

Rethinking perceptual organization: The role of uniform connectedness

STEPHEN PALMER and IRVIN ROCK
University of California, Berkeley, California

A principle of perceptual organization, called *uniform connectedness* (UC), is described, and a theoretical approach to perceptual organization is proposed in which this principle plays a fundamental role. The principle of UC states that closed regions of homogeneous properties—such as lightness, chromatic color, texture, and so forth—tend to be perceived initially as single units. We demonstrate its effects and show that they occur even when opposed by powerful grouping principles such as proximity and similarity. We argue that UC cannot be reduced to such grouping principles, because it is not a form of grouping at all. We then propose a theoretical framework within which UC accounts for the initial (or *entry level*) organization of the visual field into primitive units. Classical principles of grouping operate after UC, creating superordinate units consisting of two or more basic-level units. Parsing processes also operate after UC, dividing basic-level units into subordinate parts. UC in the retinal image is proposed to be a necessary, but not a sufficient, condition for unit formation, since connected elements on the retina that are perceived to lie in separate depth planes fail to be perceived as units. This fact, together with other evidence that the Gestalt principles of grouping are based on *perceived* (rather than retinal) relations, suggests that the organization of visual stimulation into UC objects is ultimately achieved within a relatively late, postconstancy representation of environmental surfaces. The implications of this possibility are discussed in light of present theories of visual perception.

In 1923, Max Wertheimer raised the fundamental question of perceptual organization: Why is it that people perceive an organized visual world consisting of discrete objects coherently arranged in space? Without some process of organization, he argued, perception should consist of a chaotic juxtaposition of different elementary sensations based on the intensities and colors registered by individual retinal receptors. Wertheimer recognized that perception cannot be explained by the mere presence of an image containing two-dimensional (2-D) regions that correspond to external objects. Rather, processes of organization within the visual nervous system must be hypothesized as appropriately unifying receptor stimulation that arises from a single object and as separating that which arises from two or more different objects.

Wertheimer's major contribution to the solution of this problem was his proposal that perception organizes stimulation by grouping elements together according to certain properties present in the retinal image. He initially inves-

tigated this hypothesis by using a single line of equally spaced dots (Figure 1A) in which there was no discernible perceptual grouping of the dots into larger groups. By manipulating certain stimulus factors, however, he was able to demonstrate a number of powerful principles of grouping. For example, when he changed the spacing between the dots, making some closer and others farther apart, the closer ones grouped strongly together into pairs (Figure 1B). Wertheimer called this factor of relative closeness *proximity*. Some of the other principles (or "laws") of grouping that he proposed are also illustrated in Figure 1. Figures 1C and 1D show the operation of the principle of *similarity*: all else being equal, the most similar elements (in color and size, for these examples) tend to be grouped together. The law of *common fate* (Figure 1E) is that, all else being equal, elements that move simultaneously in the same direction and with the same speed tend to be grouped together. Another powerful factor that Wertheimer uncovered was the tendency to perceive line segments that are smooth continuations of one another as parts of the same line, and segments that contain abrupt changes of direction as separate lines (Figure 1F). He called this factor *good continuation*. Still another principle he identified is *closure*, which can be seen as overcoming good continuation in Figure 1G, because here the same segments that were grouped according to good continuation in Figure 1F are grouped into closed regions.

Wertheimer's account of perceptual grouping was augmented by the principles of figure-ground organization

This research was supported in part by National Science Foundation Grant BNS-83-19630 to the first author, a National Institute of Mental Health Research Scientist Award to the second author, and National Institute of Mental Health Grant 1-RO1-MH46141 jointly to both authors. We wish to thank John Ceraso, Alan Gilchrist, Philip Kellman, Arien Mack, Mary Peterson, James Pomerantz, Anne Treisman, Carl Zuckerman, and an anonymous reviewer for their perceptive comments on prior drafts of this paper. Reprints should be requested from S. Palmer, Psychology Department, University of California, Berkeley, CA 94720 (e-mail: palmer@cogsci.berkeley.edu).

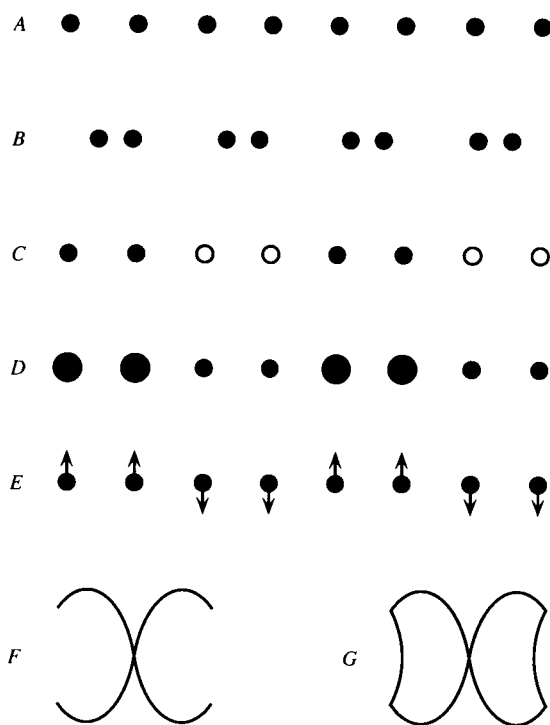


Figure 1. Demonstrations of the classical principles of grouping: ungrouped dots (A), grouping by proximity (B), similarity of color (C), similarity in size (D), common fate (E), good continuation (F), and closure (G).

first described by Rubin (1921) at about the same time. If an array consists of two or more regions, ambiguity remains about which regions constitute *figures* ("things" or objects) and which constitute *ground* (background). To resolve this ambiguity, the contour bounding two adjacent regions must be "assigned" as belonging to one side or the other. The side to which it belongs takes on shape and becomes the figure; the other side becomes a shapeless ground that appears to extend behind the figure as though occluded by it. Certain principles enunciated by Rubin and others (e.g., Kanizsa & Gerbino, 1976) determine which region tends to be perceived as figure and which as ground. All else being equal, the region that is surrounded, is symmetrical, is convex, is oriented vertically and/or horizontally, and contrasts more strongly with the background tends to be perceived as figure.

This Gestalt work on perceptual organization has been widely accepted as identifying crucial phenomena of perception, yet it has had curiously little impact on and integration with modern perceptual theory. Recent textbooks on perception usually present a prominent discussion of both grouping and figure-ground organization, but it is surprisingly similar to the original writings of Gestalt psychologists in the 1920s and oddly detached from the aspects of perception emphasized in more recent work. Indeed, from reading such textbook treatments, it is almost as though Gestalt ideas about organization have existed

in a state of suspended animation for more than 60 years. Although important strides have been made on problems related to perceptual organization (see Banks & Krajiček, 1991, for a recent review), they have yet to be interrelated within a coherent theoretical framework. It is particularly unclear how classical perceptual organization might be related to modern computational theories of vision, which tend to concentrate heavily on seemingly unrelated topics such as the detection of edges, the segregation of textures, the recovery of depth from disparity and shading, and object recognition. Marr's (1982) influential book, for example, mentions grouping and figure-ground phenomena only briefly as laboratory phenomena and assigns them no particularly important role in real-world vision.

The purpose of this article is to discuss a fundamental principle of perceptual organization that we call *uniform connectedness* (UC) and to explore its relation to both Gestalt and computational approaches to vision. Although the Gestaltists did not include this principle in their discussions, and although it has not (to our knowledge) been explicitly discussed since then within the psychological literature on perceptual organization—except for some brief remarks by one of us (Rock, 1975, 1986)—we argue that the principle of UC is strongly presupposed in the Gestalt writing about visual organization. It appears also to be strongly presupposed in modern computational theories of vision (see, e.g., Barrow & Tennenbaum, 1986; Marr, 1982). In fact, we see it as forming a bridge between the original Gestalt work and modern information processing work.

THE ORGANIZATIONAL PRINCIPLE OF UNIFORM CONNECTEDNESS

Simply stated, the proposed principle of UC is that a connected region of uniform visual properties—such as luminance or lightness, color, texture, motion, and possibly other properties as well—strongly tends to be organized as a single perceptual unit. In the present paper, we will first argue for the theoretical relevance of UC from the standpoint of what it contributes to perceptual organization. We will then demonstrate its effects in displays such as Wertheimer used. Next, we will attempt to integrate these ideas with each other and with well-known organizational phenomena by presenting a larger theoretical framework. In this theory, the principle of UC plays the crucial role of providing *entry-level* units into the part-whole hierarchy of perceptual organization. We will discuss its relation to other important concepts in visual perception, including the classical principles of grouping, parsing, edge detection, texture segregation, depth perception, and motion perception. Finally, we will discuss levels at which various principles of organization might have their effects, particularly with respect to the achievement of perceptual constancy.

Functional Importance

The relevance of UC to perceptual organization is relatively straightforward, if one presumes that the primary

function of perceptual organization is to designate areas in the image that correspond to unified, distinct objects in the 3-D environment (see, e.g., Bregman, 1981; Shepard, 1981). Certainly one of the most important features of what people tend to perceive as a single object is that it is a connected piece of matter, one that could be moved (at least in principle) as a unit. For our present purposes, the important empirical fact about such connected objects—or about significant parts of them—is that they tend to have relatively uniform surface characteristics. That is, they tend to have (approximately) the same lightness, color, texture, and possibly other properties as well. Taken together, then, the connectedness and surface uniformity of environmental objects and parts suggest a seemingly obvious early step in finding the corresponding objects and parts in the image: simply identify UC regions in the retinal image. The rationale is that, all else being equal, UC objects or parts tend to project UC regions onto the retina, so that UC image regions tend to correspond to UC objects or parts in the environment.

Retinally UC regions are not fool-proof evidence of connected objects or parts, however, because the projection from environmentally UC objects and parts to retinally UC regions is not necessarily one to one. Violations occur in the form of both many-to-one and one-to-many mappings.

Many-to-one mappings. Sometimes more than one distinct UC object in the environment projects to a single UC region in the image. This occurs mainly in cases of successful camouflage when, for example, part of an animal merges with part of its background to form a uniform retinal region that corresponds partly to the animal and partly to the background. Camouflage notwithstanding, many-to-one mappings are relatively infrequent.

One-to-many mappings. Much more frequently, a single UC object in the environment fails to project a single UC region in the image. This can happen in three different ways, depending on which condition fails to be satisfied: uniformity, connectedness, or both.

Disconnectedness. UC objects or parts often project to two or more uniform retinal regions that are not connected due to occlusion by another object. The occluding object's projection then separates the projection of the occluded object into two or more distinct regions which must be unified by some later organizational process.

Nonuniformity. UC objects or parts also frequently project to a single connected region that is not uniform in its sensory qualities, because of cast shadows, reflective highlights, and/or changes in surface orientation. When such changes are relatively abrupt (e.g., cast shadows), they result in several distinct regions that must be unified by some later organizational process. When they are gradual (e.g., the shading of a curved surface), they result in regions characterized by slowly changing gradients that may still be identified as single regions (see below).

Despite these more complicated cases, UC of retinal regions is such a useful heuristic for perceptual organization that it is difficult to imagine the visual system fail-

ing to exploit it. At the end of this article, we will discuss how the many-to-one and the one-to-many problems might be solved by further organizational processes.

Demonstrations of Uniform Connectedness

We now offer some examples of the principle of UC at work in displays similar to Wertheimer's original ones. Figure 2A shows that a line of equally spaced dots is strongly grouped into pairs when subsets of the dots are connected by short horizontal line segments. Figures 2B–2D show how powerful UC can be in comparison with some of the classical principles of grouping. Figure 2B demonstrates that it can overcome the law of proximity, and Figure 2C demonstrates the same for size similarity. Figure 2D shows that it can even overcome the combined effects of both factors, which, without the connecting segments, produce the opposite grouping very strongly.

Perhaps the most salient feature of these displays is that each UC element in Figure 2 appears to constitute a single perceptual object with multiple parts. We take this phenomenological fact very seriously, claiming that UC is such a strong factor in perceptual organization that it virtually *defines* perceptual elements. Surprisingly, Wertheimer never analyzed the basis for the "elements" on which grouping factors operate. In fact, most (but not all) of his elements are actually UC regions. This is true for all the examples with arrays of separated elements, as shown in Figures 1A–1E. The other "elements" he assumed in his analysis were actually *parts* of connected regions, such as segments of connected line drawings whose ends were defined by terminations, intersections, or abrupt changes in direction (e.g., Figures 1F–1G). In our view, the latter elements would have to be analyzed as divisions of a single perceptual unit into two or more parts which could then be grouped in different ways. We will discuss such cases in detail later in this article. The important point for the present discussion is that in neither case does UC enter explicitly into Wertheimer's analysis, although it seems to be strongly presupposed.

Relation to Classical Grouping Principles

It is more than a little surprising that Wertheimer failed to provide an explicit analysis of the stimulus basis for the elements in his displays. One possibility is that Wertheimer believed they were somehow given in the stimulus itself. If so, he would have been guilty of making the very "experience error" for which the Gestaltists so frequently faulted others.¹ Even sophisticated psychologists who accept the logic of the experience error, at least insofar as it applies to Wertheimer's grouping or Rubin's figure-ground patterns, may balk at the argument that principles of organization are required to explain why a dot or line segment appears as a perceptual unit, for here the ambiguity (and thus the alternative percept) is not obvious. How else, one might ask, could a dot or line be expected to appear?

The fallacy in this way of thinking is, again, precisely the experience error, for the image of a dot or line is itself no more organized than the image of one of Wer-

theimer's arrays of dots or lines. That is, there is no more intrinsic relation among rays of light that happen to come from the same homogeneous object in the world than there is from points of light some of which come from that object and some of which come from another object. All perceptual organization is thus properly conceived as an achievement of the viewer's visual system, albeit an achievement based on structure *implicit* in the stimulus image. Hence, even the perception of a homogeneous dot against a uniform background requires explanation in terms of some process of organization.²

A second possible explanation for Wertheimer's not providing an explicit analysis of the stimulus basis for the elements in his displays is that he believed his grouping principles to be sufficient to explain their perception. One might maintain, for example, that UC is nothing more than a special case of proximity. After all, when elements are connected, as in Figure 2, it necessarily creates maximum proximity between the connector and the connected. Nevertheless, there is a clear topological distinction between connectedness and mere proximity,³ and the burden of proof should therefore rest on the advocates of reducing connectedness to proximity. Köhler (1926) attempted to provide such an argument, offering the progression from A to B to C in Figure 3 as evidence. The relative proximity of elements within each of these clusters increases, he argued, until one reaches the case of complete homogeneity in C. But this argument is compelling only if one accepts the premise that there are component *elements* in the case of Figure 3C smaller than the entire UC region. We argue that such smaller elements simply do not exist—at least not phenomenologically—in the case of a UC region. Although Köhler showed that a continuum of proximity could be used to *construct* this sequence of images, this does not necessarily imply that there is a corresponding continuum involved in the mechanisms underlying their *perception*.

In any case, we suggest that calling connectedness a special case of proximity is putting the cart before the

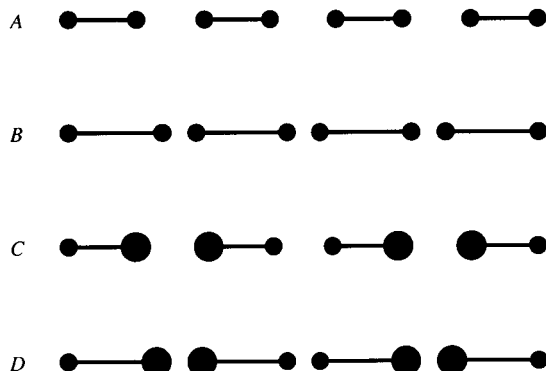


Figure 2. Demonstrations of uniform connectedness as a principle of perceptual organization (A) and of its ability to overcome such classical grouping factors as proximity of the dots (B), similarity in size (C), and their combination (D).

