University of Kent

Kent Academic Repository

Palmer, Stephen E., Brooks, Joseph L and Nelson, Rolf (2003) *When does grouping happen?* Acta Psychologica, 114 (3). pp. 311-330. ISSN 0001-6918.

Downloaded from

https://kar.kent.ac.uk/29191/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1016/j.actpsy.2003.06.003

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

This is the author's final, peer-reviewed version that was accepted for publication. The final publisher's formatted version of the article can be found at:

http://dx.doi.org/10.1016/j.actpsy.2003.06.003

WHEN DOES GROUPING HAPPEN?

Stephen E. Palmer, Joseph L. Brooks, and Rolf Nelson

Please send correspondence to:

Stephen E. Palmer
Psychology Department
University of California
Berkeley, California 94720

510-642-7135 (office)
510-525-8816 (home)
510-642-5293 (fax)

palmer@cogsci.berkeley.edu

ABSTRACT

Recent research on perceptual grouping is described with particular emphasis on identifying the level(s) at which grouping factors operate. Contrary to the classical view of grouping as an early, two-dimensional, image-based process, recent experimental results show that it is strongly influenced by phenomena related to perceptual constancy, such as binocular depth perception, lightness constancy, amodal completion, and illusory contours. These findings imply that at least some grouping processes operate at the level of phenomenal perception rather than at the level of the retinal image. Preliminary evidence is reported showing that grouping can affect perceptual constancy, suggesting that grouping processes must also operate at an early, preconstancy level. If so, grouping may be a ubiquitous, ongoing aspect of visual organization that occurs for each level of representation rather than as a single stage that can be definitively localized relative to other perceptual processes.

PsychINFO: 2323

Keywords: perceptual organization, grouping, perceptual constancy, binocular depth perception, lightness constancy, amodal completion, illusory contours

Perceptual grouping refers to the processes that are responsible for determining how the part-whole structure of experienced perceptual objects (such as people, cars, trees, and houses) is derived from the unstructured data in retinal images. When an observer views a scene containing an automobile, for example, how is it perceived as a single object? Why are the tires seen as belonging with the doors, windshield, hood, and trunk rather than as entirely separate objects or as belonging with the road?

Gestaltist psychologists were the first to recognize the ubiquity and centrality of this "grouping" problem to perception. Max Wertheimer took a giant step forward in his ground-breaking 1923 article by determining some of the stimulus factors that govern this phenomenon, which are the famous "laws" (or, more accurately, "principles" or "factors") of grouping. Figure 1 illustrates several of these principles -- including proximity, similarity (of color, size, and orientation), common fate, good continuation, and closure -- in demonstrations similar to the ones Wertheimer originally used. The principles of grouping he articulated are among the best known, yet least understood, phenomena of visual perception.

Recent findings have added several new principles of grouping to this list, including common region, element connectedness, and synchrony. Common region (see Figure 2A) is the tendency for elements that lie within the same bounded area to be grouped together, as the spots of a leopard are grouped within its contours (Palmer, 1992). Element connectedness (see Figure 2B) is the tendency for elements that share a common border to be grouped together, as the head of a hammer is grouped with its handle (Palmer & Rock, 1994). Synchrony is the tendency for elements that change at the same time to be grouped together (e.g., Lee & Blake, 1999; Palmer & Levitin, 1998). It is related to the classical principle of common fate, except that the simultaneous changes do not have to involve motion or to be "common" in any sense. Why events should be grouped by synchrony is somewhat mysterious from an ecological standpoint, because everyday examples of grouping solely by synchrony are difficult to find.

There are a variety of different theories about how and why grouping arises. The original Gestalt ideas about these issues centered on their articulation of the principle of Prägnanz -- that grouping provided a percept that was in some sense "simpler" and "better structured" than the corresponding ungrouped (or differently grouped) percept. Unfortunately, they were not very clear about just what this meant. Even so, the rather fuzzy Gestalt idea of Prägnanz has been sharpened and extended in clearer, better specified theories, most notably Leeuwenberg's structural information theory (e.g., Buffart, Leeuwenberg & Restle, 1981; Leeuwenberg, 1971; van der Helm &

Leeuwenberg, 1991; van Lier, van der Helm, & Leeuwenberg, 1994). It explicates formal rules for determining which of all possible interpretations (in this case, groupings) are the "best" in the sense of having minimal information content according to well-defined criteria.

In its classical form, Leeuwenberg's structural information theory (SIT) is what Marr (1982) called a "computational-level" theory: It does not attempt to specify the actual processes that produce grouping in perception, but only the input-output mapping between images and organizations. The primary question we address in this article is not the computational-level question of which theory might be most compatible with known grouping phenomena, but the process-oriented question of when grouping occurs relative to other perceptual processes. The question we address is related to certain issues in computational-level theories, however, such as the nature of the representation on which grouping operations are based. For example, do the redundancy-elimination rewrite rules of SIT operate on the retinal properties of physical stimulation (e.g., image-based luminance and size) or on the perceived properties of visual objects (e.g., surface lightness and 3-D size)?

