Common-Fate Grouping as Feature Selection

Brian R. Levinthal and Steven L. Franconeri
Northwestern University

Abstract

The visual system groups elements within the visual field that are physically separated yet similar to each other. Although grouping processes have been intensely studied for a century, the mechanisms of grouping remain elusive. We propose that a primary mechanism for grouping by *common fate* is attentional selection of a direction of motion. A unique prediction follows from this account: that the visual system must be limited to forming only a single common-fate group at a time, and that attempts to find a particular common-fate group among other groups, or among nongroups, should therefore be highly inefficient. We show that this is true in searches for vertically oriented groups of moving dots among horizontally oriented groups (Experiment 1) and in searches for motion-linked groups among nonlinked objects (Experiment 2). Feature selection may limit the visual system to the construction of only one common-fate group at a time, and thus the experience of simultaneous grouping may be an illusion.

The human visual system breaks an incoming image into separate elements, but also reassembles these elements into groups. Grouping processes have been studied for a century, and much of this work has focused on classifying different types of groups, determining when grouping occurs (S. E. Palmer, 2002), and measuring the relative strength of different forms of grouping (Kubovy & van den Berg, 2008). Although neurally plausible mechanisms have been proposed for some types of grouping (e.g., contour grouping; Roelfsema, 2006), for many other types of grouping, such explanations remain elusive. We propose a mechanism for the Gestalt principle of *common fate*, according to which objects appear grouped when they display the same pattern of motion. We argue that a primary mechanism underlying this powerful form of grouping is the selection of a direction of motion.

This parsimonious explanation is consistent with previous demonstrations that the selection of a direction of motion (as well as of features such as color and orientation) can occur in parallel across the visual field (Maunsell & Treue, 2006; Saenz, Buracas, & Boynton, 2002). Selecting a direction of motion could produce a spatially organized map of locations in the visual field that correspond to the selected direction (see Fig. 1); "peaks" in this map would correspond to the locations of all elements moving in that direction (see Huang & Pashler, 2007, for a similar account). The initial selection of a direction of motion could be based on a statistical summary of the motion patterns within a display (Williams & Sekuler, 1984) or on the direction of motion of a single selected object. Once a group has been created, changes to its direction of motion could be updated by a simple feedback loop that modifies the selected direction of motion (see Martinez-Trujillo, Cheyne, Gaetz, Simine, & Tsotsos, 2007, for evidence of motion-direction updating during the tracking of translating objects).

Such a map would provide two critical components of the grouping process. First, it would help isolate processing of features and identities to the objects at the locations of the peaks in the map (Treisman & Gelade, 1980). Simultaneously selecting these locations might lead them to appear to belong together, even though they are spatially separated (Xu, 2008). Simultaneous selection might also provide a summary representation (e.g., Alvarez & Oliva, 2009) of other features at these locations, leading to a holistic representation of the otherwise separate objects. For example, selecting a set of rightward-moving objects could facilitate judgments about the number of those objects (Halberda, Sires, & Feigenson, 2006) or the distribution of their sizes (Chong & Treisman, 2005).

Corresponding Authors:
Brian R. Levinthal, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208
E-mail: b-levinthal@northwestern.edu

Steven L. Franconeri, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208
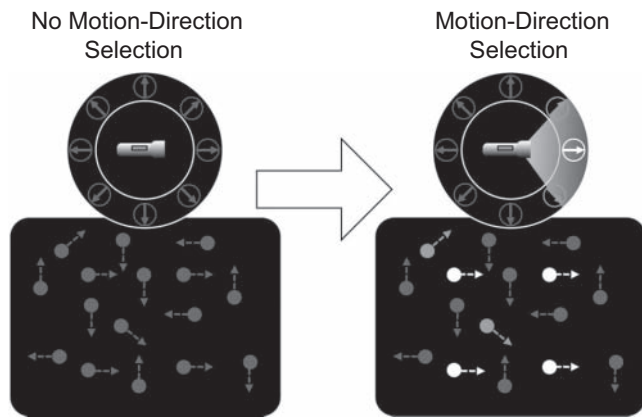E-mail: franconeri@northwestern.edu

No Motion-Direction Selection

Motion-Direction Selection

Fig. 1. Illustration of a feature-selection mechanism for common-fate grouping. The flashlights at the top illustrate whether or not a direction of motion has been selected. If an observer did not select a direction of motion (illustration at the left), all moving elements in a display would be processed together, and there would be no common-fate grouping. If an observer selected rightward motion (illustration at the right), target elements would be enhanced and form a common-fate group. Changes in the direction of selected motion could be updated via feedback connections within motion-sensitive areas of the visual system (e.g., Martinez-Trujillo, Cheyne, Gaetz, Simine, & Tsotsos, 2007).