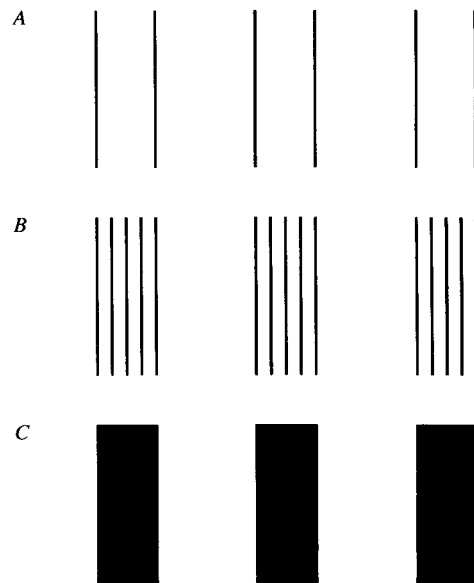


Figure 3. Köhler's argument that uniform connectedness can be reduced to extreme proximity. Grouping by proximity in part A is strengthened by adding more lines in part B, until eventually a homogeneous connected region is formed in part C (Köhler, 1926).

horse: if the present proposals are correct, proximity is more appropriately considered an approximation to connectedness. As we argued earlier, it actually makes good a priori sense for connectedness to be the foundation of perceptual organization, because what people consider to be distinct physical objects are actually *defined* by physical connectedness, at least at a macroscopic level. What really "goes together" in the strongest physical sense is a single connected piece of matter, not separate ones that happen to be near each other in space. Parts of objects that are actually connected are thus much more closely coupled in their physical behavior than are two unconnected objects that are nearby, no matter how nearby they may be. This argument is couched in terms of the properties of the distal stimulus, but, as we have argued, connectedness (and proximity) will be reflected in the structure of the image on a contingent, probabilistic basis.

The other classical principle of grouping to which one might attempt to reduce connectedness is similarity, substituting "similarity" for "uniformity" or "homogeneity" in our statement of this principle. The concepts are indeed related, but the crucial difference is based on the distinction between separate as opposed to continuously connected regions of similar properties. If elements are separated, as in Wertheimer's examples (see Figures 1C and 1D), one can speak meaningfully of similarity in color among elements. Within a single element, however, to speak of "grouping by similarity" seems inappropriate, because (we claim) there are no separate elements within a UC region. It is simply a single element on its own. Moreover, the organization achieved within a single region of uniform lightness or color is phenomenologi-

cally different from grouping together several such regions, as one can appreciate by comparing A and B versus C in Figure 3. How does this *qualitative* difference arise if degrees of proximity and similarity underlie their perception? This is a puzzle for which the grouping account has no obvious answer.⁴

More importantly, there is a simple and straightforward reason to reject the claim that UC can be reduced to a grouping principle: *UC is not a principle of grouping at all*. At least according to our analysis, grouping requires previously defined elements that can be put together into a higher order perceptual unit. But UC does not require any such “putting together” of separate elements, for (we claim) there are no such separate elements prior to its application. Rather, the operation of finding UC regions implies both *defining* a connected region of uniform perceptual quality in the unorganized image and *designating* it as a single entity. It is this operation of initial unit formation—however it might be implemented—that produces perceptual units in the first place. Thus we see a profound difference between organization by UC and the classical principles of grouping: UC operates on an uninterpreted sensory image, whereas classical principles of grouping require prior organization into elements or units. (We do not claim that all possible approaches to grouping necessarily rest on the assumption that grouping occurs by putting together distinct perceptual elements, but only that this is a sensible definition within the present theoretical framework.) It is for this reason that we use the term *organization* instead of “grouping” to describe the principle of UC.

Relation to Edge Detection

As we have described the “principle” of UC, it can be considered part of what Marr (1982) called a “computational level” constraint on a theory of object formation. That is, we have as yet made no attempt to define the algorithm by which UC regions might actually be found: we have merely hypothesized that UC regions provide important information for designating objects or significant parts of objects.

There are at least two plausible algorithms for identifying UC regions. Perhaps the most obvious, given the appellation of “uniform connectedness,” is *region growing*: explicitly detecting the *sameness* (or high similarity) of adjacent pixels in the image and “growing” regions outward until that sameness diminishes abruptly at a region boundary. Yet although this is a logical possibility, there are good reasons to believe that the visual system actually uses the more efficient, but indirect method of *edge detection*: the process of explicitly locating the spatial positions at which sharp differences in image luminance arise along the boundaries between regions (see, e.g., Hubel & Wiesel, 1968; Marr & Hildreth, 1980).

Edge detection algorithms can thus be construed as the “front end” of a boundary-based theory of finding UC regions. Whenever luminance edges form a closed contour, they necessarily define two regions: the inside and the outside. An image can therefore be segmented into

a set of connected regions by using an edge-detection algorithm to locate closed contours. This observation suggests that edge detection constitutes a very early stage in perceptual organization by implicitly defining UC regions. Once such regions have been explicitly designated by further processing, they can be subjected to figure-ground organization and classical grouping processes, as we will discuss at length later in this article. Edge detection can thus be understood as an early step in perceptual organization, one that begins the process of partitioning the image into connected regions of uniform sensory properties.

It is worth remarking that boundary-based algorithms for region segmentation are radically different from the sort of processes Gestalt theorists seemed to envision as underlying grouping phenomena. Wertheimer’s demonstrations and discussion of grouping by similarity suggest a process of “putting together” a set of discrete object-like elements, perhaps by some process of “mutual attraction” due to common properties. This is quite unlike the boundary-finding approach to region segmentation in which there are no initial “elements” at all—only an unstructured image—and in which the regions that are formed are essentially “by-products” of detecting edges. We believe that this distinction between the Gestalt conception of “aggregating separate elements” and the modern conception of “dividing an unorganized image” is theoretically important, and that both sorts of processes probably occur, albeit in different situations. We will return to this issue later.

The claim that edge detection is the principal mechanism for segmenting the image into regions is supported by a number of phenomena, but most dramatically by evidence from *stabilized images*. It is well known that the visual system adapts to optical structure in the retinal image if there is no change over time, with perception disappearing completely after a short period of stabilization. The relevance of stabilized images to the claim that edge detection is the mechanism of finding UC regions is the demonstration that experiences of regional shape and color arise from edge information. Among the most striking demonstrations is one by Krauskopf (1963). He presented observers with a central red circle surrounded by a green annulus. Under normal viewing conditions, this is just what observers perceived. But when Krauskopf stabilized the inner contour between the red and green regions of the visual field, while keeping the outer one unstabilized, the inner red disk disappeared and was “filled in” by green to create the perception of a single large green circle (Krauskopf, 1963). This dramatic result is easily explained if edge detection is the basis of region perception and if adaptation to stabilized edges effectively “erases” them.

A potentially important observation bearing on the proposal that regions are defined by the principle of UC is that regions do not have to be strictly uniform in luminance or color to be perceived as units. Figure 4 shows an example in which a rectangular region containing a dark-to-light luminance gradient is surrounded by another region containing the same dark-to-light luminance gra-

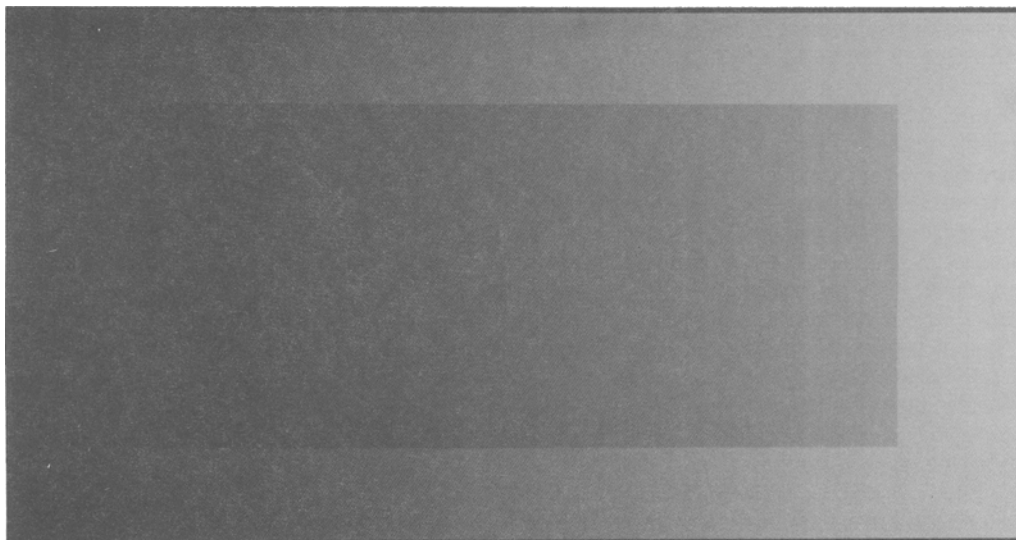


Figure 4. An example of perceptual organization into two regions, both of which contain smooth luminance gradients rather than uniform areas of lightness.

dient displaced laterally. Neither region is uniform, yet they are clearly perceived as two distinct regions. It seems likely that areas containing slow gradients are identified as unified regions simply as a by-product of the edge-detection procedure for finding UC regions. In this view, slow gradients are treated much the same as UC regions because the edge-detection algorithm is simply not sensitive to them. This account is thus compatible with—although not predicted by—the principle of UC as described above.

If edge detection is indeed the first step in the algorithm for finding UC regions, it might be tempting to think that there is no need for the principle of UC in a theory of vision. There are several reasons why this is not the case, however. First, edge detection cannot be equated with the principle of UC because they are at different conceptual levels. As stated above, edge detection is only one possible way in which the principle of UC might be realized. If there are other possible realizations, UC must be a superordinate theoretical concept. Second, edge detection cannot be equated with the principle of UC because the former is only *part* of the processes required by the latter. Although edge detection may well be the earliest and best understood step in the identification of UC regions, it is by no means the only one. Local edges must be integrated into global contours, closed contours must be identified, and the regions they enclose must be designated as potential objects. Modern computational theorists have concentrated heavily on formal methods for reliably detecting luminance edges (e.g., Marr & Hildreth, 1980). Some have discussed additional processes that would be necessary in identifying UC regions, such as linking local edges (e.g., Zucker & Hummel, 1979) and designating the interior of closed contours as regions (e.g., Grossberg & Mingolla, 1985; Ullman, 1984). But there have

been few, if any, discussions of why these processes are important in the larger context of organizing images into perceptual objects.

We believe that the principle of UC is the missing link between the psychological literature on perceptual organization and the computational literature on scene understanding. We have already argued that UC is necessary to make sense of the Gestalt literature on grouping processes. UC appears also to be strongly presupposed within computational approaches to vision,⁵ but has seldom been explicitly discussed. Perhaps this is because it seems too obvious to mention. In contrast, however, we believe it is too important *not* to mention, because it provides a major rationale for processes such as edge detection, edge integration, region coloring, and the like. Without explicating something akin to the principle of UC, it is easy to miss the forest for the trees.

As the foregoing comments imply, our primary goal in the present paper is to articulate a coherent theoretical framework for perceptual organization that encompasses both Gestalt and modern computational traditions. Many, if not all, of the necessary building blocks are present in the literature already in one form or another. What is missing is the big picture: how they are interrelated and why they go together in this particular way. We believe that the theoretical structure we describe provides a clearer and more coherent framework for understanding perceptual organization than has previously been available.

The Question of Levels

The foregoing discussion raises an important question about organizational processes: at what level of the visual system do they occur? Surprisingly, this problem has received relatively little attention in the years since the Gestaltists first proposed the concept. The “standard”

view, based on Wertheimer's (1923) original paper, appears to be that the initial units of perceptual organization are determined by the classical principles of grouping, although it is far from clear how this might occur. In contrast, we have argued that region-finding processes implementing the principle of UC perform the initial structuring of the image into perceptual units. This alternative view raises the question of whether classical grouping processes actually occur as early in vision as has commonly been supposed.

Grouping processes are usually assumed to be relatively primitive, "low-level" perceptual operations that work on an early, image-based representation. The logic behind this view is that because organization into elements must occur before any process that requires discrete elements as input, and because Gestalt grouping is presumed to be the basis of this organization, grouping must operate very early. This view was implicit in Wertheimer's original discussion and has seldom been seriously questioned. The present claim that the identification of UC regions constitutes the first stage of organization, however, implies that grouping processes need not be the early image-based processes they have previously been assumed to be. Fortunately, the level at which grouping processes occur is an empirical question that can be answered by appropriate experiments.

In fact, the preponderance of experimental evidence conflicts with the standard "early" view of grouping. Rock and Brosgole (1964), for example, found that proximity grouping was based on the *perceived 3-D separation* of elements rather than on their proximity defined at the level of the retinal image. They used a 2-D rectangular array of luminous spheres (Figure 5A) which was presented to the observer in a dark room either in the frontal plane (perpendicular to the line of sight) or slanted in depth so that the horizontal dimension was foreshortened to a degree that depended on the angle of slant. The beads were actually closer together vertically, so that when they were viewed in the frontal plane, as illustrated in Figure 5B, observers always reported them as grouped vertically into columns rather than horizontally into rows.

The crucial question was what happened when the same lattice of beads was presented to the observer slanted in depth so that the beads were closer together horizontally when measured in the retinal image, as depicted in Figure 5C. When observers achieved veridical depth perception by viewing the slanted display binocularly, they reported seeing it as organized into *columns*, just as they did in the frontal viewing condition. Thus, when depth information is perceived veridically, grouping is consistent with a relatively late, postconstancy representation rather than with a relatively early preconstancy representation. This result therefore supports the hypothesis that grouping occurs after constancy due to stereoscopic depth perception.

Rock, Nijhawan, Palmer, and Tudor (1992) addressed the analogous issue in achromatic color perception: namely, is the important factor in color similarity grouping the retinally measured *luminance* of elements or is

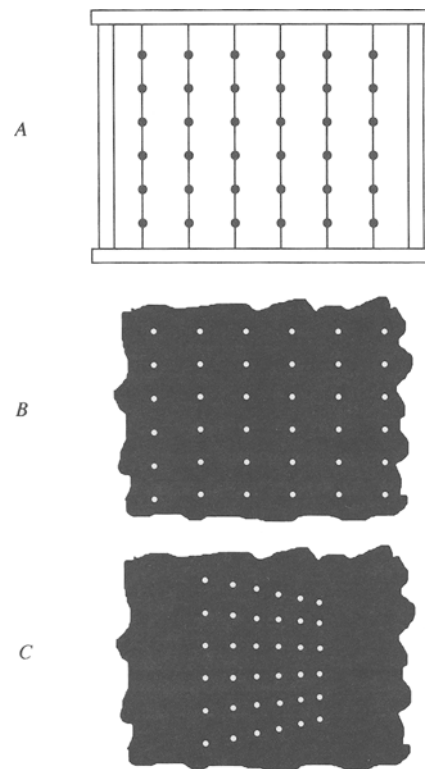


Figure 5. Apparatus used by Rock and Brosgole (1964) to investigate whether grouping by proximity is based on retinal or perceived proximity. Small luminous spheres were strung on vertical wires attached to a frame (part A). The spheres were spaced closer together vertically than horizontally. When presented in a dark room, only the spheres were visible. Part B shows the appearance of the display when the frame is in the frontal plane. Part C shows the appearance when the frame is swiveled about a vertical axis so that the retinal images of the spheres are closer together horizontally. (See text for details.)

it their phenomenally perceived *lightness* after lightness constancy has been achieved? They used cast shadows and translucent overlays to decouple image luminance and perceived lightness. Observers were shown displays like the one illustrated in Figure 6 and were asked to indicate whether the central column of elements was grouped with the ones on the left or on the right. The critical display was carefully constructed so that the central squares were identical in reflectance to the ones on the left (i.e., they were made of the same shade of gray paper), but were seen under a shadow cast by an opaque vertical strip hanging nearby. As a result, the luminance of the central squares was identical to the luminance of the squares on the right. Thus, if grouping were based on relatively early processing of image structure, the central squares would be grouped with the luminance-matched ones on the right. If it were based on relatively late processing after veridical perception of shadows had been achieved, they would group with the reflectance-matched ones on the left. The results showed that grouping followed the predictions of the postconstancy grouping hypothesis: similarity group-

ing was governed by the perceived lightness of the squares rather than by their retinal luminance. The same result was obtained when the central squares were displayed behind a translucent strip of plastic instead of in shadow. Control conditions effectively ruled out an alternative explanation that luminance ratios were responsible for the obtained results.

