for its control condition, $t(13)$ = −0.70, p = .496, Cohen's d = 0.19, 95% CI = [−0.35, 0.71]. Most straightforwardly, PSEs were smaller than 24 for all spatial-grouping cues, $t(13) > 3.94$, $p < .002$, Cohen's $d > 1.05$, and not significantly different from 24 for all similarity-grouping cues, $t(13) < 1.57$, $p > .14$, Cohen's $d < 0.42$ (see Fig. 1a).

Experiment 2

Experiment 1 showed that similarity grouping had no influence on number estimation. But perhaps those similarity groups were created in parallel but were simply not as perceptually strong as the spatial groups. Whereas the set of colors, shapes, and orientations in Experiment 1 was chosen to create the strongest possible groupings, Experiment 2 further redundantly combined color and shape cues to maximize feature-grouping strength (Nothelfer & Franconeri, 2016) and contrasted those with common regions, which showed the weakest underestimation illusion in Experiment 1. In addition, past research has suggested that similarity grouping can take longer to form (which is also consistent with the present account) than spatial-grouping processes for proximity and connectedness (Han, Song, Ding, Yund, & Woods, 2001), which might put similarity cues at a disadvantage in Experiment 1. In Experiment 2, we therefore doubled the presentation duration to 500 ms.

Method

Participants. Twelve Northwestern University undergraduates and members of the Evanston, Illinois, community participated in exchange for $5 or course credit. All participants were required to have normal or corrected-to-normal visual acuity and normal color vision. The number of participants was specified on the basis of a power analysis of Experiment 1: With an average effect size of 1.16 for the estimation bias in spatial-grouping conditions, 12 participants were needed to detect a similar estimation bias in the current experiment with a power of 95%.

Stimuli. Common-region-grouped displays and their control displays were identical to those in Experiment 1. Similarity-grouped displays were created in a similar manner to those in Experiment 1, except that each similarity group contained items that shared the same shape and color. Corresponding control displays consisted of homogeneous items (same shape and color) spatially dispersed in the same way as in the grouped displays. The shape or color used for a control display was not present on the grouped display for that same trial.

Procedure. Each trial started with a 500-ms fixation screen, followed by two successive displays, both of which were presented for either 200 ms or 500 ms (in separate blocks). A 500-ms mask followed each display.

The second mask was followed by a question screen on which participants were instructed to press a key to indicate which display contained more items—the first or the second. The same staircase procedure as in Experiment 1 was implemented. PSEs were calculated for both conditions at the two presentation durations.

Results

Common-region grouping again showed significant underestimation. For the 200-ms duration, PSEs were 23.1 (SE = 0.60) for the common-region-grouped condition and 17.4 (SE = 0.66) for its control condition, $t(11) = 5.82$, $p < .001$, Cohen's $d = 1.65$, 95% CI = [0.75, 2.51], whereas for the 500-ms presentation duration, PSEs were 23.4 (SE = 0.48) for the common-region-grouped condition and 18.7 (SE = 0.75) for its control condition, $t(11) = 5.50$, $p < .001$, Cohen's $d = 1.58$, 95% CI = [0.71, 2.43]. Meanwhile, similarity-grouping cues failed to influence number estimation and, in fact, produced estimates that were in the opposite direction of the trend (higher for the grouped displays and higher than 24). PSEs for the similarity-grouped condition and its control condition were 25.2 (SE = 0.71) and 24.9 (SE = 0.53), respectively, $t(11) = 0.29$, $p = .777$, Cohen's $d = 1.68$, 95% CI = [−0.49, 0.65], for the 200-ms presentation duration and 24.6 (SE = 0.49) and 23.9 (SE = 0.62), respectively, $t(11) = 0.84$, $p = .421$, Cohen's d = 0.08, 95% CI = [−0.34, 0.81], for the 500-ms presentation duration. There was no difference between long and short display durations and no sign of an interaction between display duration and grouping type. PSEs were lower than 24 at both display durations for common-region grouping, $t(11) > 5.99$, $p < .001$, Cohen's $d > 1.73$, and not significantly different from 24 for both display durations for similarity-grouping cues, $t(11) < 1.29$, $p >$ .23, Cohen's $d < 0.37$ (see Fig. 1b).