Second, because any area of the visual field containing the selected feature would result in a peak in the spatially organized map, the complex distribution of the selected locations would convey the shape of the group (Huang & Pashler, 2007).

The mechanism we propose is consistent with a powerful property of common-fate grouping: A group can be constructed in a massively parallel operation in which all elements across the visual field that share a feature are selected. However, our account also suggests that common-fate grouping should have a salient weakness: Peaks on a spatially organized map should be distinguishable only by their location, so the visual system should be able to construct only one group at a time. This prediction contrasts with observers' experience that they see multiple common-fate groups simultaneously. We report the results of two visual search experiments that demonstrate that there are severe limitations on the number of common-fate groups that can be constructed at once.

Experiment 1

In Experiment 1, we asked participants to perform a visual search for a vertically oriented group among horizontally oriented groups; groups consisted of pairs of dots that were moving in the same direction. Because searches for objects with a unique orientation are typically efficient (e.g., Treisman & Gelade, 1980), this search task is well suited for determining whether common-fate groups can be formed in parallel. To evaluate participants' efficiency in performing this task, we included a control condition, in which the elements in each common-fate group were connected by a 1-pixel line.

Method

Ten participants (ages 18 to 21) took part in this study. All were either volunteers from the Northwestern University student population or paid participants from the Evanston, Illinois, community.

Stimuli were presented on a 17-in. CRT monitor (resolution of 1024 × 768 pixels, 85-Hz refresh rate) and were generated using MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997).

Displays consisted of pairs of moving dots against a static background of 250 randomly positioned dots (approximately 0.3 dots per square degree). The two dots in each pair moved within a square region measuring 5.2° on each side, maintained a constant distance (2°) from each other throughout the trial, and moved at a constant velocity (2.65°/s). Each dot pair was assigned an initial random angle of motion and moved along a straight path until one of the dots reached the invisible boundary of the square region, at which point the pair's angle of motion was reflected. On each trial, two, four, or eight dot pairs were placed in adjacent positions along an imaginary circle with an 8° radius.

Trials were blocked into two conditions. In the grouped condition, participants were asked to determine whether a target (i.e., a vertically oriented dot pair) was present or absent in each display; the dots in a display continued to move until participants responded via key press. In the connected (control) condition, the task was identical, except that the dots within each pair were connected by a 1-pixel line. Figure 2a depicts the experimental task. In target-absent trials (50% of trials), all dot pairs were arranged horizontally. In target-present trials (50% of trials), a single pair of dots was arranged vertically. Movies of sample trials are available online at http://viscog.psych.northwestern.edu/.

