Perceptual Grouping: It's Later Than You Think

Stephen E. Palmer[1]

Psychology Department, University of California, Berkeley, California

Abstract

Recent research on perceptual grouping is described with particular emphasis on the level at which grouping factors operate. Contrary to the standard view of grouping as an early, two-dimensional, image-based process, experimental results show that it is strongly influenced by binocular depth perception, lightness constancy, amodal completion, and illusory figures. Such findings imply that at least some grouping processes operate at the level of conscious perception rather than the retinal image. Whether classical grouping processes also operate at an early, preconstancy level is an important, but currently unanswered question.

Keywords

grouping; organization; Gestalt; constancy

Gestalt psychologists are justly famous for their groundbreaking work on understanding how images on the retina are organized into the objects of perceptual experience. When you view an automobile parked behind a telephone pole, for example, you perceive a single, unified car rather than seeing the left and right halves as two disconnected, independent objects. The Gestaltists were not only the first to make useful contributions toward solving this problem, but also the first to recognize that it even existed. Simply stated, the problem of perceptual organization is that the objects of conscious perception are not directly given in any simple or direct way in the retinal image, but must be constructed through activity of the visual nervous system. How does this happen?

Wertheimer (1923/1950) took the first step toward an answer by trying to determine which stimulus factors govern "what goes with what" in the retinal image. The results of his studies are the famous Gestalt laws, or principles of grouping, which state that objects are grouped together when they are close, similar, moving together, and so forth. Grouping is among the best known, yet least understood, phenomena of visual perception.

Recent demonstrations from my own laboratory have added to this list the principles of common region, element connectedness, and synchrony. According to the principle of common region, elements that lie within the same bounded area tend to be grouped together, as the spots of a leopard are grouped within its contours (Palmer, 1992). The principle of element connectedness is that elements that share a common border tend to be grouped together, as are the head and handle of a hammer (Palmer & Rock, 1994). The principle of synchrony is that elements that change at the same time tend to be grouped together (Palmer & Levitin, 2002). It is related to the principle of common fate, but the simultaneous changes do not have to involve motion or to be "common" in any sense. Why grouping by synchrony should occur is somewhat mysterious, because everyday examples are hard to find.

The perceptual processes underlying classical grouping phenomena have generally been assumed to be relatively primitive, low-level operations that work on some early, two-dimensional (2-D) representation and create an initial set of discrete elements on which subsequent perceptual operations are performed (e.g., Marr, 1982). It is generally thought that such initial elements are required to achieve what is called perceptual constancy, which refers to the fact that observers usually perceive the constant, unchanging properties of physical objects despite wide variations in their projected optical images due to different viewing conditions. If this is the case, then grouping operations should occur before the processes that support constancy, including binocular depth perception, surface lightness perception, and the completion of partly occluded objects. I call the extreme version of this idea—that grouping occurs *only* at an early, preconstancy level—the "early-only" view of grouping. Whether it stands up to theoretical and empirical scrutiny is the primary focus of this article.

THEORETICAL CONSIDERATIONS

In a previous article, Rock and I challenged the early-only view of grouping on purely theoretical grounds (Palmer & Rock, 1994). First, we pointed out that although Wertheimer's demonstrations of grouping involved putting together two or more discrete elements—for example, grouping the dots in Figure 1a into vertical columns—he never actually said where the elements themselves came from. Presumably he believed that they were somehow derived from the grouping principles he articulated, but we argued that they arise from a different kind of organizational principle that we call *uniform connectedness*. Uniform connectedness is the principle by which the visual system partitions the image into connected regions having uniform (or smoothly changing) properties, such as luminance, color, motion, and texture, in much the way that a stained-glass window has regions with different visual properties.

Regions defined by uniform connectedness do not acquire the status of distinct visual elements until figure-ground organization determines which ones correspond to perceived objects and which to backgrounds or spaces between objects. Once figural regions have been designated as entry-level perceptual elements, they can then be aggregated into larger, superordinate units by principles of grouping or divided into smaller, subordinate units by parsing at places at which the contour curves sharply inward.