Experiment 3

The previous experiment found underestimation for the weakest spatial cue but not for the redundant combination of color and shape. In Experiment 3, we used an independent task to match the strength of similarity-grouping cues (again, redundant color and shape groups) with the strength of proximity, a spatial-grouping measure whose strength can be quantitatively titrated by adjusting the distances between grouped objects (Huang, 2015).

Method

Participants. Sixty-two Northwestern University undergraduates and members of the Evanston, Illinois, community completed the strength-matching task, and 12 participated in the number-estimation task in exchange for $5 or course credit. All participants were required to have normal or corrected-to-normal visual acuity and normal color vision. The number of participants for the number-estimation task was decided on the basis of the previous experiment. The number of participants for the strength-matching task was arbitrarily determined as a large sample of 60 to capture and average over potentially substantial variance in participants' perception of relative strength between similarity and proximity grouping. The experiment included 62 participants because too many signed up.

Stimuli. For the strength-matching task, we created an array of items that could be grouped into columns tilting 45° to the left or right when they were grouped by similarity or proximity (see Fig. 3). The array was viewed behind a round aperture with a radius of 6.1° and consisted of the same shape types as used in Experiment 2. The initial distance between two neighboring items that

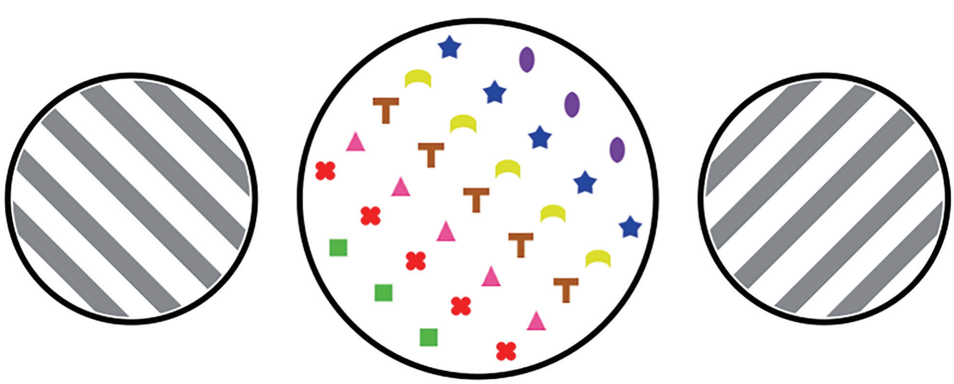

Fig. 3. Example of an experimental display for the strength-matching procedure of Experiment 3. The objects in the array can be perceived as being grouped into columns tilting 45° to the right if grouped by proximity and to the left if grouped by similarity.

were within each proximity group was 1.0°, and the initial distance between two neighboring proximity groups was 2.4°. Proximity groups were randomly selected to be tilting left or right on a trial-to-trial basis, and similarity groups were tilting in the orthogonal direction.

In the number-estimation task, similarity-grouped displays and control displays were generated similarly to those in Experiment 2. For the proximity-grouped displays, we set the between-group distance to be 1.6 times the within-group distance. This ratio was acquired through a strength-matching measurement before the number-estimation task, described in more detail in the following section. Detailed information about how this ratio was implemented can be found in the Supplemental Material. The corresponding control display consisted of dots interspersed within the same area as the proximity-grouped display. Their positions were chosen in the same manner as the control displays for proximity grouping in Experiment 1.