Perceptual grouping by shape similarity also appears to occur relatively late in perception. Palmer (1994) showed that such grouping is affected by visual completion, a process that is generally thought to occur fairly late in processing. Using the same type of displays as Rock et al. (1992), Palmer constructed a display in which half circles in the center column were perceived as whole circles partly occluded by a vertical strip, as shown in Figure 7A. An early view of grouping would predict that the central elements will still be seen to group with the half circles on the left because both have the retinal shape of a half circle. A late view of grouping would predict that they will group with the full circles on the right because they will be perceived as completed behind the occluding strip. The central figures group themselves to the right with the full circles, indicating that grouping is based on similarity of completed shape rather than on retinal shape. The possibility that this outcome was determined by the presence of the occluding strip, which divides the elements into two regions according to grouping by "common region" (Palmer, 1992), can be ruled out by the control

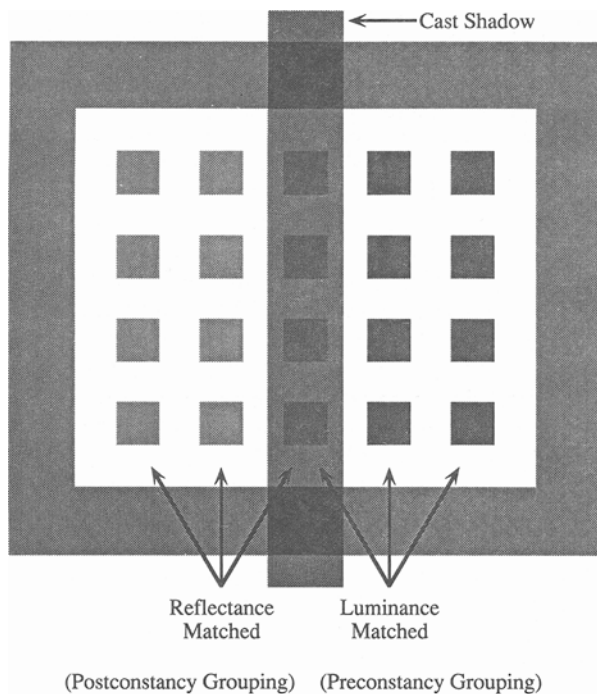


Figure 6. Stimuli used by Rock, Nijhawan, Palmer, and Tudor (1992) to demonstrate that grouping by achromatic color is based on perceived lightness rather than on retinal luminance. See text for details. From "Grouping Based on Phenomenal Similarity of Achromatic Color," by I. Rock, R. Nijhawan, S. Palmer, and L. Tudor, 1992, *Perception*, 21, 782. Copyright 1992 by Pion, Ltd., London. Reprinted by permission.

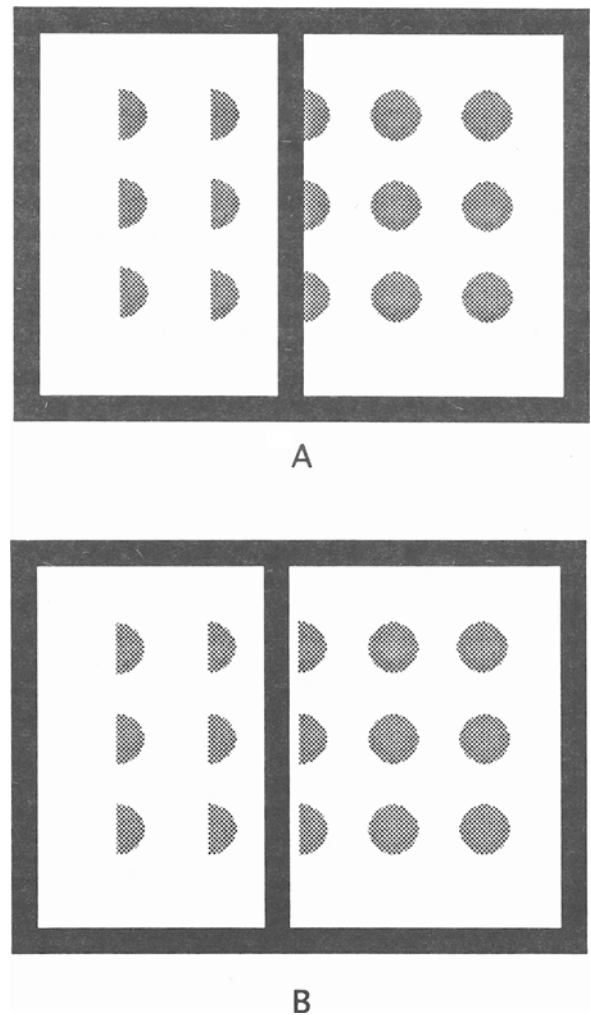


Figure 7. Stimuli used by Palmer (1994) to demonstrate that grouping by shape is based on the completed, perceived shape rather than on the retinally projected shape. (See text for details.)

condition illustrated in Figure 7B. The occluding strip is simply moved a little farther to the side to reveal the entire contour of the central elements, allowing their half-circular shape to be perceived unambiguously. Now the same central elements tend to be perceived as grouped with the half circles on the left.

Peterson and her colleagues have recently reported results consistent with the possibility that figure-ground processes operate at a fairly late stage of processing (Peterson & Gibson, 1991, in press; Peterson, Harvey, & Weidenbacher, 1991). They have found that figure-ground perception can be influenced by the "denotivity" (roughly, the meaningfulness) of the regions being viewed. For example, subjects are more likely to initially perceive meaningful regions (e.g., the silhouette of a woman's figure or a person's face) as figure, and they do so for a longer period of time than they do with nonmeaningful regions (e.g., the same silhouettes presented upside-

down). In order for figure-ground perception to be influenced by such a presumably late process as object recognition, it seems likely that figure-ground segregation also occurs relatively late in perception.

The evidence just reviewed thus favors the hypothesis that the final processes of organization (grouping and figure-ground organization) do not occur at an early, preconstancy level in which the properties of the retinal image govern the outcome, but at some relatively late, postconstancy level of perception, after stereoscopic depth information, lightness constancy, and occlusion information have been extracted.⁶ Thus organization based on the classical Gestalt principles seems to be governed by the *perceived* structure of surfaces and objects in the 3-D world rather than by the retinal structure of the proximal stimulus. Of course, it is entirely possible—perhaps even likely—that organizational processes may operate at *both* pre- and postconstancy levels, with the later results dominating conscious perception. We are currently undertaking experiments designed to determine whether early effects of grouping are present as well as late ones.

The following phenomenon leads us to believe that UC must also be influenced by relatively late perceptual processes. To the extent that it is, organization by UC should not be identified *solely* with 2-D image processing, such as detecting closed luminance edges, although this may indeed be the first important step in the process.

The phenomenon, illustrated in Figure 8, demonstrates that the final organization according to the principle of UC occurs after binocular depth perception. It shows that if dots are retinally connected by arcs in another depth plane as well as physically connected by arcs in their own depth plane, one perceives the dots as grouped according to their perceived connectedness within their own depth plane. To experience this phenomenon, trace or photocopy the disconnected U-shaped arcs in Figure 8A onto a clear plastic transparency and slide it over the display of dots shown in Figure 8B. You should perceive a smooth undulating curve with equally spaced dots on it, as illustrated in Figure 8C. No selective grouping of the dots according to either the upright or inverted Us is perceived. (Although good continuation of the arcs is a fac-

tor here, it is irrelevant to the demonstration.) To perceive the depth effect, simply lift the transparency a short distance above the page, being careful to keep the arcs aligned with the dots by alternately closing each eye. When this depth display is finally viewed with both eyes at once, the dots spontaneously organize themselves into units consisting of the pairs connected on the page via the inverted Us. The same demonstration can be reversed—that is, trace the dots and arcs of Figure 8B and place them over Figure 8A—to show that it is really perceived connectedness that matters, not just which arcs are in the back as opposed to the front plane.

From this demonstration, we conclude that retinal connectedness is not a *sufficient* condition for perception of a single element, because the perceived 3-D structure of the scene can override or at least influence it. Thus, the situation with respect to UC appears to be analogous to the previously mentioned classical principles of grouping: it is not *exclusively* retinal connectedness that governs initial organization into elements, but perceived connectedness at some higher level in the visual system, one that occurs at or after stereoscopic depth perception.

It is important to exercise care in interpreting such postconstancy effects. One possibility is that phenomena of UC may occur exclusively at a later, postconstancy level. This leads to a logical problem, however, in that it requires the postconstancy representation to be constructed without the benefit of identifying regions defined by UC in the retinal image. This seems highly unlikely, given our current understanding of how constancy might be achieved. A more reasonable possibility is that UC may work both at an early image-processing level, prior to achieving depth and constancy, and also at some later surface-processing level, after stereoscopic depth and constancy have been achieved. The same two interpretations are, of course, possible for the grouping phenomena described earlier in this section: grouping might occur only at the later, postconstancy level or it might occur at both levels, with the later postconstancy level determining the ultimate perceptual experience. We will return to these issues later in this article.

TOWARD A THEORY OF PERCEPTUAL ORGANIZATION

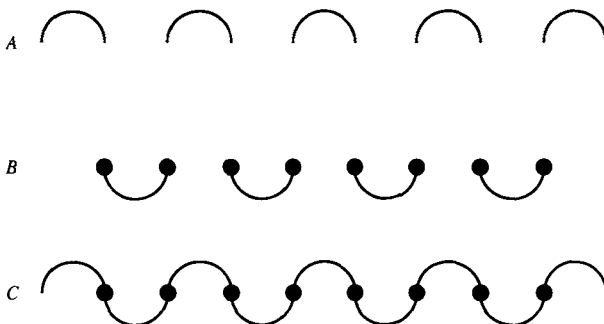


Figure 8. A demonstration that organization according to connectedness occurs during or after stereoscopic depth perception. (See text for details.)

Logic dictates that some amount of visual organization must occur at an early stage in visual processing and that it must occur preattentively. As Neisser (1967) and Treisman (1986) have put it, discrete perceptual elements of some sort must first be present to serve as candidates for further element-based processing. Only after such elements are present can we attend selectively to one or another, recognize them if they are familiar, learn about them, remember them, and so on. We have seen, however, the possibly contradictory fact that some of the classical Gestalt principles of grouping—and perhaps the principle of UC as well—operate at or beyond the level at which depth and constancy are achieved. To the extent that these observations are correct, they require a thor-

ough reexamination of the theoretical underpinnings of perceptual organization.

In the present section, we will sketch the outline of a theory that seems consistent with the evidence we have described and that places the principle of UC at center stage. It is based on two major ideas:

1. Entry-level organization: UC occupies a privileged position in perceptual organization by virtue of defining the primary or *entry-level* units on which further organizational processes operate. Once established, entry-level units can then be either parsed into *subordinate* units (parts) or grouped into *superordinate* units (wholes or aggregates).

2. Postconstancy representation: Consciously perceived organization reflects the structure of a relatively late representation, in which depth, occlusion, and the various constancies have already been achieved. Organization at this level supersedes any achieved at earlier, preconstancy levels in the visual system.

We will discuss both of these ideas in turn and elaborate on some of their implications for a theory of perceptual organization.

Entry-Level Units

We begin with the premise that the result of perceptual organization is a part-whole hierarchy of perceived units or elements at different size scales (see, e.g., Biederman, 1985; Marr & Nishihara, 1978; Palmer, 1975, 1977; Winston, 1975). These units may correspond to individual texture elements, object parts, whole objects, aggregations of many objects, or even complex extended scenes. All of the entities in this hierarchy are thus perceptual units of some description, yet they must arise from an initially unstructured retinal image. How might this be accomplished?

In our theory of perceptual organization, UC plays the special role of providing entry into the part-whole hierarchy directly from the stimulus image, without any requirement that there be previously defined "elements" or "units." We will call any level from which image information provides direct access into the part-whole hierarchy an *entry level*. Eventually, we will propose that there may be several entry levels, all of them defined by the principle of UC as applied to different image properties. But it will be easier right now to pretend that there is a single entry level.

Once the entry-level units have been constructed, other levels in the part-whole hierarchy can be derived by further processes that operate on these entry-level units. They can be further structured in the upward (larger scale) direction by putting several units together via grouping processes. They can also be further structured in the downward (smaller scale) direction, by dividing a basic level unit into subparts via parsing processes. Both possibilities are illustrated in Figure 9 for a simple example. The claim is that the first organization to be achieved comprises the six double-lobed regions, plus the single background region, as defined by the principle of UC. Figure-ground processes would then be required to designate the

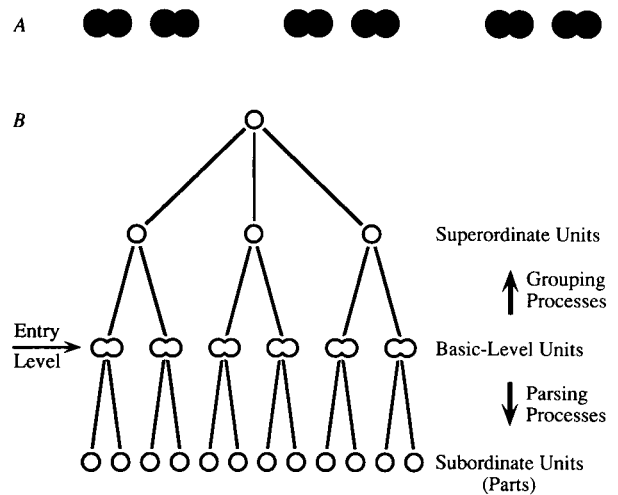


Figure 9. The proposed relation among entry-level organization, grouping processes, and parsing processes within the part-whole hierarchy. The nodes in the hierarchical network correspond to units involved in the perception of the pattern of bilobed elements shown at the top. These elements can then be divided into subordinate units (parts) by parsing processes or unified into superordinate units (aggregations) by grouping processes.

six black regions as figures and the surrounding white one as ground. The six entry-level figures can then be divided into subordinate level units—the component convex lobes in Figure 9—via *parsing processes* like those proposed by Hoffman and Richards (1984). The same entry-level elements can also be grouped into superordinate units—the pairs of nearby bilobed shapes in Figure 9—via *grouping processes* such as Wertheimer (1923) proposed in his classic paper. We will discuss each of these further operations in detail shortly.

Both parsing and grouping operations produce phenomenal outcomes that are qualitatively different from entry-level organization. When perceptual elements at the entry level are divided into constituent parts, as in the two lobes of the black figures in Figure 9, the incomplete separateness of the parts is qualitatively different from the complete separateness of the dots in Wertheimer's classic proximity example (Figure 1B). In other words, the "parts" of the objects in Figure 9 are clearly perceived as subsets of a single connected object, which is simply not the case for the dots in Figure 1. Similarly, when several separate entry-level elements are grouped together, as in Figure 1B, the resulting aggregation of objects is qualitatively distinct from the component objects themselves. We have argued, for example, that the most basic elements in Wertheimer's displays are the UC regions, such as dots and lines. After grouping has taken place, these entry-level elements continue to have a strongly perceived identity despite their organization into a looser, superordinate aggregation. The dots of Figure 1B, for instance, are indeed perceived as grouped into pairs, yet one continues strongly to perceive the individual dots as separate objects. This fact clearly implies a

part-whole hierarchy of perceptual organization in which the individual dots make up the lower level and the groupings of dot pairs make up the higher level. All these observations about the distinctiveness of entry-level elements support the fundamental nature of organization by UC.