Results

Figure 2b depicts the results for Experiment 1. According to our account, if only one common-fate group is created at a time, adding more groups to the display should substantially increase response times. Indeed, response times in the grouped condition increased with the number of distractor pairs, and the search slope of 55 ms per pair was significantly different from 1, $t(9) = 17.3$, $p < .001$. When common-fate groups were replaced by connectivity groups (connected condition), which can be processed in parallel (Franconeri, Bemis, & Alvarez, 2009; Rensink & Enns, 1995), adding more groups to the display did not substantially affect response times (search slope = 6 ms/pair), $t(9) = 2.7$, $p = .023$. Furthermore, searches among common-fate groups were significantly less efficient than searches among connectivity groups, $t(9) = 12.3$, $p < .001$. Thus, the costs of searching through the common-fate groups were attributable to how they were constructed, as the two conditions did not differ in display complexity or the shape-identification requirements of the search task.
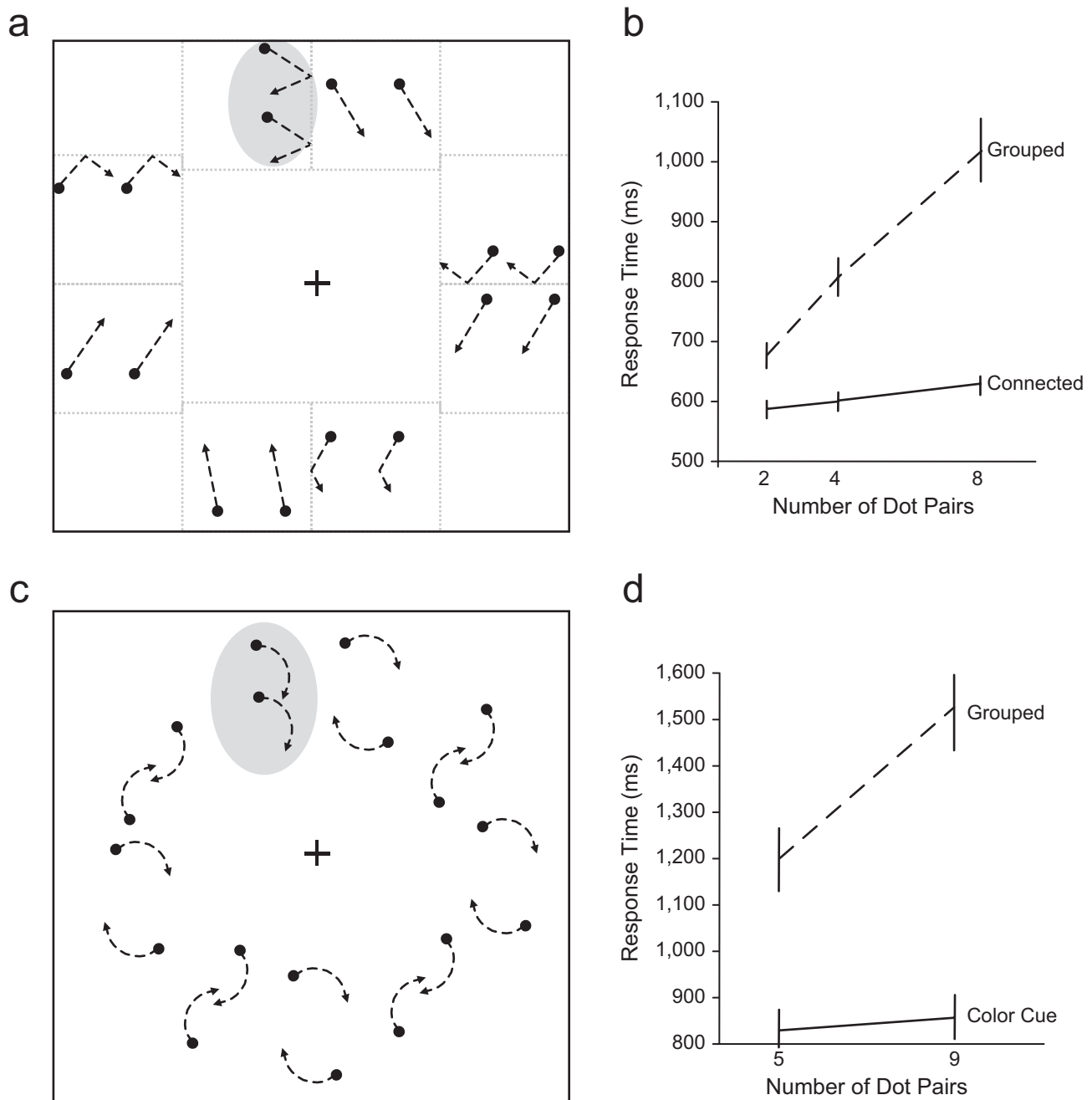
a

b

c

d

Fig. 2. Experimental stimuli and results. In Experiment 1, each display contained two, four, or eight dots pairs defined by common movement (a), and participants searched for a vertically oriented pair (highlighted here by the gray shading) among horizontally oriented pairs; the graph (b) shows response time in the two conditions as a function of the number of dot pairs in the display. In Experiment 2, each display contained five or nine dot pairs; the dots in distractor dot pairs moved 180° out of phase, and the dots in target pairs moved in phase (c). Participants indicated whether a target dot pair was absent or present on each trial. The graph (d) shows response time in the two conditions as a function of the number of dot pairs in the display. Error bars indicate standard errors of the mean.

Experiment 2

Experiment 1 showed that the capacity for extracting the shape of a common-fate group is severely limited. However, even if shape information is not available, the visual system may be able to detect coarse properties of groups in a display (e.g., the number of clusters of grouped elements; see

Trick & Enns, 1997). Although discrimination of the shape of common-fate groups may be inefficient, observers may nonetheless be able to efficiently detect the existence of such groups. To test this possibility, we altered the visual search task so that participants judged whether a single common-fate group was present in a display filled with nongrouped objects (see Fig. 2c).