Notice that this theory puts classical perceptual grouping operations farther along the chain of visual information processing than has generally been assumed, after region segmentation and figure-ground organization have already provided a set of perceptual elements. Because figure-ground processing can be viewed as a form of depth perception that uses so-called pictorial cues to determine what is in front of what and to which region the boundaries belong (Palmer, 1999), our analysis suggests that grouping may occur after depth perception and constancy. The level at which grouping processes operate is ultimately an empirical question, however. Despite the importance of this question, until recently few experiments have been directly concerned with answering it.

BINOCULAR DEPTH EFFECTS

The influence of binocular depth perception on proximity grouping has been examined by asking whether the distances that govern proximity grouping are defined in the 2-D retinal image or in perceived 3-D space. In one such experiment (Rock & Brosgole, 1964), observers who were located in a dark room viewed a 2-D array of

Published by Blackwell Publishing Inc.

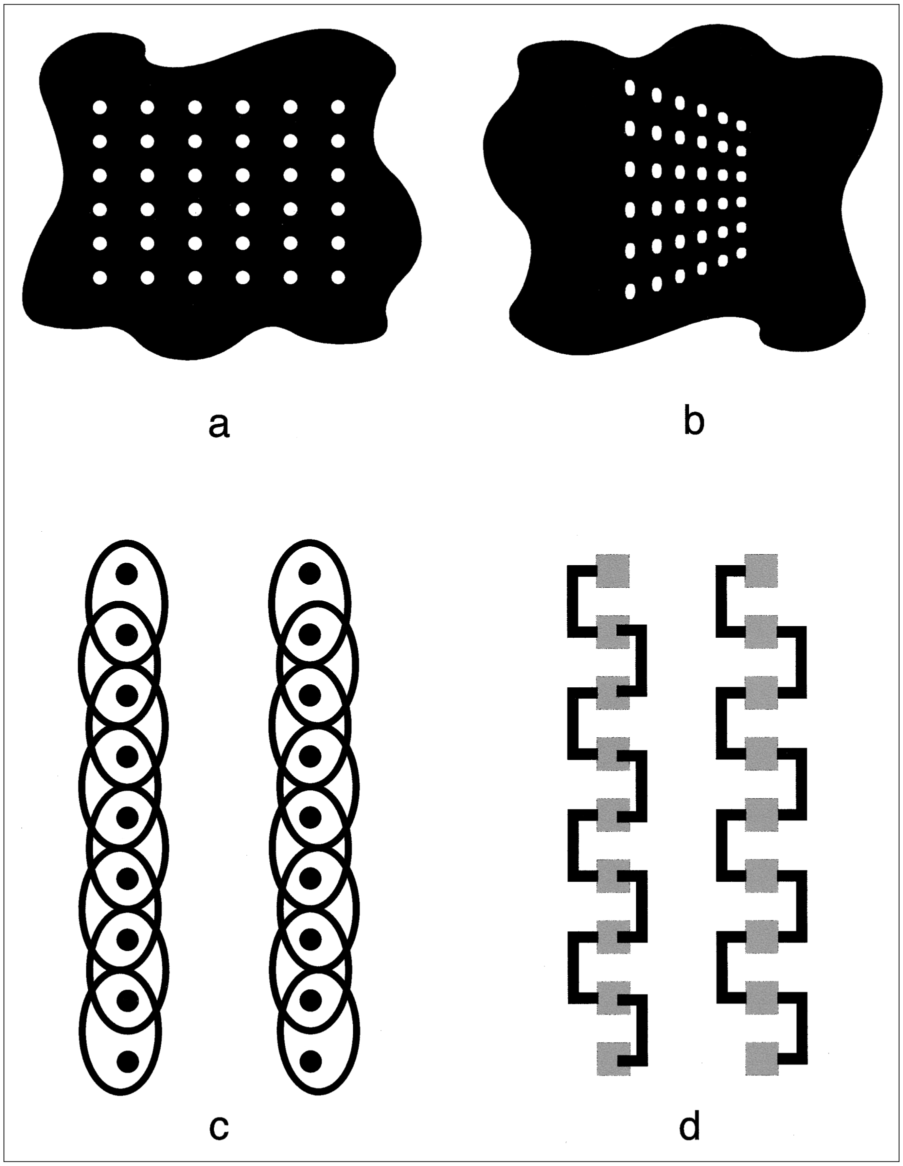

Fig. 1. Depth effects in perceptual grouping by proximity (a, b; adapted from Rock & Brosgole, 1964), common region (c), and element connectedness (d). To fuse the two images in (c) and (d) binocularly with crossed disparity, look between the two images and try to cross your eyes. Fixating on your finger or a pencil point held above the page may help. Cross your eyes to a degree that produces three distinct versions of the original pair. Moving your finger or pencil closer to your eyes or to the page should help you achieve this. The central image is the binocular one and should appear to separate into two distinct depth planes. Panel (a) is reprinted with permission from Palmer (1999, p. 264).

luminous beads either in the frontal plane (perpendicular to the line of sight; see Fig. 1a) or slanted in depth so that the horizontal dimension of the array was compressed more than the vertical (Fig. 1b). The beads were actually closer together vertically than horizontally, so that when they were viewed in the frontal plane, observers always reported seeing them organized into columns rather than rows. The crucial question was how the beads would be grouped when the same lattice was viewed slanted in depth. When the slanted lattice was viewed with both eyes open, the beads were closer together horizontally than vertically as measured in both retinal images, but they were closer together vertically than horizontally in the observer's perception of the physical situation. In this condition, observers still reported seeing the beads organized into columns. Thus, the results showed conclusively that grouping is based on the phenomenally perceived distance between the beads rather than on their retinal distance, supporting the hypothesis that final grouping occurs after binocular depth perception.