Thus we are suggesting that the entry-level organization of the two-dimensional retinal image is initially accomplished by applying the principle of UC—that is, by identifying connected regions of homogeneous sensory qualities, such as luminance and/or chromatic color. (Other properties will be considered later, including texture, motion, and depth). As we have already noted, this “region-finding” process is not a “grouping” process like that which Wertheimer discussed, because, prior to its operation, there are no independent “elements” or “units” to be grouped; there is simply an unstructured image. Later processes of organization, then, are responsible for constructing higher and lower levels of the organizational part-whole hierarchy by dividing or unifying the entry-level units defined by UC. As we mentioned earlier, the organization of an image into entry-level units may well take place via the operation of edge- and line-detection algorithms analogous to those devised by computer vision theorists (e.g., Marr & Hildreth, 1980). But for our purposes, it is not crucial what mechanism leads to the designation of UC regions. What matters is only that the rough uniformity of stimulation within regions is detected by the perceptual system, however that may be accomplished.

Figure-Ground Processes

The mechanisms underlying UC divide the stimulus image into some number of discrete regions, but none of them yet constitutes a perceptual “element” or “entry-level unit,” for this status implies that they have been further identified as figures rather than background. We therefore conclude that the next organizational process at work must separate figures from ground. The principles identified by Rubin (1921) and others (e.g., Kanizsa & Gerbino, 1976) represent at least the beginning of a functional (or “computational”) theory concerning the informational basis of figure-ground segregation. They do not specify an algorithm by which this might be accomplished, however. A great deal of further work is required to define the processes that underlie this important step. A method for integrating the results from applying the various different factors is also required. We will not attempt to solve such problems here, but merely note that such processes must occur after initial regions have been identified in order to isolate the perceptual units at the entry level of the part-whole hierarchy. Once such entry-level units have been found, they can be subjected to grouping and/or parsing.

Recent results found by Mack, Rock, and their collaborators (Mack, Tang, Tuma, Kahn, & Rock, 1992; Rock, Linnett, Grant, & Mack, 1992) on the role of attention in perceptual organization are broadly consistent with the

present theory. They have found that some process of element individuation—that is, designating figures against ground—appears to occur without voluntary attention, but that classical grouping does not. In their inattention paradigm, subjects are given an explicit task of deciding whether the horizontal or vertical line of a large, centrally presented cross is longer. On the first several trials, only the cross is presented. Then, on a single inattention trial, subjects are presented with an additional element (or elements) of which they have neither prior knowledge nor expectation. Under these conditions, subjects can report with above-chance accuracy the color, location, and approximate number of unattended objects, but not their shape (Rock, Linnett, Grant, & Mack, 1992). The authors interpret this finding as consistent with the view that the visual system registers the presence of “shapeless” individuated UC regions which they call “blobs”⁷ without voluntary attention. Similar experiments on grouping were conducted by Mack et al. (1992). In the first few trials, the cross was surrounded by ungrouped small elements, followed by an inattention trial in which the elements were configured into grouping patterns. Subjects were unable to report the pattern of grouping according to proximity, lightness similarity, or orientation similarity. This result supports the contention that grouping processes occur only after UC elements have been detected and after attention has been deployed to a set of such previously defined elements.

Grouping Processes

Once entry-level units are established, they can be amalgamated into larger sets by the action of at least some of the classical grouping processes Wertheimer proposed: proximity, similarity, common fate, and so forth. Within our theoretical framework, however, they are applicable only *after* perceptual elements exist. We will now consider some examples that illuminate the interplay between entry-level unit formation (i.e., the application of UC) and classical grouping processes.

In Figure 10A, most people perceive a rectangle consisting of two connected squares of different lightnesses. The present account of this outcome requires that the two squares first be designated as separate, entry-level units against an intermediate gray background by the principle of UC and the operation of figure-ground processing. These two units can then be aggregated into a single superordinate unit on the basis of grouping principles. Good continuation of the contours is clearly an important factor, because eliminating it greatly weakens the superordinate unit, as is illustrated in Figure 10B. This example makes the important point that entry-level units are merely the *entry point* into a part-whole hierarchy and do not necessarily determine which level corresponds to the perceived *objects* in the scene. Indeed, there is often no single level that can be identified as “the objects” in a scene, since what is perceived as “the objects” can change with attention and volition. In the present case,

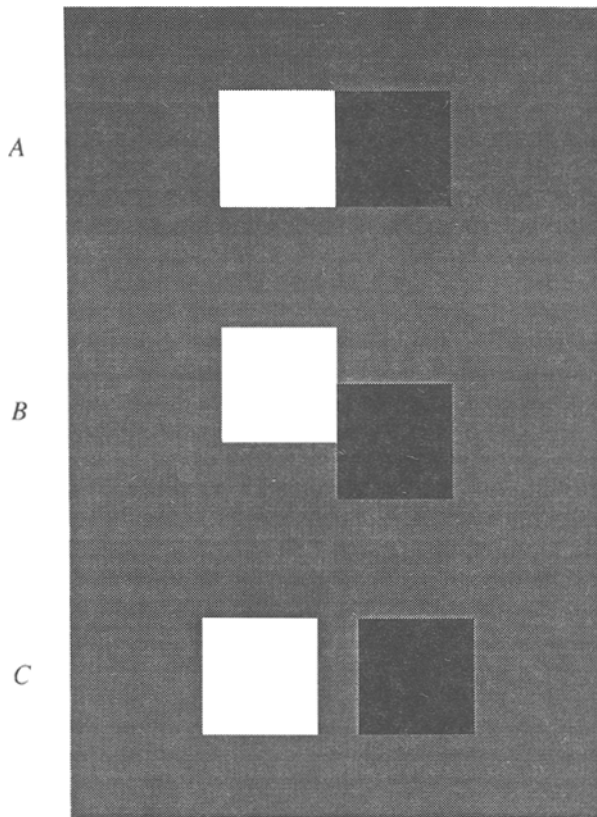


Figure 10. Examples of grouping two entry-level units based on good continuation and element connectedness. Two squares constitute different entry-level units owing to different lightnesses, but are unified into a single rectangle by both good continuation and element connectedness (A). Element connectedness without good continuation (B) weakens the unity of the composite, and good continuation without element connectedness (C) virtually eliminates it.

for example, either the squares or the rectangle can be seen as objects, depending on the observer's set and context.

Although grouping based on common fate is both powerful and ubiquitous, it should also be understood as the result of unifying separate entry-level units. Motion-based grouping is perhaps most obvious when stationary arrays do not provide the necessary information via connected uniformity, as when cloud formations or camouflaged objects change positions over time. Indeed, infants seem to perceive the grouping of object parts on the basis of common fate before they acquire any other grouping factor (Kellman & Spelke, 1983). It is instructive to realize, however, that common fate often arises primarily because of the connectedness of objects in the physical world: retinally disconnected regions arising from the same connected object move in the same way precisely because they are physically connected. Cases in which common fate works for disconnected objects are few and far between: airplanes or birds flying in group formations, fish swimming in schools, or dancers executing a choreographer's instructions.

The end result of grouping several perceptual elements together is a set of superordinate units (*aggregations* or *groups*), which can then be grouped further through successive applications of grouping principles. This allows the formation of a nested hierarchy of perceptual units, with entry-level units usually occupying a lower level than aggregate units. These grouping processes are exemplified in Figure 9 by the superordinate units representing pairs of lobed figures resulting from proximity grouping.

Element Connectedness

We have discussed UC as an organizational factor, conceptually distinct from the classical principles of grouping, that operates before they do. To reiterate, the central distinction is that UC requires no prior units upon which to operate, whereas the classical grouping principles do require such prior units. We now argue that there is another principle of connectedness at work in perceptual organization, one that can be construed as a grouping factor in the classical sense. To distinguish it from UC, we call it *element connectedness*, because it presupposes the existence of two or more prior elements on which to operate.

Element connectedness is the tendency of two or more previously distinguished elements to be grouped together into a single superordinate unit when they constitute two (or more) connected elements against a background. Unlike UC, element connectedness *can* reasonably be characterized as a limiting case of classical proximity grouping, because the only distinction is that between zero and nonzero distance, separating two distinct elements. It could therefore be argued that element connectedness is not a qualitatively new principle of grouping, but a special case of proximity.

We find it more useful to discuss this type of grouping in terms of element connectedness than proximity because of its relation to connectedness in environmental objects. Distal objects often consist of several connected parts that have different surface characteristics. A typical pencil, for example, has a yellow body, silver erasure holder, red eraser, tan point, and black tip. Such multipart objects will obviously not project to a single UC region on the retina, but to several such regions. The fact that these parts are (usually) connected in the environmental object is reflected in the image by the fact that these distinct regions are connected to one another in the image. Again,

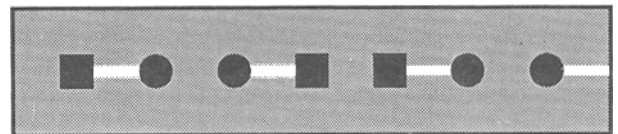


Figure 11. Further examples of element connectedness as a grouping factor. Equally spaced circles and squares are perceived as grouped into pairs when they are connected by distinct entry-level units. This occurs despite the fact that without the white connecting regions, the opposite grouping would be perceived because of shape similarity.

it is connectedness that is crucial, rather than mere proximity, for the same parts would not be perceived to constitute a single object were they disconnected. If the eraser is separated from the pencil, for instance, it is seen as a separate entity, even if it is located right next to the rest of the pencil. We choose to distinguish between element connectedness and standard proximity grouping because of their different implications for objecthood. Indeed, it seems more appropriate to consider proximity as an approximation to element connectedness than the other way around.

The importance of element connectedness within our theoretical framework can be appreciated by noticing that good continuation is not the only factor at work in Figure 10A. The other is that the two figural regions are connected along a common border. Its importance in this example can be discerned by comparing Figure 10A with Figure 10C, in which the two squares have been separated. Even though their borders are still collinear, as in Figure 10A, and even though the two regions are still in close proximity, they are clearly and unambiguously perceived as two distinct objects because they are not connected.

Another example requiring an analysis in terms of element connectedness is shown in Figure 11. It demonstrates element connectedness overcoming grouping by shape similarity. UC cannot directly account for the perceived organization of this example, because the dots and connecting bars do not have the same lightness and therefore will result in separate entry-level units. How can their unitary appearance then be explained? As with the other examples of classical grouping principles, we suggest that at least three distinct stages of organization must be involved. First, the differently colored circles, squares, bars, and background are defined as separate regions of the image. Next, figure-ground processes must determine that the black and white regions are figures against the gray background, thereby defining them as distinct entry-level elements. Finally, these units can be grouped into superordinate units according to the principle of element connectedness as just described. Thus, the strong perception of several multipart objects in this case depends on element connectedness unifying multiple entry-level units. The resulting superordinate objects are strong enough to determine the level of the hierarchy corresponding to the phenomenally perceived object.

The effectiveness of element connectedness in forming perceptual objects out of differently colored regions has recently been demonstrated in patients with a clinical condition known as Ballint's syndrome (Humphreys & Riddoch, 1993). These patients have bilateral damage to the parietal cortex, resulting in a serious perceptual and/or attentional deficit in which they remain fixated on a single perceptual object in their visual field. Their inability to shift attention from one object to another is so severe that they are unable to discriminate systematically between arrays containing many circles of just one color (either all red or all green) and arrays in which half of the circles are red and the other half green. This perceptual dis-

ability can be explained by the hypothesis that such patients are only able to perceive one object (a single circle) in the visual field. However, if pairs consisting of one red circle and one green circle are connected by lines, the same patients are suddenly able to make the discrimination between one-color and two-color arrays (Humphreys & Riddoch, 1993). Unifying a pair of circles through element connectedness thus appears to cause these patients to perceive them as a unitary perceptual object, enabling them to perceive two circles at once, an impossible feat for them in the absence of the connecting line.

Figure-Ground Revisited

Figure-ground organization is an important process involved in all the examples of element connectedness just described and in other examples of classical grouping as well. It is required within the present theoretical framework because once UC regions have been defined in the two-dimensional image, *all* such regions are necessarily interconnected, like the pieces in a stained glass window.

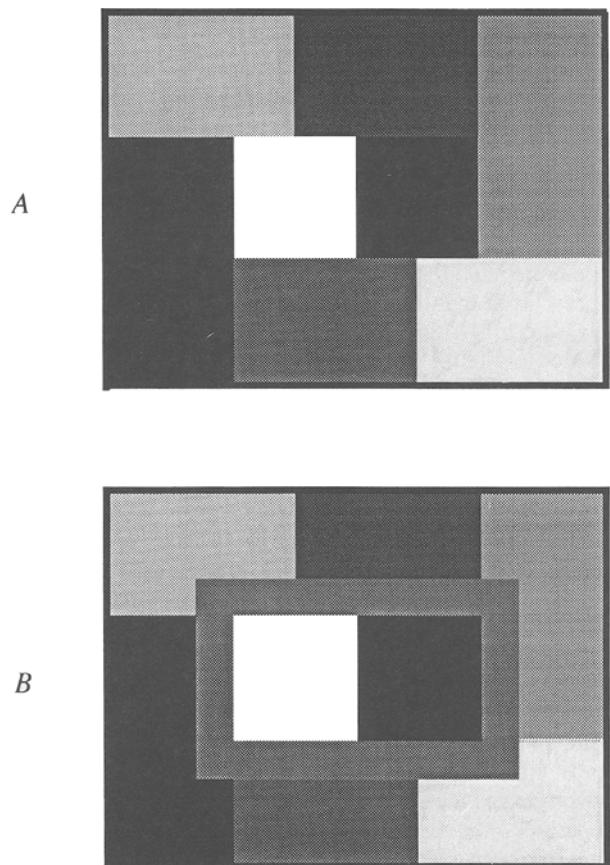


Figure 12. A demonstration of the importance of figure-ground relations in grouping operations. The dark and light squares are not easily perceived as constituting a unified rectangle in part A because of the lack of a single region that constitutes their ground. In part B, the rectangle is perceived more easily, because of the addition of a unified ground.

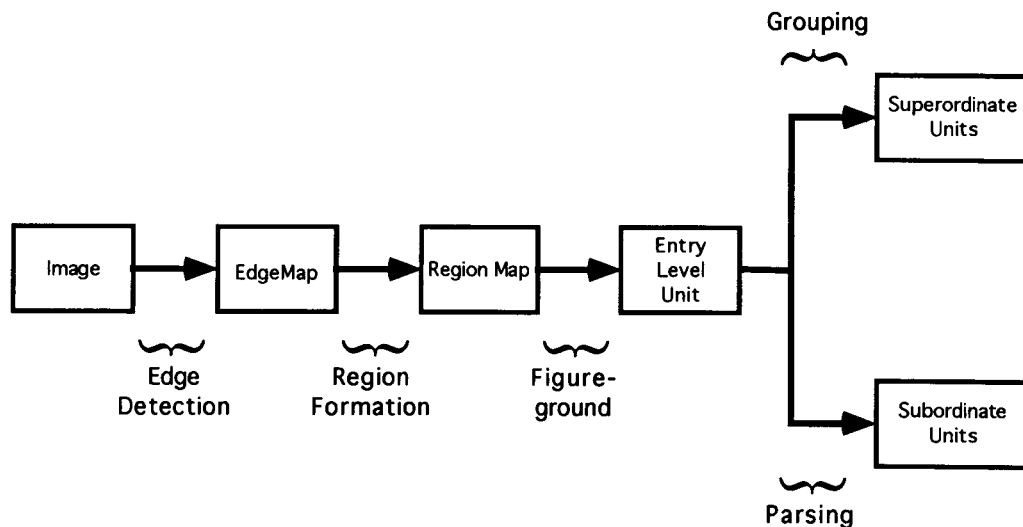


Figure 13. A flowchart representation of the relations among processes proposed to be involved in perceptual organization.

This provides no differential basis on which element connectedness can work to define connected groups. However, if figures are separated from their ground after being organized into UC regions, the entry-level elements can indeed be grouped into superordinate units via element connectedness.