Method

Twelve participants (ages 18 to 21) took part in this experiment. All were either volunteers from the Northwestern University student population or paid participants from the Evanston, Illinois, community.

The displays in Experiment 2 consisted of pairs of moving dots against a black background. Either five or nine dot pairs were displayed on each trial. Each dot orbited an imaginary circle with a 1° radius at a rate of 6.28°/s. In distractor dot pairs, the two dots revolved around separate loci and moved 180° out of phase. Paired distractor dots never approached closer than 0.5° and were never separated by more than 3.5°. The dots in a target pair also revolved around separate loci, but their movement was in phase, and therefore the dots maintained a constant separation of 2°. A target pair was assigned a unique phase, and distractor dots in the same display were always at least 75° out of phase from the dots in the target pair. Trials were blocked into two conditions. In the grouped condition, all dot pairs were white. In the color-cue condition, distractor dot pairs were white, and the target dot pair was magenta (on each target-absent trial, a random distractor dot pair was magenta). Participants were instructed to determine whether a target was present or absent in each display; motion persisted in the display until participants indicated their response via key press. Movies of sample trials are available at http://viscog.psych.northwestern.edu. Figure 2c depicts the experimental task.

Results

Response times in the grouped condition increased with the number of distractor pairs, and the search slope of 82 ms per pair was significantly different from 0, $t(11) = 12.3$, $p < .001$ (Fig. 2d); this result suggests that the capacity for detecting the presence of common-fate groups is severely restricted. Results for participants in the color-cue condition (Fig. 2d) confirmed that the search costs associated with additional distractors were not simply due to increased display complexity: When the target group was cued with a unique color, adding more distractors to the display did not affect response times (search slope = 7 ms/pair), $t(11) = 1.49$, $p = .16$. Also, search slopes in the two conditions differed significantly from each other, $t(11) = 7.45$, $p < .001$.

Although Experiment 2 showed that the capacity for detecting common-fate groups is severely limited, there may be ways in which an observer can detect groups efficiently in certain kinds of displays. For example, consider a common-fate group consisting of coherently moving elements among randomly moving elements. As the number of objects within the group increases (relative to the total number of objects on the screen), the group's direction of motion should begin to dominate the global distribution of motion directions in the entire display. In this case, competition among multiple directions of motion might be resolved through simple feedback mechanisms in the visual system (Chey, Grossberg, & Mingolla,

1997; Martinez-Trujillo et al., 2007) such that the dominant direction of motion would tend to be selected first. This selection would result in an efficient and seemingly automatic representation of a common-fate group. By contrast, Experiment 2 shows that when a common-fate group's direction of motion does not dominate that of other groups by a significant ratio, the common-fate group cannot be efficiently detected.

Discussion

We have argued that a core mechanism of common-fate grouping is the selection of a direction of motion. This selection would enhance the neural activity associated with similarly moving elements across the visual field, providing the visual system with a map of locations for further processing. If this account is correct, it should be possible to form only one common-fate group at a time. We found evidence consistent with this prediction in two experiments. In Experiment 1, participants searched for a vertically oriented group among horizontally oriented groups. When common-fate grouping was required to perform the task, searches were highly inefficient. However, when elements were connected by lines, the vertical group could be identified instantly. Similarly, in Experiment 2, when participants searched for a single common-fate group among nongrouped objects, searches were highly inefficient.

We argue that the inefficiency of these searches was due to the need to sequentially select the current direction of motion of each group. We note that two alternative explanations are theoretically possible. First, it is possible that common-fate groups are formed in parallel, but the search process cannot access these representations efficiently. This alternative is unlikely because previous studies have shown performance on visual search tasks to be a sensitive measure of visual grouping processes (e.g., Trick & Enns, 1997; Rensink & Enns, 1995). Second, searches for common-fate groups might be inefficient because the parallel grouping process has a limited capacity (see J. Palmer, 1995, for discussion). This alternative is also unlikely, because it is unclear why such a resource limitation might arise; this account would also require additional mechanisms to maintain the representations of a potentially unrestricted number of common-fate groups. Compared with these alternatives, our account of sequential selection has the powerful advantage of parsimony. Because previous work has shown that directions of motion can be selected in a global fashion (Saenz, Buracas, & Boynton, 2002), we can concretely specify how this known mechanism could support common-fate grouping. The burden of proof therefore falls on alternative accounts to identify experimental effects that cannot be explained by serial feature selection.