Analogous conclusions about the effects of binocular depth are supported for grouping by the factors of common region (Fig. 1c) and element connectedness (Fig. 1d). Each half of Figure 1c alone exhibits no differential grouping of the black circles with one versus the other set of overlapping ellipses. But when the two images are cross-fused binocularly (consult the caption for instructions), the resulting binocular perception shows that the black circles group strongly within the ellipses in the same depth plane and not with the ellipses that float above them in the closer plane. Figure 1d demonstrates the analogous effect for grouping by element connectedness. Once binocular fusion is achieved, the gray squares are seen to group according to the connecting bars in the same depth plane (those on the left side), with the other bars (on the right) floating in a plane above them. Clearly, what matters most is the enclosure and connectedness of the elements in 3-D perceived space, rather than in 2-D retinal space.

LIGHTNESS CONSTANCY

The corresponding question in the domain of lightness perception is whether the important factor in grouping by lightness similarity is preconstancy retinal luminance or postconstancy perceived lightness.

It was answered by studying the effects of cast shadows and translucent overlays on grouping of light and dark squares (Rock, Nijhawan, Palmer, & Tudor, 1992). Observers were shown displays containing five columns of squares (see Fig. 2a) and asked to report whether the central column grouped with those to the left or right. The critical display was carefully constructed so that the central squares were identical in reflectance to those on the left—they were made of the same shade of gray paper—but were seen behind a strip of translucent plastic that rendered their retinal luminance identical to that of the squares on the right (see Fig. 2a). Thus, if grouping were based on retinal luminance, the central squares would be grouped with the luminance-matched ones on the right, but if grouping were based on processing after the transparency of the strip had been perceived, the central squares would group with the reflectance-matched ones on the left. The latter result was obtained. In another condition, the same luminances were achieved by casting a shadow over the central column of squares, and the results were similar. The results for both the transparency and the shadow conditions thus support the postconstancy hypothesis: Grouping is based on perceived lightness rather than on retinal luminance.