Procedure. Before the start of Experiment 3, a separate group of 62 participants completed the strength-matching task across multiple pilot studies. A pattern similar to the center circle in Figure 3 was presented in the center of the screen for 200 ms, followed by a 500-ms mask. Participants were instructed to make an intuitive judgment about whether the pattern was tilting to the left or right. We used a staircase procedure in which the ratio of the between-proximity-group distance to the within-proximity-group distance increased by 0.1 if participants saw similarity grouping as stronger, and that ratio decreased if participants saw similarity grouping as weaker. The staircase ended when the number of reversals reached 20. The threshold ratio was calculated by averaging the ratios at the last 5 reversal points. For observers (8 of 62) whose ratio dropped to 1, they saw the organization consistent with proximity grouping while the between-proximity-group distance equaled that of the within-proximity-group distance; we terminated the staircase procedure and recorded those participants' threshold ratio as 1. For participants (5 of 62) whose ratio reached 3, they saw that similarity grouping dominated even when the items started to overlap; we terminated the staircase procedure and recorded their ratio as 3. The median of the ratios, including data from participants who failed to complete the staircases, was 1.62 (SE = 0.07). The average of the ratios, excluding those data, was 1.84 (SE = 0.06). We used the more conservative ratio of 1.60 to create the proximity-grouping displays in the number-estimation task to avoid making the proximity-grouping strength too strong.

In Experiment 3, participants completed the same number of comparison tasks as in Experiment 2 and went through the same staircase procedure. Each display was presented for 200 ms, followed by a 500-ms mask. Conditions and their associated staircases were randomly interleaved. After completing the number-estimation task, participants completed the same strength-matching task used in the pilot studies. We set the sequence this way to avoid getting participants into a mind-set of looking for visual structures during the number-estimation task.

Results

Proximity grouping again showed a reduction in number estimates: The PSE was 22.0 (SE = 0.54) for the proximity-grouped condition and 24.1 (SE = 0.67) for its control condition, $t(11)$ = 2.67, p = .02, Cohen's d = 0.77, 95% CI = [0.11, 1.41]. However, no such underestimation effect was found for similarity grouping: The PSE was 26.0 (SE = 0.56) for the similarity-grouped condition and 24.6 (SE = 0.68) for its control condition, $t(11)$ = 1.56, p = .15, Cohen's d = 0.45, 95% CI = [0.15, 1.04]. PSEs were robustly smaller than 24 for proximity grouping, $t(11)$ = 3.27, p = .007, Cohen's d = 0.94, 95% CI = [0.24, 1.61], and larger than 24 for similarity grouping, $t(11)$ = 3.84, p = .003, Cohen's d = 1.11, 95% CI = [0.37, 1.82] (see Fig. 1c).

Three participants from this set failed to complete the strength-matching procedure because they saw the organization as consistent with proximity grouping while the between-proximity-group distance equaled that of the within-proximity-group distance. Their ratios were recorded as 1. The average ratio, excluding those participants, was 1.64 (the median ratio, including those participants, was 1.47, which is still comparable with the ratio of 1.60 implemented here).

General Discussion

Feature-based selection provides a parsimonious mechanism for gestalt similarity grouping, but it counterintuitively predicts that groups can be formed only serially. Past results showing response-time increments for creating additional groups have been consistent with this account (e.g., Levinthal & Franconeri, 2011; Yu, Tam, & Franconeri, 2019), but these increments could also be due to decisions made about the groups, instead of the grouping process per se.

We present strong evidence for this account, using a task that should isolate the grouping process—a number-underestimation illusion that should not appear for a serial-grouping process. Our experiments replicated previous findings showing number underestimation for spatial-grouping cues, such as connectedness (Franconeri et al., 2009; He et al., 2009), common region (He et al., 2015), proximity (Frith & Frith, 1972), and even connections from illusory contours (Kirjakovski & Matsumoto, 2016). But this effect was systematically absent for the similarity-grouping cues of color, shape, redundant color and shape combined, and orientation across multiple exposure durations, as well as for a procedure designed to equate the grouping strength of the spatial and similarity cues.