In support of this analysis, consider the figure-ground sensitivity of perceiving two connected light and dark squares as being parts of a larger rectangle as initially demonstrated in Figure 10A. If the same squares are not presented against a uniform surrounding background region, but within a mosaic of other regions as is illustrated in Figure 12A, their perceptual unity is considerably weakened, if not lost altogether. In this case, the interconnectedness of all regions interferes with their perception as a unified entity different from the rest of the mosaic. The whole mosaic, however, is perceived as a single interconnected figural unit against a homogeneous background. If we now change the mosaic so that the two squares in question lie within a single larger region, as is illustrated in Figure 12B, they are again strongly perceived as figures against a uniform surrounding ground.

Note that within the present theoretical framework, figure-ground organization must logically play the same role in demonstrations of all classical grouping factors. Otherwise, element connectedness among all the initially defined UC regions would cause them to be grouped as a single unit. However, if figure-ground organization operates *after* the formation of UC regions but *before* element connectedness, grouping principles will work as expected because only grouping among figural regions (elements) can occur. From this, we infer that figure-ground organization must occur after organization into UC regions, but before grouping.

The logical and empirical constraints discussed thus far suggest the rough flowchart of processes in perceptual organization presented in Figure 13. First, edge-detection

processes operate on the sensory image to define an *edge map*. This edge map is then subjected to region-formation processes, which divide the image into nonoverlapping areas in the *region map*. These regions are then interpreted as figures versus grounds by the action of laws of figure-ground organization, resulting in *entry-level units*. Once formed, these units can be either grouped into superordinate units, as described above, or parsed into subordinate units, as we are about to discuss.

Parsing Processes

Parsing processes subdivide entry-level units into subordinate units or "parts." Hoffman and Richards (1984) have proposed, for example, that objects are divided into parts at concave discontinuities and deep concavities (see also Biederman, 1985). In the example shown in Figure 9, each entry-level unit is divided into two essentially circular underlying parts by parsing them at the paired cusp points. Hoffman and Richards' analysis also applies to dividing complex, connected line configurations at intersections

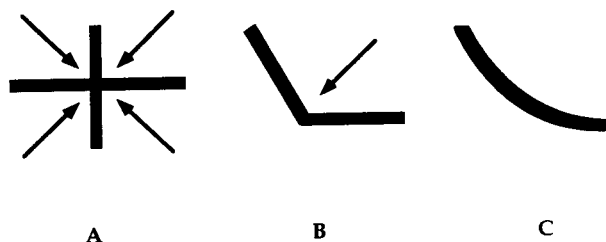


Figure 14. Examples of parsing a single entry-level unit into parts, using Hoffman and Richards (1984) heuristic of dividing contours at concavities. Part A shows an intersection at which multiple paired concave discontinuities strongly indicate division into parts. Part B shows that a single concave discontinuity indicates division less strongly, and part C that a single broad concavity does so very weakly, if at all.

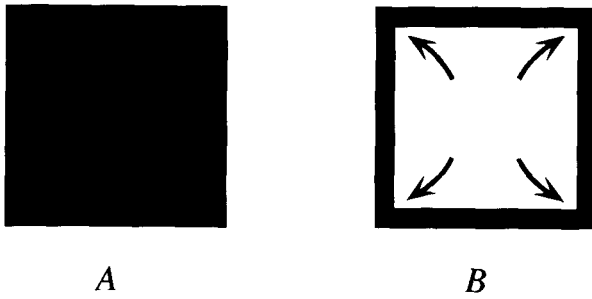


Figure 15. Differences in the perception of solid versus outline squares. The solid square (A) resists parsing into components because of the lack of concavities, whereas the outline square (B) can be readily divided into four component line segments at the interior concave discontinuities.

and other concave discontinuities, as illustrated in Figure 14A. In Figure 14B there is only a single concave discontinuity rather than paired ones, so the division into parts is somewhat weaker than in part A. In Figure 14C, however, the curve has *no* concave discontinuities—although there is a broad concavity on one side—so that parsing into subordinate parts is less likely to occur than in either A or B.

Note that subordinate parts of entry-level units often take on a persuasively independent identity as perceptual

units in their own right. Indeed, it was the powerful independence of parts that enabled us to introduce UC in the manner of a standard principle of grouping in Figure 2, for the parts of the connected figures (consisting of two dots and the line connecting them) are perceived as relatively autonomous units in their own right. It was this same independence that enabled Wertheimer to appeal to line segments within connected figures as perceptual elements (e.g., in Figure 1F). Nevertheless, the present analysis implies that these figures are initially organized as a single entry-level unit by virtue of UC. They must therefore be subsequently parsed into multiple subordinate parts by dividing them at concave discontinuities or deep concavities. We will consider some evidence supporting this proposal shortly.

One interesting example of the effects of parsing at concavities concerns the organizational differences between solid and outline polygons, such as the squares shown in Figure 15. Each figure forms a UC region, and so each constitutes a single entry-level unit as we have defined it. The solid square contains no deep concavities, and it therefore resists being divided into subordinate components: it is perceived as a simple, unitary figure. The outline square, however, contains sharp concavities at each interior corner which support dividing it into four parts: namely, the line segments that constitute its four “sides.” Analyzing the outline square into subordinate parts thus

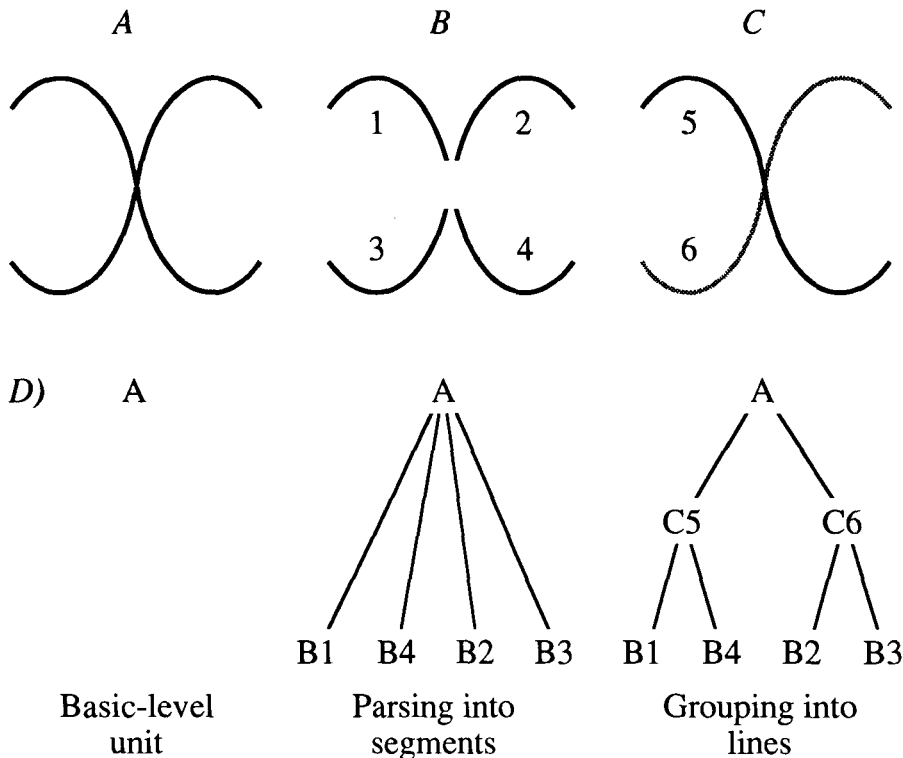


Figure 16. An illustration of the sequence of processes hypothesized to occur in perceiving the X-shaped figure as consisting of two intersecting lines. The initial organization (A) is a single, connected, entry-level unit, which is then divided into four segments (B) via parsing operations. Two pairs of these segments are then unified by virtue of good continuation (C).

has a phenomenal reality that is absent in the closely related case of the solid square.⁸

Parsing processes are required if one is to account for a problem that arises in the present theory. In many cases, the principle of UC predicts the initial perception of a *single* unit where *multiple* units are typically perceived. In fact, we have already presented one good example of this phenomenon, but did not comment specifically on this aspect of it. In the illustration of good continuation in Figure 1F, the principle of UC predicts that an observer will initially perceive a single unit—namely, the entire curvilinear X. Although this organization does appear to be experienced as a cohesive unit, it also seems clear that we perceive two intersecting lines as distinct subparts of the X, as Wertheimer's analysis implies. This level of perceptual unit is only possible in the present theory if entry-level units can be divided into subordinate parts.

A full account of this particular example within the present theoretical framework actually requires both parsing and grouping processes to operate, as is depicted in Figure 16. First, the entire X-shaped figure is constructed as the entry-level unit (see Figure 16A). Then it is parsed into four component line segments by dividing it at deep concavities (see Figure 16B). Once these four distinct subordinate units exist, they become candidates for unification by the action of grouping principles. In this case, good continuation is obviously the factor that favors unifying opposite segments into the two longer lines (see Figure 16C). Thus, the full account of how this simple figure is perceived within the present theoretical framework requires the three hierarchically nested levels of organization depicted in part D of Figure 16. They are constructed in real time as just described: the top level first, the bottom level next, and the middle level last, as is indicated by the sequence in Figure 16. Note that the "standard" account just in terms of good continuation (Wertheimer, 1923) differs from ours in supposing that the lowest level organization, consisting of the four separate segments, occurs first, the middle level next, and the highest level last. Although it is not yet clear exactly how to determine which possibility is true, it is an example of a differential prediction made by the present theory.

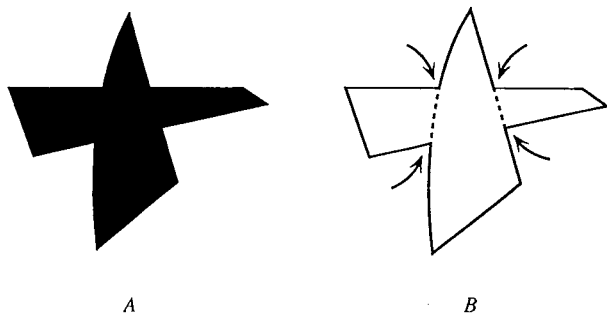


Figure 17. Part A shows an example of a figure that is spontaneously perceived as the occlusion of one figure by another. For this to occur in the present theory, the unitary initial figure must be parsed into appropriate regions, as indicated in Part B.

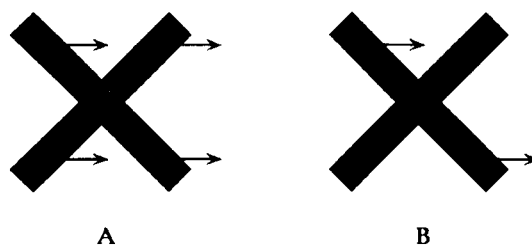


Figure 18. Illustrations of the relation between UC and common fate. When the entire figure moves as a unit (A), common fate is consistent with UC. When part of the figure moves differentially (B), however, the two principles conflict, and common fate dominates the final perception.

Another instructive example is shown in Figure 17. Again, the principle of UC alone does not predict the usually described perceptual outcome—namely, the superposition of two separate figures of the same color, one occluding the other. The problem is that the present theory implies the initial organization to be that of a single entry-level unit comprising the entire UC region. If the entry-level unit is parsed into three parts at the paired concavities indicated in Figure 17B, however, the opposite parts can be grouped into unified figures by virtue of good continuation of their contours and can thereby be perceived as occluded ends of the same connected object (Kellman & Shipley, 1991). Thus the logic of this example is the same as that of the curvilinear X in Figure 16. The theory predicts that the initial perception of such displays consists of a single, undifferentiated figure, which must then be split into the two overlapping components. If this splitting process requires enough time, it should be detectable with suitably sensitive experimental procedures.

Examples can also be constructed in which both common fate and UC are present, yet conflict in their implications for the most appropriate perceptual units. Consider the case of the moving X-shaped object in Figure 18. If it moves as a rigid object, as is indicated in Figure 18A, then its motion changes nothing about its perceptual organization, and we suppose that its *initial* structuring will be as a single unit, which is consistent with its analysis in terms of UC. However, suppose that one of the two intersecting lines were moving harmonically back and forth, as is indicated in Figure 18B, without moving so far that the two lines become disconnected in the retinal image. Then the perception of two separate lines would be strong enough to block the perception of a single unitary figure based on connectedness.

This is an interesting case, because the differential motion of the lines indicates that the retinally connected region *cannot* be connected in the world. Rather, it would typically indicate that the retinal connectedness of the lines is merely coincidental, with one line actually lying behind the other and the perspective accidentally resulting in their retinal connectedness. As several prominent theorists (e.g., Biederman, 1985; Lowe, 1985; Rock, 1983;

Shepard, 1981) have pointed out, the visual system avoids coincidence, favoring nonaccidental explanations of sensory information. Thus, this example is akin to a case in which stereoscopic depth information at a later stage of processing overcomes an initial organization in terms of UC. We will now analyze such cases as we consider the operation of UC based on higher-order properties such as motion, texture, and depth.

UNIFORM CONNECTEDNESS DEFINED BY OTHER STIMULUS PROPERTIES

Thus far, we have discussed the application of the principle of UC almost exclusively within the domain of luminance structure in the image. We suggested at the outset, however, that it also defined regions based on other surface properties, such as chromatic color, texture, and possibly other properties as well. We will now discuss these possibilities and consider the implications of including them within the present framework.

In the present section, we examine the possibility that UC can define entry-level image regions for other properties that the visual system processes in parallel over the entire retinal image. Within the edge-detection paradigm for region segregation, the only requirements are (1) a retinotopic map of cells sensitive to a given image property and (2) *edge detector* cells that are selectively sensitive to rapid spatial changes in that property. The changes detected by such cells would signal the transition from one UC region to another with respect to the given domain. Texture, motion, and binocular disparity are all second-order properties currently believed to be computed in parallel over the entire visual field and on which edge detection appears to take place (Cavanaugh, 1987, 1988). We will now briefly consider how UC might be applied to each of these properties.

Chromatic Surface Color

The generalization from UC regions defined by luminance to those defined by chromatic color is relatively

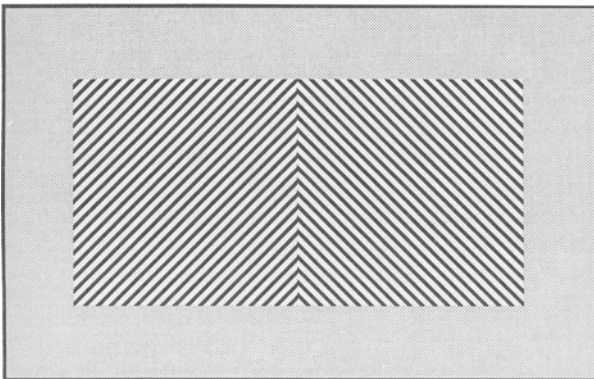


Figure 19. An example of hierarchically nested relations of color-defined regions within luminance-defined regions. The red (right diagonal striping) and green (left diagonal striping) equiluminant regions are perceived as a unified rectangle within the lighter background.

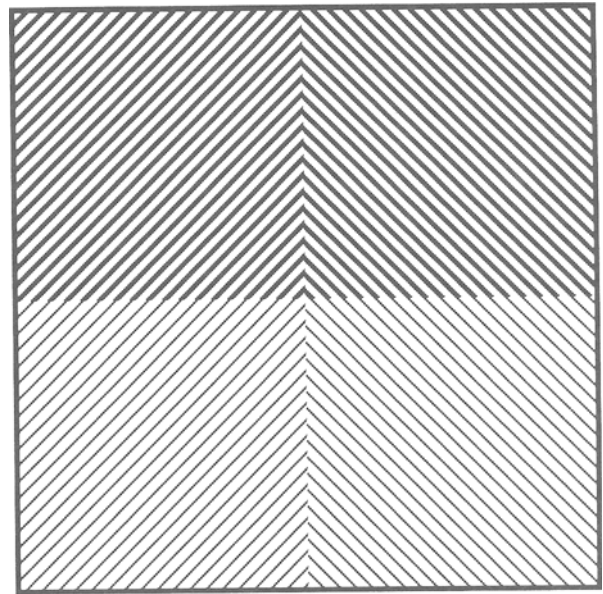


Figure 20. An example of competition between color-defined and luminance-defined organizations. The left diagonal stripes represent red and the right diagonal stripes green, with stripe density representing luminance. This pattern occurs most frequently when a luminance edge (from a shadow or change in surface orientation) intersects a chromatic reflectance edge.

straightforward. The fact that chromatic color alone is sufficient for the perception of regions has been demonstrated by studies of perception at equiluminance (Cavanaugh, 1987), although it is often quite difficult to perceive certain properties of equiluminant regions.⁹

For the present purposes, the first important observation is that if both color and luminance are used to define first-order UC regions, there will be two region maps for the image: one for luminance and another for chromatic color. This immediately raises the question of the relation between the two. In many cases, they will be essentially the same, because differences in color tend to be strongly correlated with differences in luminance for normal images of the natural environment. Logically, however, they are independent. Thus, the regions defined by luminance and color can potentially be consistent with one another, hierarchically nested, or competing.