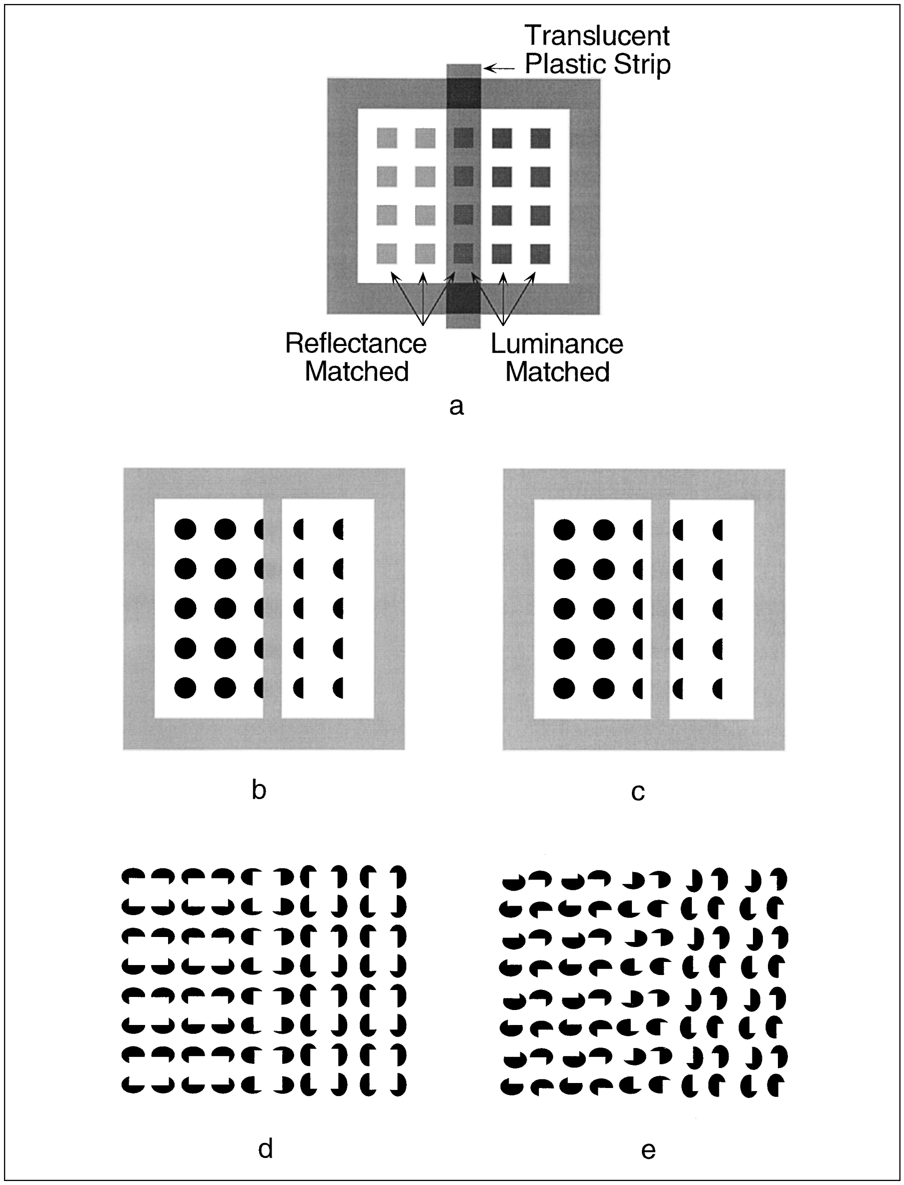

Fig. 2. Stimulus displays used to show that grouping is influenced by lightness constancy (a), amodal completion (b and c), and illusory figures (d and e). In the investigation of lightness constancy (a), the central column of squares matched the reflectance of the squares on one side, but this column was covered with a translucent strip of plastic or viewed with a shadow cast over it, so that its luminance matched the luminance of the squares on the other side (Rock, Nijhawan, Palmer, & Tudor, 1992). The amodally completed half-circles in (b) group with the full circles, but this effect is confounded by common region; the display in (c) shows that moving the occluder slightly further to the side reduces this effect (Palmer, Neff, & Beck, 1996). The vertical illusory rectangles in the central column in (d) group to the right with the other vertical illusory rectangles, rather than according to the orientation of the inducing elements; in a control condition in which the same inducing elements have been rearranged (e), no clear grouping is evident (Palmer & Nelson, 2000). Panels (a), (b), and (c) are reprinted with permission from Palmer (1999, p. 265), and panels (d) and (e) are reprinted with permission from Palmer and Nelson (2000, pp. 1324–1325).