This account's counterintuitive prediction that multiple feature groups cannot be constructed at the same time might appear to contradict our explicit experience of a simultaneously organized world (note that Huang, Treisman, & Pashler, 2007, posited a version of this account in which the viewer can see the complementary set, e.g., all of the not-red objects, as a group at the same time). Future work should ensure that this pattern of results generalizes to more spatially extended displays that contain more complex real-world objects. Future work should also strive to rule out the possibility that a properly motivated participant could be induced to construct similarity groups in parallel within an enumeration task.

Serial access to similarity groups does not entail that we can access information about only one feature (e.g., red) at a time. Rather, the visual system is capable of providing statistical information about which features exist, and in what relative quantities, creating mental representations analogous to a histogram over feature spaces such as hue, orientation, or size (Haberman & Whitney, 2012). The visual system can produce these summaries within natural worlds, informing us, for example, that an orchard contains a splash of red among ample green, extracted from apples and leaves (Oliva & Torralba, 2006), and within artificial worlds, informing us, say, that a scatterplot contains a roughly equal amount of red and green, extracted from two color-coded classes of data points (Szafir, Haroz, Gleicher, & Franconeri, 2016). These features are not grouped into visual structures until selection mechanisms link them to their corresponding spatial locations (Franconeri, 2013; Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012; Luck & Ford, 1998; Treisman & Gelade, 1980). Even if green and red objects are not each simultaneously grouped at any perceptual moment, this information could provide, for example, the relative number of objects across multiple hue values (Halberda, Sires, & Feigenson, 2006; Poltoratski & Xu, 2013). This histogram could guide serial selection of feature values via rapid cycles of selection and inhibition across its modes (Bruce & Tsotsos, 2006).

If similarity grouping requires serial selection of features—one of many operationalizations of attention—then why do some past results show that similarity grouping is possible even when a secondary task engages attention elsewhere (Kimchi & Razpurker-Apfeld, 2004; Moore & Egeth, 1997; Russell & Driver, 2005; Shomstein, Lee, & Behrmann, 2010)? Although these studies constrained participants' spatial attention to one region of the display, we must predict that participants could still apply feature-based selection to other parts of the display to allow similarity grouping. Another past result that at first glance might seem in conflict shows that the repeated appearance of particular color pairs biases observers to underestimate the total numerosity of the display (Zhao & Yu, 2016). However, this underestimation effect was not due to the pair groupings, per se, but to their status as implicitly recognized repeated patterns over time. Indeed, in conditions in which other nongrouped patterns repeated over time, those conditions showed equal or greater underestimation even in the absence of color groups.

In addition to the prediction tested here, the feature-based selection account makes other falsifiable predictions. First, if one could constrain participants' feature-based attention, for example, by adding a dual task that required filtering for red while simultaneously measuring grouping of objects that were green, the present account predicts that the green grouping would fail. Second, given a neuroimaging method that detects feature-based attention with sufficient temporal resolution (e.g., event-related potential in humans or single-cell recording in primates), one should be able to track serial attention to each feature if the task requires grouping based on those features (for a discussion, see Woodman & Luck, 2003).

In conclusion, the present experiments provide a parsimonious and falsifiable account of the underlying mechanism for gestalt similarity grouping. Although this account's prediction of low capacity—just one grouping at a time—may seem like a deficit, we see it as a feature. It allows for similarity grouping that is flexible in a world in which color, shape, and orientation may provide conflicting groupings and the choice of group often depends on current goals.

Action Editor

Philippe G. Schyns served as action editor for this article.

Author Contributions

All the authors contributed to the study design. X. Xiao and D. K. Bemis completed pilot experiments for the study. D. Yu conducted the testing and data collection, analyzed and interpreted the data, and drafted the manuscript. S. L. Franconeri provided critical revisions. All the authors approved the final manuscript for submission.