There is obviously no problem if luminance and color are consistent: they simply converge on the same regions and the same entry level into the hierarchy. If they are hierarchically nested, the regions defined are compatible, although they define entry into the hierarchy at different levels. An example would be a pair of connected, equiluminant, but differently colored squares on a lighter background, as is indicated in Figure 19 through the use of textures rather than colors. The luminance structure of such an image defines just two regions: the lighter background and the darker rectangle. The color structure, however, defines three distinct regions: the two differently colored squares and the lighter background. Thus, two of the three regions defined by color are nested within

one of the two regions defined by luminance, and the remaining unit is the same.¹⁰

The present theory also allows for situations in which color and luminance will lead to conflicting partitions of the image, as is illustrated in Figure 20. This situation seldom occurs for surface lightness and color, except in man-made designs such as painted surfaces and dyed fabrics. Although one seldom finds *surfaces* whose reflectances differ in this way, however, one does frequently find *image* regions with precisely this structure. In particular, the juxtaposition of regions shown in Figure 20 occurs whenever an illumination edge crosses an equiluminant chromatic reflectance edge, as when a shadow falls across portions of a surface containing an equiluminant red/green border. This example again raises the question of levels, for this pattern of regions typically arises at the preconstancy level of image structure rather than the postconstancy level of surface structure. We will consider this issue again later.

This example demonstrates that one frequently *does* find "equichromatic" regions of different luminance juxtaposed, as in the left and right portions of Figure 20. In such cases, entry-level luminance regions will be nested within entry-level chromatic regions, because the hue and saturation of the surface will be approximately constant across the illumination contours. This structure is therefore important information that there is a single chromatically uniform surface present under differential illumination (Gilchrist, 1979; Rubin & Richards, 1984). The regions of chromatic UC will necessarily be larger, corresponding to the visible surfaces of the uniformly colored object.

Although the alternative entry-level organization provided by chromatic information in some sense complicates the theoretical picture—because there is no longer a *single* partition of the image or a *single* entry level into the part-whole hierarchy—it provides an ecologically important level of structure for the visual system to use in defining objects and parts that have uniform surface qualities.

Surface Texture

Texture is a visual property for which an account of region formation appears to be possible in terms of classical grouping of texture elements. That is, one can conceive of arriving at a regional organization in terms of texture by the operation of factors such as element proximity and various forms of element similarity. Figure 21 illustrates an example in which there is strong segregation between regions containing differently oriented line segments, and strong grouping within regions containing like-oriented segments. Here Wertheimer's analysis would seem to apply in force: the texture elements appear to be perceived as discrete entities, yet they are also perceived as grouped together into larger regions of similar elements. And it has frequently been interpreted in just this way: as grouping by similarity of texture elements (Beck, 1966; Rock, 1986). Texture segregation thus appears to be an example for which an analysis in terms of UC sim-

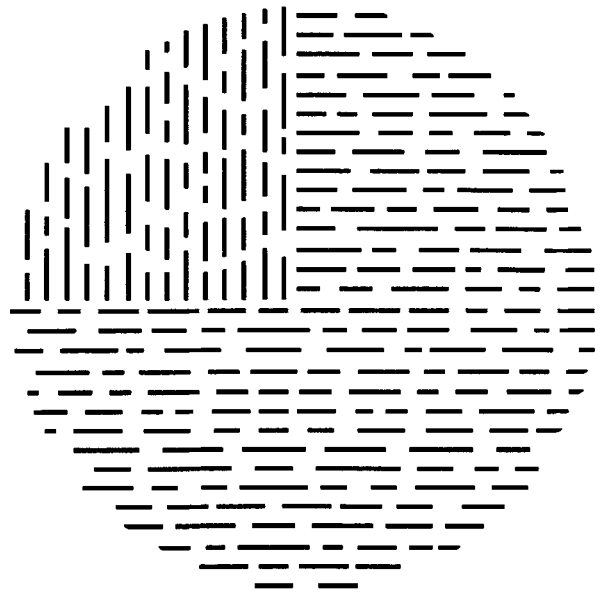


Figure 21. An example of organization from texture. One quadrant of the circle is distinguished from the others by virtue of its segments' being oriented vertically rather than horizontally.

ply does not apply—except, of course, for the perception of the individual texture elements based on luminance and/or color.

But there is another way of viewing texture segregation in which an analysis in terms of UC applies quite directly. As we remarked briefly at the outset, the principle of UC is potentially useful for organizing image regions that correspond to connected objects (or parts) that have uniform surface characteristics, presumably including their texture as well as their lightness and chromatic color. Therefore, one can justifiably claim that texture segregation may occur via visual mechanisms that identify regions of homogeneous texture, in much the same way that such mechanisms find regions of homogeneous lightness or color (Julesz, 1975, 1981).

According to this view of texture segregation, the organizations defined by UC with respect to luminance and texture can differ from each other. The UC regions of luminance (or color) in the image would correspond to individual texture elements as organizational units. The UC regions of texture, however, would be identified with much larger regions. These two levels of organization exist at quite different spatial scales. The individual elements corresponding to UC regions of lightness and/or color exist at a relatively small spatial scale; the areas corresponding to UC regions of texture, at a relatively large spatial scale.

As a result, the entry-level regions of luminance (or color) and texture tend to be compatible with each other, forming a hierarchically nested structure. The resulting hierarchy of small versus large units of organization is an important, if not a defining, characteristic of texture perception. It actually represents a corresponding hier-

archical structure in the environment, since textured surfaces are typically composed of many smaller objects or surfaces: grassy fields of many blades of grass, brick walls of many identical bricks, tiled floors of many small tiles, and so on.

For the reasons just cited, the principle of UC applies to the texture of perceived objects. To compute such regions, the only requirements are (1) that the visual system process image characteristics corresponding to a surface's texture in parallel, and (2) that there be mechanisms to detect boundaries at which these textural properties change. Indeed, the most recent and successful computational theories of texture segregation accomplish this by employing essentially the same paradigm that was developed in computer vision for finding regions based on luminance or color: boundaries are detected between regions by computing differences in the output of what are essentially particular sorts of "texture filters" based on spatial frequency channels (Beck, Sutter, & Ivry, 1987; Bergen & Landy, 1991; Graham, Beck, & Sutter, 1992; Malik & Perona, 1990). Homogeneous connected regions of texture can then be identified with the interiors of closed texture boundaries, thereby incorporating this domain within the general framework of organization by UC.¹¹

As in the case of luminance and color, it is not easy to say a priori whether the perceptual units corresponding to the smaller, luminance-defined elements or the larger, texture-defined regions will be perceived first. It is tempting to suppose that the elements must be perceived first, since, logically, texture would not exist without the elements of which it is composed. But this is not necessarily an accurate statement of the perceptual situation, for the texture elements may not be perceived *as elements* prior to texture analysis. That is, the elements may be present in the stimulus without yet being perceptually differentiated as individual elements. The reason is that both the luminance analysis and the texture analysis of UC can take place directly from the image, and it is possible that the texture analysis might finish first.

Note that we have now given two quite different accounts of texture segregation: one in terms of classical grouping and the other in terms of UC of texture. The present theory allows for both possibilities. It is interesting to conjecture, however, that there might be particular conditions favoring one mechanism over the other. For example, Wertheimer's original demonstrations of grouping might, in principle, be accounted for by finding texture edges, but we suspect that the dominant mechanism will depend on the number and size of the elements. When there are relatively few, relatively large elements, as there often are in grouping demonstrations (e.g., Figure 1), the classical "putting together" explanation may be more appropriate. When there are many relatively small elements, as there usually are in standard texture segregation displays (e.g., Figure 21), the boundary-finding mechanism may be more appropriate. (See Goldmeier, 1972, and Kimchi & Palmer, 1982, 1985, for further discussion of the closely related perceptual distinction between "form" and "material.")

Other Properties

The same type of analysis is possible for other visible properties, such as motion and depth. For motion, the basic assumption is that the visual system attempts to find objects or significant parts in terms of those that undergo uniform motion. At the early, preconstancy, level of image structure, entry-level regions can be identified by computing connected regions of uniform (or gradually changing) image motion via detecting motion edges. That image motion alone is sufficient to define visual elements can be nicely demonstrated with the movement of random dots. If a region of dots moves uniformly against a stationary background of otherwise identical dots, for example, the moving region is effortlessly perceived as a figure with a well-defined shape, distinct from the motionless ground of dots behind it. This fact cannot be explained by information from luminance, color, or texture because the moving region is identical to the background in all these (momentary) characteristics.

It might at first be thought that UC of motion is equivalent to the classical grouping principle of common fate, but this is not the case. The crucial distinction is that common fate does not require the elements in question to form a connected region, whereas UC of motion does. Suppose, for example, that two spatially separated regions of dots are moving in the same direction and at the same speed on a stationary background of motionless dots. This organization of the percept can be explained by the detection of UC with respect to motion, because the two separate regions of moving dots can be identified as two distinct perceptual units. Somewhat surprisingly, common fate cannot explain this perception, because all of the elements in *both* groups are undergoing common motion. Common fate therefore predicts the perception of a single group undergoing uniform motion rather than two distinct regions against an unmoving background. There is no basis within common fate for distinguishing between the two separate groups—unless, of course, one appeals to proximity as well.

The situation in perceiving motion structure among many individual elements is precisely analogous to that of perceiving texture, in the sense that two quite different accounts can be given of such perceptual organization into regions. One is the "classical" account, in terms of grouping by common fate: the velocity of each individual element can be registered as one of its properties. The motion of all such elements can then be used to construct higher order perceptual units (groups) consisting of those elements that have the same (or similar) motion. As in the case of grouping-based accounts of textural organization, the individual elements must be designated as elements before they can be grouped into superordinate units of uniform motion.

The alternative account is based on organization arising from the *direct* computation of UC regions as defined by motion. For this to occur, the visual system must be able to compute image motion in parallel over the entire visual field and to detect spatial gradients in that field of motion at both large and small scales of resolution. The

closed contours of such large-scale regions can then be determined from the output of "motion-edge detectors" in just the same way as for uniform luminance, color, and texture. In the case of random dot motion, there will be regions of uniform motion at both small and large scales, corresponding to the small regions of uniform motion defined by the individual dots (which will coincide with the small regions of uniform luminance defined by the dots) and the large regions of uniform motion defined by textures of dots.

Again, there is no good evidence we know of for deciding which account of organization by motion is correct. As in the case of texture, it may be that both mechanisms are used under somewhat different circumstances. A few relatively large elements moving with the same velocity may be first perceived as individual elements and then grouped by common fate, whereas numerous small elements may be organized into regions of uniform motion by direct detection of spatial gradients of motion.

The final property that we will consider on which the principle of UC might operate is depth. The basic claim is simply that connected regions at a uniform (or continuously varying) distance from the viewer will tend to be organized as a single element.

A basic proximal stimulus property that carries depth information, at least for nearby objects, is binocular disparity. Disparity, of course, is a property not of an environmental surface, but of displaced pairs of image points within the same scene. We have already mentioned an example that demonstrates the effect of binocular disparity in the principle of UC. Figure 8 showed that retinally connected regions of uniform luminance are organized according to depth when viewed binocularly. The action of UC in the domain of binocular disparity is sufficient to account for this phenomenon. Without it, there is no reason to expect that there will be any differential organization of the dots.

A property of environmental surfaces that is closely related to disparity and depth information is their orientation in 3-D space. A slanted plane, for instance, has uniform (constant) orientation, even though its depth varies continuously across space. Any curved surface has nonuniform slant, of course, but if its curvature is continuous, surface orientation will vary continuously. Thus, surface orientation is potentially another property that might prove useful if analyzed in terms of UC. There may well be others we have not yet considered.

THE QUESTION OF LEVELS REVISITED

Postconstancy Representation

To accommodate the evidence that grouping is influenced by relatively late perceptual processes, we suggest that some phase of perceptual organization—possibly even all of it—takes place within a representation that incorporates depth and perceptual constancy. This representation reflects the perception of environmental properties of surfaces themselves, rather than the retinal regions to which they project. To conform to the present theoretic-

cal framework, we suggest that this representation would have to include phenomenal surface properties such as their lightness, chromatic color, and texture as well as observer-relative properties such as their motion, distance, and orientation. That is, this representation would have to incorporate the results of texture analysis as well as constancies of lightness, size, orientation, shape, and motion.

Let us call this the *postconstancy representation*. We conceive of it as being somewhat like Marr and Nishihara's (1978) construct of the 2½-D sketch, in that it should contain information about the distance to a local patch of surface and its slant with respect to the viewer. It differs in that it should also include other properties of the surface, such as its lightness, color, texture, and motion. It also bears a close resemblance to some of Barrow and Tennenbaum's (1978) proposals about "intrinsic images." We intend it to be theoretically more neutral, however, in that we do not propose a specific representation. We merely use this term to refer to whatever internal structures support conscious perception of the properties of object surfaces in 3-D space.

The advantages of defining perceptual units within such a postconstancy representation are considerable. Many thorny problems can be avoided if organization occurs at this level instead of or in addition to an earlier, image-based representation. In the environment, for example, connected homogeneous objects have component surfaces of the same reflectance, but they often project to retinal regions of different luminances. This complication results from the fact that the object's surfaces are frequently illuminated differently, whether because of their varying orientation relative to the light source (or sources), because of the presence of shadows cast by other surfaces from the same or other objects, or both. The variations in luminance due to shadows and shading can be surprisingly wide, often much greater than the luminance differences between white and black surfaces under equal illumination.

If organization according to uniformity of *retinal* luminance were the whole story, the visual system would face a daunting task in perceiving any complex 3-D object as a single unit, because luminance differences would necessarily cause each entry-level region to be defined as a separate unit within the early, preconstancy representation. The numerous entry-level units constituting the image of an object of uniform reflectance could conceivably then be unified by the action of grouping principles such as element connectedness, chromatic similarity, good continuation, and so forth. But there may be an easier way.

Consider how much simpler the problem would be if a later organizational process operated on the postconstancy representation as we have suggested. Because the entire object would be represented as having uniform lightness at this level, it could easily be organized into a single connected unit or object by applying the principle of UC to its postconstancy lightness representation. That is, the surfaces of an object with uniform reflectance

would be represented as having uniform *perceived* lightness in this postconstancy representation, so that finding the connected region of uniform perceived lightness would correspond veridically to the homogeneous object (or part) in the environment, despite many luminance variations due to differences in lighting and/or surface orientation. Note that this possibility is entirely consistent with the evidence reviewed earlier. For example, grouping by achromatic similarity in color is based on perceived lightness rather than retinal luminance when cast shadows and partly transparent surfaces are present (Rock et al., 1992). Likewise, grouping by proximity depends on perceived 3-D distance in the environment rather than registered 2-D distance on the retina (Rock & Brosgole, 1964).

Once such a postconstancy representation has been constructed, all that is needed is an algorithm that extracts connected regions of constant perceived lightness (or chromatic color, texture, etc.) from this representation and defines them as perceptual units. Note that the connectedness alluded to in this representation is more than 2-D, because it occurs after at least some depth information has been taken into account. It is quite possible, therefore, that two regions could be connected and uniform in both 2-D images, yet not perceived as continuous in 3-D if binocular disparity indicated separation in depth. An example is given in Figure 22 as a stereoscopic pair.