AMODAL COMPLETION

Another experiment examined whether grouping is influenced by amodal completion of partly occluded objects (Palmer, Neff, & Beck, 1996). When a simple object is partly occluded by another, its complete shape behind the occluding object is perceived, although perception does not produce sensory experience of the occluded part of the shape. This process is referred to as *amodal completion* (as opposed to *modal completion*, which produces the illusory figures discussed in the next section) and is widely believed to occur relatively late in perception, presumably during or after the determination of relative depth relations among objects based on the pictorial cue of interposition or occlusion. Is grouping by shape similarity determined by the retinal shape of uncompleted elements, as predicted by the early-only view, or by the perceived shape of completed elements, as predicted by a late view?

Grouping effects involving amodal completion can be measured using the central-column grouping task with a display in which the central column contains half-circles. These are usually perceived as whole circles completed amodally behind an occluding object (see Fig. 2b). The early-only view predicts that the central elements will group with the half-circles on the right; a late view predicts that they will group with the full circles on the left. Clearly, they group with the whole circles. Unfortunately, common region also predicts the same result, so it is necessary to decouple these two factors. This can be done by moving the occluding strip a little further to the side so that the half-circular shape of the central elements can be unambiguously perceived (see Fig. 2c). The results of an experiment varying these two factors independently showed that they both influence perceived grouping. This finding supports the conclusion that grouping by shape similarity is strongly influenced by the perceived shape of amodally completed objects.

ILLUSORY FIGURES

Illusory figures are perceived where inducing elements, such as the notched ovals in Figure 2d, are positioned so that their contours align to form portions of the edges of a closed figure. The completed perception is of a figure that has the same surface characteristics as the background and occludes parts of the inducing elements. Thus, in Figure 2d, the observer perceives an array of white rectangles that occlude parts of black ovals. The crucial question is whether grouping occurs only before the perception of illusory figures, as would be predicted by the early-only view, or whether it occurs afterward, as expected from a late view.

Recent experiments have demonstrated that grouping can occur after perception of illusory figures (Palmer & Nelson, 2000). In the experimental condition, the task was to decide whether the central column in the array shown in Figure 2d groups to the right or the left. The inducing elements were horizontal ovals in the left six columns and vertical ovals in the right four columns. In their unnotched versions, the central two columns of ovals unequivocally group to the left. When the ovals were notched so that illusory rectangles were perceived, the central column of vertical illusory rectangles grouped strongly to the right, with the other vertical illusory rectangles, opposite to the grouping of the inducing elements themselves. A control condition (see Fig. 2e) was tested to be sure that this grouping was not due simply to the nature of the individual notched elements themselves. In this condition, the same elements were slightly rearranged, and equal numbers of observers saw the central columns group to the left and right. The striking difference between the grouping evident in Figures 2d and 2e can be attributed to the fact that grouping is strongly affected by the perception of illusory figures.

THEORETICAL IMPLICATIONS

All of these findings point to the same conclusion: Phenomenally perceived grouping—that is, the final result of underlying grouping processes—is ultimately governed not by the structure of early, preconstancy retinal images, but by the structure of relatively late, postconstancy perceptions. This fact categorically rules out the validity of the early-only view, according to which grouping processes occur only at a 2-D, preconstancy level. The most critical unresolved problem is to determine which alternative theory is correct. There are three types of alternatives, all of which are consistent with the findings described here.