ORCID iD

Dian Yu https://orcid.org/0000-0003-4171-8775

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618822798

Open Practices

All data and materials have been made publicly available via the Harvard Dataverse and can be accessed at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/9KSHOE. The design and analysis plans for these experiments were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797618822798. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

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

Bruce, N., & Tsotsos, J. (2006). Saliency based on information maximization. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (pp. 155–162). Cambridge, MA: MIT Press.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*, 425–428.

Franconeri, S. L. (2013). The nature and status of visual resources. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 147–162). New York, NY: Oxford University Press.

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*, 1–13.

Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. (2012). Flexible visual processing of spatial relationships. *Cognition*, *112*, 210–227.

Frith, C. D., & Frith, U. (1972). The solitaire illusion: An illusion of numerosity. *Perception & Psychophysics*, *11*, 409–410.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). New York, NY: Oxford University Press.

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

Han, S., Song, Y., Ding, Y., Yund, E. W., & Woods, D. L. (2001). Neural substrates for visual perceptual grouping in humans. *Psychophysiology*, *38*, 926–935.

He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, *16*, 509–517.

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences, USA*, *112*, E5647–E5655.

Huang, L. (2015). Grouping by similarity is mediated by feature selection: Evidence from the failure of cue combination. *Psychonomic Bulletin & Review*, *22*, 1364–1369.

Huang, L., & Pashler, H. (2002). Symmetry detection and visual attention: A “binary-map” hypothesis. *Vision Research*, *42*, 1421–1430.

Huang, L., & Pashler, H. (2007). A Boolean map theory of visual attention. *Psychological Review*, *114*, 599–631.

Huang, L., Treisman, A., & Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science*, *317*, 823–825.

Kimchi, R., & Razpurker-Apfeld, I. (2004). Perceptual grouping and attention: Not all groupings are equal. *Psychonomic Bulletin & Review*, *11*, 687–696.

Kirjakovski, A., & Matsumoto, E. (2016). Numerosity underestimation in sets with illusory contours. *Vision Research*, *122*, 34–42.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What’s new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Suppl.).

Levinthal, B. R., & Franconeri, S. L. (2011). Common-fate grouping as feature selection. *Psychological Science*, *22*, 1132–1137.

Luck, S. J., & Ford, M. A. (1998). On the role of selective attention in visual perception. *Proceedings of the National Academy of Sciences, USA*, *95*, 825–830.

Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 339–352.

Nothelfer, C., & Franconeri, S. (2016). Feature redundancy benefits in different attentional modes. *Journal of Vision*, *16*(12), Article 682. doi:10.1167/16.12.682

Nothelfer, C., Gleicher, M., & Franconeri, S. (2017). Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1667–1676.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, *4*, 118–123.

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

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, *13*(8), Article 6. doi:10.1167/13.8.6

Russell, C., & Driver, J. (2005). New indirect measures of “inattentive” visual grouping in a change-detection task. *Perception & Psychophysics*, *67*, 606–623.

Shomstein, S., Lee, J., & Behrmann, M. (2010). Top-down and bottom-up attentional guidance: Investigating the role of the dorsal and ventral parietal cortices. *Experimental Brain Research*, *206*, 197–208.

Szafir, D. A., Haroz, S., Gleicher, M., & Franconeri, S. (2016). Four types of ensemble coding in data visualizations. *Journal of Vision*, *16*(5), Article 11. doi:10.1167/16.5.11

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, *25*, 168–199.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Trick, L. M., & Enns, J. T. (1997). Clusters precede shapes in perceptual organization. *Psychological Science*, *8*, 124–129.

von Ehrenfels, C. (1988). On 'gestalt qualities.' In B. Smith (Ed. & Trans.), *Foundations of gestalt theory* (pp. 82–117). Munich, Germany: Philosophia. (Original work published 1890)

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*, 1172–1217.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495–501.

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 121–138.

Yu, D., Tam, D., & Franconeri, S. L. (2019). Gestalt similarity groupings are not constructed in parallel. *Cognition*, *182*, 8–13.

Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217–222.