Thus, UC may well operate on both the early, preconstancy representation and the late, postconstancy representation. This raises the question of the relationship between the two analyses. Is the initial organization of the 2-D representation carried over into the later 3-D representation in some way as the "default" organization, or is it merely used as the basis of the processes that extract depth information and result in perceptual constancy? In other words, is organizational processing simply started afresh in the later postconstancy representation? Unfortunately, we do not have a definitive answer to this difficult question. Our belief, based on examples such as Figure 22, is that it may indeed be started afresh, since the unity of the X figure vanishes once binocular fusion is achieved.

Preconstancy Representation

The proposal that consciously perceived organization is governed by processes that occur in the postconstancy

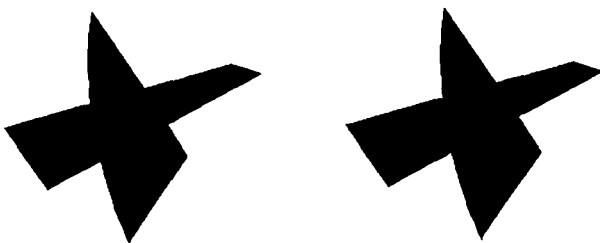


Figure 22. An example of dividing a single entry-level unit on the basis of binocular disparity. These two figures constitute a stereoscopic pair which, when fused, necessitate dividing a homogeneous connected region into two separate surfaces, one of which appears closer.

representation opens the question of what organization is present prior to that level in what we will call the "preconstancy representation." Before raising the issue of the postconstancy representation, we talked as though the 2-D image representation undergoes a full-blown organizational analysis, including the formation of entry-level units, their grouping into superordinate units, and their parsing into subordinate units. But introducing the idea that organizational processes also operate on the postconstancy representation requires that we reconsider this hypothesis. Does *all* organization occur at the postconstancy level, or must some (or all) also be initiated at the early, preconstancy level? And if some organization is required earlier, which kinds specifically?

It seems almost certain that there must be *some* organization at an earlier stage, for depth and constancy processing are often thought to require organized perceptual units on which to work. Stereoscopic depth, for example, is generally thought to operate by finding matches between discrete image elements, such as edges and blobs, in the two images (see, e.g., Marr & Poggio, 1979). Perceiving depth from pictorial sources of information, such as occlusion, perspective, relative size, and shading also seems to require prior organization into relevant perceptual units. Although these arguments are not logically compelling—for the simple reason that even strongly held current beliefs may ultimately turn out to be in error¹²—they are relatively plausible, so we tentatively conclude that there is organization prior to the postconstancy representation. But now let us consider what sort of organization this might be.

Although many proposals are possible concerning the organization of the preconstancy representation, it is useful to view them as lying along a continuum from maximal to minimal 2-D organization. At the maximal end is the hypothesis that all the classical principles of perceptual organization—to which we can now add UC and element connectedness—operate on the early, preconstancy representations of images. This possibility implies that the visual system arrives at a full-blown, yet preliminary, analysis of image-based organization, which can then be modified as necessary once depth and constancy information become available. In general, organized units in the later postconstancy representation will tend to be unifications of units in the preconstancy representation, when factors like illumination conditions (shadows and highlights) and depth are incorporated. There are, to be sure, some cases in which a 2-D connected region must be divided when illumination and depth information come into play, as in the perception of two distinct objects in Figure 22. But such cases are relatively rare under normal circumstances, because they result from accidental retinal alignment of distinct surfaces that happen to be reflecting the same amount of light to the eye.

At the minimal end of the continuum is the more radical possibility that preconstancy organization is far less organized than has generally been supposed. Perhaps only rudimentary organizational units—for example, lines and edges (Hubel & Wiesel, 1968) augmented by blobs and

terminators (Marr, 1982)—are present in the preconstancy 2-D image representation. For this to be a viable hypothesis, however, the rudimentary organizational units would have to be sufficient to support the achievement of perceptual constancy. This particular hypothesis can be examined in light of current computational theories to determine whether lines and edges might be necessary and sufficient for achieving the postconstancy representation.

Lightness constancy, for example, is currently thought to be an essentially edge-based process in which luminance edges are classified into reflectance versus illumination edges (e.g., Gilchrist, 1988). To do this for an entire scene, however, requires that local edge information be integrated over extended areas of the retinal image (Land, 1977). Given the central role played by edges in such computations, they appear to be a necessary organizational entity in achieving lightness constancy. Whether they are sufficient or not is a more difficult question. To our knowledge, there are presently no viable theories of lightness constancy that propose processes working directly on an unorganized image. Indeed, given that lightness must be assigned to *some* differentiated perceptual entity, it is unclear how this could happen unless it were on the basis of individual pixels.

The situation is more complex for other aspects of constancy. Stereoscopic depth algorithms have been devised that work on the basis of the location and orientation of local edges (e.g., Marr & Poggio, 1979), but others do not require the presence of such edges as elements (e.g., Marr & Poggio, 1977). Additional edge-based computations have been proposed as necessary for the completion (or interpolation) of partly occluded surfaces, based on the “reliability” of edge discontinuities (Kellman & Shipley, 1991). Thus, current computational theories suggest, albeit inconclusively, that organization into luminance edges is required for arrival at the postconstancy representation.

The Importance of Regions in the Preconstancy Representation

Even if edge-based computations turn out to be *necessary* for achieving a postconstancy representation, it is unlikely that they are *sufficient* for this purpose. Certainly *local* edge detection of the sort many believe to be performed by simple cells in cortical area V1 (e.g., Hubel & Wiesel, 1968) are inadequate to the task. As mentioned above, for example, lightness constancy requires that local luminance edges be interpreted as either illumination or reflectance edges that must be integrated globally. Similarly, depth information from factors such as linear perspective, relative size, and occlusion almost certainly require more global structuring of edge information into candidate objects in order for reasonable performance to be achieved. We therefore reject the idea that local edge information is, in and of itself, sufficient for constancy to be achieved.

The question we now want to address is whether organization into *regions* is also required in the preconstancy representation, thus implicating the principle of

UC in constructing this early representation. Marr (1982, pp. 270–272) apparently believed that it was not. We believe that it is. Indeed, we suggested at the outset that edge information may be extracted in part to identify UC regions in the image.

Kellman and Shipley’s (1991) theoretical framework for understanding the perception of partly occluded objects provides an instructive example. They propose computing the “reliability” of edge discontinuities to determine which objects are occluded and which occluding. This process requires determining potentially global relations among extended edges, such as deciding for a given occluded edge which other edge in an entire scene, if any, is to be taken as its continuation. For example, in Figure 23A the occluded edges on the right are good continuations of the occluded edges on the left, and thus are highly reliable by Kellman and Shipley’s analysis.

We claim that for the edges to be properly matched, however, it is essential for the system to know which edges constitute boundaries of which image regions. After all, it is actually the alignment of environmental surfaces (which correspond to regions) that is crucial. Edges are a means to this end, however, because if regions are indeed uniform, edges provide the only available information about surface alignment. If the edges align but the surfaces (regions) do not, perception of occlusion should not occur. In Figure 23B, for example, the edges of the right-hand occluded region match up precisely with two edges of the left-hand ones, yet people typically do not spontaneously perceive them to be connected behind the occluder as they do in Figure 23A. The reason is presumably that the otherwise highly reliable edges on the left are boundaries of regions that do not relate well to those on the right. Thus regional information seems crucial in determining the perception of completion under partial occlusion.

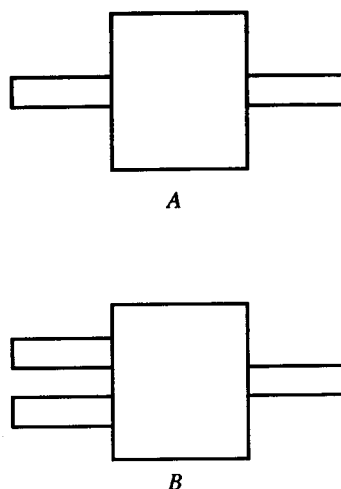


Figure 23. An example demonstrating that regions and figure-ground information must be used in unifying regions according to good continuation (or “reliability”). In part A, the aligned edges are perceived as corresponding, whereas in part B they are not.

Figure-ground organization is another important aspect of perceptual representation that logically appears to require regional information rather than just edge information. Figure-ground relations are defined primarily by three phenomenal outcomes: (1) the figure is perceived to be closer than the ground, (2) the ground is perceived to extend behind the figure, and (3) the shared contour between figure and ground belongs to the figure, thus giving it shape. For example, if the lighter left side of Figure 24 is perceived as figure, it is perceived as having a convex shape and as being closer than the uniform black surface that extends behind it. If the black region is perceived as figure, it is perceived as having a concave shape and as being closer than the uniform white surface that extends behind it. These outcomes are an important component in assigning depth and occlusion relations to a scene, and thus are required for the postconstancy representation in depth to be achieved. Indeed, figure-ground processes are perhaps more properly considered an aspect of depth processing than of perceptual organization, although the two are closely intertwined, as we have seen already.

The crucial fact for the present discussion is that many of the factors that govern figure-ground organization refer directly to the properties of 2-D *regions*—for example, convexity, relative size, and surroundedness—rather than to the properties of edges or boundaries. These properties are simply undefined for an edge or a contour. Since figure-ground relations are governed by such region-based properties, it is necessary to postulate the existence of regions in the preconstancy representation. This is consistent with our proposal that the principle of UC must be applied to the uninterpreted image before figure-ground organization can emerge.

Parsing figural regions into subordinate parts is another process that is thought to require information about concavity/convexity (Hoffman & Richards, 1984). If parsing is accomplished by dividing at deep concavities, pars-

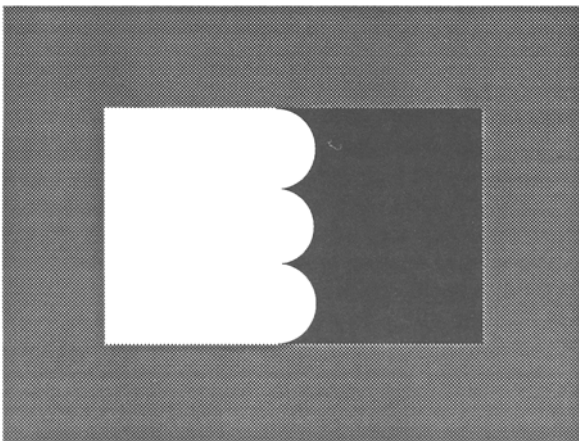


Figure 24. An example of figure-ground organization by convexity, which illustrates that regional information must be utilized. (See text for details.)



Figure 25. Rubin's (1921) vase/faces figure illustrates that parsing into parts occurs only when the region in question is figural. When the white vase is perceived as figure, this region is divided into parts at its paired concavities, but when the same region is perceived as ground, no such division occurs.

ing must also be based on a representation containing information about regions. Furthermore, we can infer that figure-ground organization must already have occurred before parsing begins. The rationale for this claim is that parsing at a "deep concavity" occurs only for figures, since the concavities of background regions do not "belong" to the background region and therefore cannot result in division. In the classical vase/faces ambiguous figure illustrated in Figure 25, for example, the vase can be divided into parts at its concavities, as Hoffman and Richards (1984) have argued. But when the same central region is perceived as ground in the faces interpretation, no such division into parts occurs, even though the same image-based concavities are present. It is perceived as a homogeneous surface behind the faces with no subordinate division into parts at all.

Thus regions appear to be an important aspect of perceptual representation and are almost certainly present at a relatively early stage in vision, prior to the achievement of constancy. Although the matter is by no means settled, we tentatively conclude that the early, preconstancy representation must include (at least) local edges, globally integrated edges, and the connected regions or areas that these edges bound, organized as figures on grounds. Without all three types of information, it seems unlikely that the postconstancy representation could be constructed, at least given our present understanding of how depth and constancy processing occur. Whether further information is required is obviously an interesting and open question.

Repairing Errors Due to UC

We have argued that UC is a useful heuristic for organizing perception in terms of likely environmental objects and have discussed several implications of this view for a theory of perceptual organization. If UC were always sufficient, of course, uniform objects (or uniform

parts of objects) could be identified within the perceptual system simply by applying the principle of UC at the level of the retinal image. As we noted earlier, however, there are numerous cases in which it fails to produce the correct mapping from connected external objects with uniform surfaces to UC regions. We now consider how such errors might be repaired by other perceptual processes within the current theoretical framework.

Many-to-one mappings occur when several objects in the environment accidentally map to a single connected region in the image. This happens only under rather unusual circumstances, such as cases of camouflage or infrequent juxtapositions of surfaces that project the same luminance to the retina. If indeed edge detection and region formation by UC of luminance and/or color were the only organizational processes at work in the early, preconstancy stage of processing, there would be no possibility of repairing such errors at an early level. Analysis of UC in motion and depth, however, can clearly overcome erroneous region formation. If binocular disparity between the two objects is sufficient to be detected, it can be used to correct an initial error, as is illustrated in Figure 22. Differences in motion between the two objects can also be used to separate them in depth, as is illustrated in Figure 18. They result in a postconstancy representation in which the principle of UC can be applied to recover the veridical organization.

The other type of error due to the heuristic nature of UC is far more common: one-to-many mappings. They result from a UC object in the environment projecting to two or more different image regions as defined by UC. As mentioned earlier, there are three different ways in which this can occur, depending on whether it is the uniformity of the corresponding regions that fails to be present, or their connectedness, or both.

The problem of uniform objects projecting to non-uniform but connected image regions arises mainly when surfaces of the same object are differently oriented or are shaded by other objects. Both circumstances are illumination conditions that cause different luminances to be projected onto the retina despite uniform reflectance. These are precisely the cases in which finding UC regions within the postconstancy representation provides a simple way of unifying nonuniform regions in the preconstancy representation. Constancy processing is hypothesized to recover the uniform qualities of the surface such that the nonuniformities are absent at the postconstancy level. The result is a strict hierarchy in which the multiple, nonuniform regions in the preconstancy representation are united into a single UC region in the postconstancy representation. The single uniform region at the superordinate level then corresponds to the UC object in the environment. Note that this "repair" of errors arising from the application of UC at the image level can be achieved simply by applying the same principle of UC to a later representation in which environmental properties of the surface have been recovered. Thus it is not really a problem for the general principle of UC that includes finding uniform surfaces in a postconstancy rep-

resentation, but merely for its initial application at the level of the 2-D image.

The problem of a connected object projecting to two or more disconnected but otherwise uniform regions is also common. It occurs whenever one or more objects partly occlude the given object in such a way that its retinal projection is discontinuous. This form of one-to-many errors cannot be repaired by any of the preceding methods, although depth and constancy processing play important supporting roles. Here the major work is done by classical Gestalt grouping factors, such as good continuation, common fate, similarity of color, and so forth. The basis for unifying such disconnected regions and interpreting their depth has recently been analyzed in some detail by Kellman and Shipley (1991), and the interested reader is referred to their article for further discussion.

It should be emphasized that even if all of these problems could be successfully overcome, it still would not represent a complete solution to the problem of defining the internal correlates of perceived objects in the environment. The reason is that many environmental objects simply do not have UC surface qualities; they often have different parts whose surfaces consist of different colors and textures. An automobile, for instance has black rubber tires, silver-colored bumpers and trim, and a body covered by one or more colors of paint. Each of these differently colored surfaces should be identified as a UC surface in the environment, but their unity as a single object cannot be determined by UC for the simple reason that the automobile is not itself uniform. This problem is therefore outside the scope of any explanation in terms of UC. Some more general notions of postconstancy connectedness must come into play—including what we have called "element connectedness"—because one of the most important features of such multipart objects is their intrinsic connectedness in both space and time.

The additional processes needed to perform such unification of multipart objects are many and varied. Certainly the classical principles of grouping in the late, postconstancy representation will do much of the work. Three-dimensional versions of the factors of good continuation, proximity, similarity, common fate, and so forth, may be sufficient in many cases. But in others it seems likely that even higher level factors come into play, such as prior experience of known classes of meaningful objects. Although early Gestalt theorists in fact mentioned this possibility—indeed, past experience was one of Wertheimer's original "laws" of grouping—it has been neglected because of the emphasis on what are now called "bottom-up" processes. This important factor has been studied only infrequently and not terribly recently (e.g., Leeper, 1935; Rock, 1975; Wallach, O'Connell, & Neisser, 1953). It is, however, currently undergoing systematic examination by Peterson and her co-workers with respect to figure-ground organization (Peterson & Gibson, 1991, 1993; Peterson et al., 1991).