- *Late-only theories*: Grouping processes may work only after constancy has been achieved.
- *Early-and-late theories*: Grouping processes may occur at two (or more) levels, both preceding and following the achievement of constancy.
- *Feedback theories*: Grouping processes may be part of a cascade of temporally overlapping processes that begins prior to constancy operations, but receives postconstancy feedback that alters the initial grouping results.

In both early-and-late theories and feedback theories, early grouping at the image-processing level would provide a preliminary organization that could be used to bootstrap the higher-level processes involved in constancy. The results of these constancy computations would then be used to modify the provisional 2-D organization that resulted from image-based grouping processes, so that the final organization conforms to the perceived, postconstancy properties.

Late-only theories could be categorically ruled out if grouping processes could be shown to operate before as well as after constancy processing. The most promising avenue for establishing this conclusion would be to demonstrate that grouping principles influence constancy processing itself. For example, the following anecdote suggests that the grouping principle of common fate strongly affects lightness constancy. One day I looked up into my gym locker and saw my shirt hanging from a hook at the top. At first, it looked like there was a dark stain on the shirt, spreading down from where it was suspended, as though it had been stained by rust on the hook. When I grabbed the shirt and lifted it, however, the edge of the dark patch did not move upward with the shirt as a stain would have, but stayed fixed relative to the locker, so that it now covered even more of the shirt. I immediately perceived (correctly) that the "stain" was actually just a shadow cast by the top of the locker. If this and similar effects of grouping on constancy stand up to rigorous laboratory tests, then grouping cannot occur only after constancy, which would rule out late-only theories as well as early-only theories. Discriminating between early-and-late theories and feedback theories will be a great deal more difficult, however.

One of the biggest challenges will be understanding the relation of these findings to physiological mechanisms. Given the well-known path of visual information in the brain and discoveries such as finding out which brain cells respond to the contours of illusory figures (von der Heydt & Peterhans, 1989), it is tempting to try to translate the terms "early" and "late" into simple brain locations. The problem is that massive backward connections from higher levels to lower levels throughout the visual system make such translations difficult, if not impossible. Processing that goes on in a given area might be functionally either early or late, depending on whether it happens without or with the benefit of feedback from higher levels and depending on the higher levels from which it might receive feedback. The precise relation between the burgeoning literature on the physiology of the visual system and the kind of functional analysis given here thus constitutes a difficult, but important, area for future research.

Recommended Reading

Palmer, S.E. (1999). Organizing objects and scenes. In *Vision science: Photons to phenomenology* (pp. 254-310). Cambridge, MA: MIT Press.
Palmer, S.E., & Rock, I. (1994). (See References)
Rock, I. (1997). *Indirect perception*. Cambridge, MA: MIT Press.

Acknowledgments—This research and the preparation of this article were supported in part by Grant 1-R01-MH46141 from the National Institute of Mental Health to the author.

Note

1. Address correspondence to Stephen E. Palmer, Psychology Department, University of California, Berkeley, CA 94720-1650.

References

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Palmer, S.E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, *24*, 436–447.

Palmer, S.E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmer, S.E., & Levitin, D. (2002). *Synchrony: A new principle of visual grouping*. Unpublished manuscript, University of California, Berkeley.

Palmer, S.E., Neff, J., & Beck, D. (1996). Late influences on perceptual grouping: Amodal completion. *Psychonomic Bulletin & Review*, *3*, 75–80.

Palmer, S.E., & Nelson, R. (2000). Late influences on perceptual grouping: Illusory figures. *Perception & Psychophysics*, *62*, 1321–1331.

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

Rock, I., & Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology*, *67*, 531–538.

Rock, I., Nijhawan, R., Palmer, S., & Tudor, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception*, *21*, 779–789.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex: I. Lines of pattern discontinuity. *Journal of Neuroscience*, *9*, 1731–1748.

Wertheimer, M. (1950). Laws of organization in perceptual forms. In W.D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 71–81). New York: The Humanities Press. (Original work published 1923)