The feature-selection account, at first glance, appears to contradict prior observations that feature-based grouping can be accomplished outside the focus of attention (Kimchi & Razpurker-Apfeld, 2004; Moore & Egeth, 1997; Russell & Driver, 2005). Those prior studies showed that when participants performed a demanding primary task at fixation,

the perceptual organization of an irrelevant background (groupings of luminance, color, or orientation) influenced performance, even though participants reported no awareness or memory of the background's content. However, any ostensible conflict between our results and those of previous studies would be due to different definitions of *attention*. Prior studies have shown a processing bottleneck that prevents perceptual groups from reaching the processing stages required for awareness or memory. Grouping by feature selection should still be possible as long as the selection of features is not incompatible with the primary task.

Although the present study focused on common-fate grouping, feature selection may be a more generalized mechanism for grouping by any type of similarity, including not only similarity of motion, but also similarity of color, shape, or orientation (Huang & Pashler, 2007). If so, grouping by similarity may be subject to the same limits as common-fate grouping: The visual system may be able to create only one group at a time. This prediction is supported by recent studies in which observers were asked to determine whether multicolored patterns were symmetrical. In these studies, judgments were faster when the displays consisted of few colors and slower when the displays consisted of more colors; these results suggest that symmetry judgments can be made for only a single color subpattern at a time (Morales & Pashler, 1999). The proposal that feature selection is the mechanism for grouping by similarity implies a substantial difference between grouping by similarity and other types of grouping. Specifically, grouping by proximity, connectivity, or common region may rely on a different mechanism that can produce discrete units in parallel (Franconeri et al., 2009; S. Palmer & Rock, 1994; Rensink & Enns, 1995). The term *grouping* is likely too broad to capture the diversity of mechanisms that cause some elements of the visual field to be associated with others.

Selection of a direction of motion may not be the only mechanism for grouping moving objects. There are almost certainly long-term representations for more specific patterns of motion, such as those produced by walking bodies, flapping wings, or mouths moving during speech (Cavanagh, Labianca, & Thornton, 2001), and these representations could be used to facilitate perception. A challenge for our proposed account, which predicts that only a single surface is available for perceptual processing at a given time, is to explain more complex grouping abilities related to common fate, such as the mental construction of rigid three-dimensional objects from a dense array of moving dots (the structure-from-motion phenomenon; Wallach & O'Connell, 1953). Although such percepts might require complex local processing of the pattern of motion (Ullman, 1979), we argue that the relatively simple mechanism described in this article may produce surprisingly rich percepts, especially when combined with other cues, such as statistical information (Alvarez & Oliva, 2009; Balas, Nakano, & Rosenholtz, 2009) about the distribution of directions of motion and edges created by motion contrast (Regan & Beverley, 1984).

For simple common-fate groups, and possibly for more complex patterns of motion, feature selection presents a parsimonious account of grouping by motion. Observers may construct a single such group at a time, and what seems like the construction of multiple groups may be an illusion of perceptual detail.

References

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

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), Article 13. Retrieved from http://www.journalofvision.org/content/9/12/13

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

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition*, *80*, 47–60.

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, *14*, 2570–2594.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.

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

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

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

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

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.

Martinez-Trujillo, J. C., Cheyne, D., Gaetz, W., Simine, E., & Tsotsos, J. K. (2007). Activation of area MT/V5 and the right inferior parietal cortex during the discrimination of transient direction changes in translational motion. *Cerebral Cortex*, *17*, 1733–1739.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*, 317–322.

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

Morales, D., & Pashler, H. (1999). No role for colour in symmetry perception. *Nature*, *399*, 115–116.

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

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

Palmer, S. E. (2002). Perceptual grouping: It's later than you think. *Current Directions in Psychological Science*, *11*, 101–106.

Regan, D., & Beverley, K. I. (1984). Figure-ground segregation by motion contrast and by luminance contrast. *Journal of the Optical Society of America*, *1*, 433–442.

Rensink, R. A., & Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review*, *102*, 101–130.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*, 203–227.

Russell, C., & Driver, J. (2005). New indirect measures of "inattentive" visual grouping in a change-detection task. *Perception & Psychophysics*, *64*, 606–623.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*, 631–632.

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.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society B: Biological Sciences*, *203*, 405–426.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*, 55–62.

Xu, Y. (2008). Representing connected and disconnected shapes in intraparietal sulcus. *NeuroImage*, *40*, 1849–1856.