Even so, an enormous amount of the early processing of perceptual organization seems to involve the principle of UC as we have described it. Its importance for a gen-

eral theory of organization is twofold. First, UC constitutes a firm foundation on which other organizational processes can rest. This is what we have called the entry level of perceptual organization. Without it, there would be no units to be grouped or parsed. Second, its logical relations to other organizational processes strongly constrain the way we have come to understand perceptual organization. We believe that figure-ground organization and depth relations must come after the application of UC, for example, and that grouping and parsing must occur later still, most likely in what we have called the late, postconstancy representation. The latter inference is consistent with a growing body of evidence that classical grouping factors are affected by relatively late, postconstancy processes in perception. We believe that understanding perceptual organization within this general framework integrates a large number of disparate phenomena and clarifies the nature of the questions that remain to be answered in future investigations.

REFERENCES

- BANKS, W. P., & KRAJICEK, D. (1991). Perception. In M. R. Rosenzweig & L. Porter (Eds.), *Annual review of psychology, 1991* (Vol. 42, pp. 305-331). Palo Alto, CA: Annual Reviews.
- BARROW, H. G., & TENNENBAUM, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (pp. 3-26). New York: Academic Press.
- BARROW, H. G., & TENNENBAUM, J. M. (1986). Computational approaches to vision. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 38-1 to 38-70). New York: Wiley.
- BECK, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science*, **154**, 538-540.
- BECK, J. (1975). The relation between similarity grouping and perceptual constancy. *American Journal of Psychology*, **88**, 397-409.
- BECK, J., SUTTER, A., & IVRY, R. (1987). Spatial frequency channels and grouping in texture segregation. *Computer Vision, Graphics, & Image Processing*, **37**, 299-325.
- BERGEN, J. R., & LANDY, M. S. (1991). Computational modelling of visual texture segregation. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253-271). Cambridge, MA: MIT Press.
- BIEDERMAN, I. (1985). Recognition by components: A theory of object recognition. *Computer, Vision, Graphics, & Image Processing*, **32**, 29-73.
- BREGMAN, A. S. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 99-118). Hillsdale, NJ: Erlbaum.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- BRUNO, N., & CUTTING, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, **117**, 161-170.
- CAVANAUGH, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, & Image Processing*, **37**, 171-195.
- CAVANAUGH, P. (1988). Pathways in early vision. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary perspective* (pp. 245-289). Norwood, NJ: Ablex.
- GIBSON, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- GILCHRIST, A. L. (1979, March). The perception of surface blacks and whites. *Scientific American*, pp. 112-126.
- GILCHRIST, A. L. (1988). Lightness contrast and failures of constancy: A common explanation. *Perception & Psychophysics*, **43**, 415-424.
- GOLDMEIER, E. (1972). *Similarity in visually perceived forms*. New York: International Universities Press.
- GRAHAM, N., BECK, J., & SUTTER, A. (1992). Nonlinear processes in spatial frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research*, **32**, 719-743.
- GROSSBERG, S., & MINGOLLA, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, **92**, 173-211.
- HOFFMAN, D. D., & RICHARDS, W. (1984). Parts of recognition. *Cognition*, **18**, 65-96.
- HUBEL, D. H., & WIESEL, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, **195**, 215-243.
- HUMPHREYS, G. W., & RIDDOCH, M. J. (1993). Interactions between object and space systems revealed through neuropsychology. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 143-162). Cambridge, MA: MIT Press.
- JONES, D., & MALIK, J. (1992). A computational framework for determining stereo correspondence from a set of linear spatial filters. In G. Sandini (Ed.), *Proceedings of the Second European Conference on Computer Vision* (pp. 395-410). Berlin: Springer-Verlag.
- JULESZ, B. (1975, April). Experiments in the visual perception of texture. *Scientific American*, pp. 34-43.
- JULESZ, B. (1981). Textons: The elements of texture perception and their interactions. *Nature*, **290**, 91-97.
- KANIZSA, G., & GERBINO, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25-32). New York: Springer-Verlag.
- KELLMAN, P. J., & SHIPLEY, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, **23**, 141-221.
- KELLMAN, P. J., & SPELKE, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, **15**, 483-524.
- KIMCHI, R., & PALMER, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **8**, 521-535.
- KIMCHI, R., & PALMER, S. E. (1985). Separability and integrality of levels in hierarchical patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 673-688.
- KÖHLER, W. (1928). An aspect of Gestalt psychology. In C. Murchison (Ed.), *Psychologies of 1925* (pp. 163-195). Worcester, MA: Clark University Press.
- KRAUSKOPF, J. (1963). Effect of retinal image stabilization on the appearance of heterochromatic targets. *Journal of the Optical Society of America*, **53**, 741-744.
- LAND, E. H. (1977, December). The retinex theory of color vision. *Scientific American*, pp. 108-128.
- LEEPER, R. (1935). A study of a neglected portion of the field of learning: The development of sensory organization. *Journal of Genetic Psychology*, **46**, 41-75.
- LIVINGSTONE, M., & HUBEL, D. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, **7**, 3416-3468.
- LOWE, D. G. (1985). *Perceptual organization and visual recognition*. Boston, MA: Kluwer.
- MACK, A., TANG, A., TUMA, R., KAHN, S., & ROCK, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, **24**, 475-501.
- MALIK, J., & PERONA, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, **7**, 923-932.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- MARR, D., & HILDRETH, E. C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London: Series B*, **207**, 187-217.
- MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.
- MARR, D., & POGGIO, T. (1977). Cooperative computation of stereo disparity. *Science*, **194**, 283-287.
- MARR, D., & POGGIO, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London: Series B*, **204**, 301-328.

- NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- PALMER, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 279-307). San Francisco: W. H. Freeman.
- PALMER, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, **9**, 441-474.
- PALMER, S. E. (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, **24**, 436-447.
- PALMER, S. E. (1994). *Late influences on perceptual grouping: Amodal completion*. Manuscript in preparation.
- PETERSON, M. A., & GIBSON, B. S. (1991). The initial identification of figure-ground relationships: Contributions from shape recognition processes. *Bulletin of the Psychonomic Society*, **29**, 199-202.
- PETERSON, M. A., & GIBSON, B. S. (1993). Shape recognition contributions to figure-ground organization and perceived depth in three-dimensional displays. *Cognitive Psychology*, **25**, 383-429.
- PETERSON, M. A., HARVEY, E. M., & WEIDENBACHER, H. (1991). Shape recognition contributions to figure-ground organization: Which route counts? *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 1075-1089.
- ROCK, I. (1975). *Introduction to perception*. New York: Macmillan.
- ROCK, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- ROCK, I. (1986). The description and analysis of object and event perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 33-1 to 33-71). New York: Wiley.
- ROCK, I., & BROSGOLE, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology*, **67**, 531-538.
- ROCK, I., LINNETT, C. M., GRANT, P., & MACK, A. (1992). Perception without attention: Results of a new method. *Cognitive Psychology*, **24**, 502-534.
- ROCK, I., NUHAWAN, R., PALMER, S., & TUDOR, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception*, **21**, 779-789.
- RUBIN, E. (1921). *Visuell wahrgenommene Figuren*. Copenhagen: Glydendalkse Boghandel.
- RUBIN, J. M., & RICHARDS, W. A. (1984). *Color vision: Representing material categories* (Tech. Rep. No. 764). Cambridge, MA: MIT, Artificial Intelligence Laboratory.
- SHEPARD, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 279-341). Hillsdale, NJ: Erlbaum.
- STEVENS, K. A. (1979). *Surface perception from local analysis of texture and contour*. (Tech. Rep. No. 512). Cambridge, MA: Artificial Intelligence Laboratory.
- TREISMAN, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 35-1 to 35-70). New York: Wiley.
- ULLMAN, S. (1984). Visual routines. *Cognition*, **18**, 97-159.
- WALLACH, H., O'CONNELL, D. N., & NEISSER, U. (1953). The memory effect of visual perception of three-dimensional form. *Journal of Experimental Psychology*, **45**, 360-368.
- WERTHEIMER, M. (1923). Untersuchungen zur Lehre von der Gestalt: II. *Psychologische Forschung*, **4**, 301-350. Partial translation in W. D. Ellis (Ed.) (1950). *A sourcebook of Gestalt psychology* (pp. 71-81). New York: Humanities Press.
- WINSTON, P. H. (1975). Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 157-209). New York: McGraw-Hill.
- WITKIN, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, **17**, 17-45.
- ZUCKER, S. W., & HUMMEL, R. A. (1979). Toward a low-level description of dot clusters: Labelling edge, interior, and noise points. *Computer Graphics & Image Processing*, **9**, 213-233.

NOTES

1. The experience error is the mistake of assuming that some aspect of one's perceptual experience on viewing a stimulus is directly avail-

able in or given by the stimulus itself and, therefore, does not require explanation in terms of perceptual mechanisms.

2. We concentrate entirely on perceptual organization in vision in this article, but organization into discrete units is necessary in other sense modalities as well. In addition, for example, Wertheimer (1923) and others since (e.g., Bregman, 1990) have invoked some of the Gestalt principles of grouping to explain how auditory sequences of tones are organized into discrete streams, melodies, and rhythms. But what is the basis for designating the tones as discrete entities? It is plausible to suppose that UC over time in audition plays an analogous role to UC over space in vision; that is, a temporally continuous sound that is uniform in its sensory qualities (e.g., pitch, intensity, timber, spatial location, etc.) will be perceived as a single auditory element. By analogy to vision, such UC events may be identified by auditory "edge detection"—the identification of abrupt changes over time (or onsets and offsets) in audible sensory qualities. See also Bregman (1990, pp. 68-73) for a brief discussion of this problem.

3. This is not to say that the strict geometrical definition of connectedness is necessarily the appropriate one for perception. A tiny gap between two otherwise uniform regions is sufficient to rule out geometrical connectedness, for example, yet it may fail to be detected perceptually, at least initially.

4. This argument against reducing UC to classical grouping principles rests on the implicit assumption that the concept of a "perceptual element" is defined in terms of phenomenological structure. We claim, for example, that the smallest perceptual elements in a given image are its UC regions, such as the dots or lines in Figure 1. This analysis is clearly consistent with the Gestalt phenomenological tradition, but it is also at odds with reducing UC to the action of classical grouping factors. Classical grouping factors can, in fact, account for organization in terms of UC, but only if one assumes that there are smaller perceptual elements within uniform regions. One possibility is that the outputs of individual receptors can be taken as the most fundamental perceptual elements. Then the unity of a homogeneous dot or line can indeed be explained as the result of grouping together receptor outputs on the basis of simple properties such as proximity and color similarity. Grouping factors such as good continuation and closure are not defined for individual receptors, however, and suggest that this is not what Wertheimer had in mind.

Receptor outputs are necessary in this account because they provide a set of smaller "elements" that can be grouped together to form UC regions. The price is steep, however: one must be willing to equate perceptual elements with receptor outputs. Although this is logically possible, it violates some of the most basic Gestalt ideas about perceptual organization. One of the Gestaltists' most urgent objections to structuralist theory was directed at this very assumption: namely, that the activity of individual retinal receptors define the "atoms" of perception. It is therefore difficult to believe that Wertheimer might have implicitly made the same assumption in accounting for the perception of elements via classical principles of grouping. Although there is no doubt that these receptors constitute the most elementary physiological unit in the visual nervous system, this fact by no means guarantees that their outputs correspond to the most basic phenomenological units. This correspondence is unlikely to be valid for a number of reasons, not the least of which is that the outputs of individual retinal receptors are probably not directly experienced at all.

5. This is not universally true, however. Marr, for example, explicitly rejected the idea of segmenting the image into regions corresponding to objects or parts as an inherently ill-formed and inappropriate goal for vision (Marr, 1982, pp. 270-272). The edges represented in his primal sketch were not used to define 2-D UC regions in the image, but to support construction of the 2½-D sketch. Although he strongly implied that segmentation into objects would occur in the 2½-D representation of surfaces' slant-at-a-distance, he did not discuss how it was to be accomplished. As a result, regions appear to play no important role in his theory.

6. Some investigators have argued that the Gestalt laws operate on early, preconstancy representations. For example, Beck (1975) compared grouping based on the projected orientation of line elements to the retina with grouping based on their perceived orientation in the 3-D environment, by slanting an entire array of line segments backward in the third dimension. He found the projected orientation of lines to be

more powerful. Although his experiment was carefully controlled, it was complex, and differing interpretations of the results are possible.

7. Mack, Rock, and their colleagues call these elements "blobs" because they appear to be visual elements of essentially indeterminate shape. They are not to be confused with the "blob" primitives in Marr's (1982) primal sketch, which must be roughly circular to excite the "blob detectors" rather than the detectors of other features, such as bars, edges, and terminators.

8. Outline polygons constructed from thin line segments are perceptually ambiguous in this respect because of an ambiguity in the interpretation of the lines themselves. They can be perceived as thin, solid regions and parsed at interior concave discontinuities as in Figure 15B. However, they can also be interpreted as representing the edges or contours of a solid figure (as in Figure 15A), in which case no parsing occurs.

9. The possibility that equiluminant regions may not support the perception of certain properties—for example, the claim that motion perception does not occur or is attenuated at equiluminance (Livingstone & Hubel, 1987)—will not concern us here.

10. It is worth noting the redundancy in organizational mechanisms provided by the present theory in accounting for this example. The final result—a hierarchy consisting of the two square regions nested within a superordinate rectangular region against the surrounding background—can be achieved by two routes. The first is the one just suggested, using UC to define regions by both luminance and color information. The second is that the three distinct regions defined by UC color could be processed to distinguish the two figural regions from a common ground, after which the two squares could be unified by element connectedness.

11. The starting point for texture analysis is, of course, the structure of the proximal image: how it projects onto the retina. As in the analogous case of uniformly light surfaces, uniformly textured surfaces do not necessarily project to uniformly textured regions in the image. Changes in surface orientation and distance conspire to alter the size, density, aspect ratio, and orientation of texture elements according to projective transformations. The usual result is that their projected regions

are defined by texture gradients rather than uniform texture (Gibson, 1950). A uniformly textured surface that is curved, for instance, projects to a region that is nonuniform in image texture, but is characterized within its boundaries by a smooth transformation in image texture. At the level of image structure, then, it makes sense that textural region analysis proceed by detecting texture edges.

While the textural structure of the image is certainly the appropriate starting place for an analysis of texture, the visual system almost certainly achieves what might be called *texture constancy* in most cases—that is, the tendency to perceive a uniformly textured surface in the environment as uniformly textured, despite differences in its projected image texture resulting from viewing conditions, such as differences in surface orientation and distance with respect to the viewer. This means that at this later postconstancy level of analysis, uniformly textured surfaces will be represented as uniformly textured, despite texture gradients at the earlier, preconstancy level. Consistent with the analysis of luminance and color given above is our suggestion that image texture receives a dimensional analysis in which certain dimensions are factored out, corresponding to perspective changes that affect the orientation, aspect ratio, and size scale of image texture, leaving intrinsic textural dimensions that correspond to the distal properties of surface texture. At this later stage, then, texture gradients are no longer a factor, and uniformly textured surfaces will be represented as uniformly textured. Although there have been studies of texture perception in the computational literature that address the issue of textural structure arising from changes in surface orientation (e.g., Stevens, 1979; Witkin, 1981), most psychological studies of texture perception have not adequately considered these effects (but see Bruno & Cutting, 1988).

12. For example, Jones and Malik (1992) have recently shown that stereoscopic fusion can be achieved efficiently on essentially objectless images by using information from a local analysis of spatial frequency.

(Manuscript received June 21, 1993;
revision accepted for publication September 18, 1993.)