The perceptual processes that underlie classical grouping phenomena have generally been assumed, although perhaps only tacitly, to be relatively primitive, low-level operations that operate on some early, 2-D representation to create a set of discrete elements on which subsequent perceptual operations are performed. This view, which has been widely held by prominent visual researchers (e.g., Kahneman & Henik, 1981; Marr, 1982; Neisser, 1967) is often justified on the grounds that grouping must occur early because the groups it produces are generally thought to be required to achieve perceptual constancy. If so, grouping logically must occur before the various processes that support perceptual constancy, such as binocular depth computations, surface lightness perception, and the completion of partly occluded objects.

The "early view" that grouping occurs prior to constancy operations can be cast in a variety of ways. The most extreme version of this idea, which we will call the "early-only" view, is that grouping occurs exclusively at an early, preconstancy level. In the first part of this article we will review recent evidence that this view cannot be correct. The opposite extreme view, which we will call the "late only" view, is that grouping occurs exclusively at a relatively late postconstancy level. Toward the end of this article we will describe some preliminary evidence that this view cannot be correct either. We will argue that the most reasonable conclusion is that grouping operates at multiple stages in visual processing, both before and after constancy processing, and that this multistage view should be considered as the basis of future theories of perceptual grouping.

Before we delve into the heart of the argument, it is perhaps worth considering why we are so preoccupied with the location of grouping relative to constancy processing. The reason is that, in our opinion, constancy provides the single most crucial landmark in visual processing. It is the set of visual operations whose presumed job it is to convert visual representations that encode image-based (retina-based) features into ones that encode environment-based (object-based) features. Although there is as yet no clear consensus about the precise nature of postconstancy representation (e.g., it might encode 2.5-D surfaces, 3-D objects, or both), many of the most crucial inferences the visual system must make are concerned with the logical leap from 2D representations to some more ecologically useful representations that contain explicit information about properties of external, environmental objects. Accomplishing these inferences according to some optimizing or satisficing criterion is the job of constancy processing, which therefore occupies a particularly prominent position in perception.

THEORETICAL CONSIDERATIONS

Palmer and Rock (1994) initially challenged the early-only view of grouping on purely theoretical grounds. First, they pointed out that although Wertheimer's demonstrations (see Figure 1) involved putting together two or more discrete elements, he never actually said where the elements themselves came from. Presumably he believed that they were somehow derived from the grouping principles he articulated in his 1923 article, but Palmer and Rock argued that they arise from a different kind of organizational principle that they called uniform connectedness. Uniform connectedness (UC) is the principle by which the visual system partitions the image into connected regions having uniform (or smoothly changing) properties, such as luminance, color, motion, and texture. The result of organization according to UC is the partition of the image into a nonoverlapping set of regions, much like a stained-glass window.

According to Palmer and Rock's theory (see Figure 3), UC regions do not acquire the status of distinct visual elements until after figure-ground organization determines which ones correspond to phenomenal objects and which ones to backgrounds or spaces between objects. Once figural regions have been designated as entry-level perceptual elements, they can then be aggregated into larger, superordinate units by principles of grouping, or they can be divided into smaller, subordinate units by being parsed at deep concavities, where the contour curves sharply inward (e.g., Hoffman & Richards, 1984).

Notice that Palmer and Rock's (1994) reasoning places classical perceptual grouping operations somewhat farther along the chain of visual information processing than had previously been assumed, after region segmentation and figure-ground organization have already provided a set of perceptual elements. Because figure-ground processing can be viewed as a form of depth perception through pictorial cues to determine what is in front of what and to which region the boundaries belong (Palmer, 1999), Palmer and Rock's analysis suggests that grouping may actually occur after depth perception and various other forms of constancy have been achieved.

Even so, the level at which grouping processes operate is ultimately an empirical question. Despite its importance, few experiments have been directly concerned with answering it until recently. In this part of the article, we review some evidence showing that grouping is not an exclusively early (preconstancy) process.

BINOCULAR DEPTH EFFECTS

The importance of binocular depth in determining perceived grouping is demonstrated rather dramatically in Figure 4. The three images can be fused with either crossed or uncrossed disparity to produce two stereoscopic images in depth. Before fusing them, however, notice that in the middle display, the central column groups more strongly with the columns on the left than with those on the right due to their closer 2D proximity. In the leftmost and rightmost displays, the central column is equally distant from the left and right sides, so that it does not group differentially with either side. Any grouping of the central column based on 2D proximity would therefore have to predict that it should group to the left. Nevertheless, once the display has been fused stereoscopically, the central column is perceived clearly and unequivocally as grouped with the elements on the right, and this is true in both of the stereoscopic images. The reason for this grouping is that binocular disparity reveals the central elements to be in the same depth plane as the ones on the right and in a different depth plane than the ones on the left. Clearly, this demonstration supports the claim that it is 3D grouping in depth that matters, once stereoscopic fusion has been achieved.

The radical difference in grouping based on viewing Figure 4 monocularly versus binocularly thus illustrates that perceived grouping is strongly influenced by stereoscopic depth. In terms of determining when grouping occurs, it implies that grouping cannot occur only before stereoscopic depth perception because, if it did, no effects of stereopsis on grouping would be possible. This conclusion leaves open several theoretical possibilities:

grouping may occur only after stereopsis, both before and after stereopsis, during stereopsis, or any combination of these alternatives. We have ruled out only the strongest form of the early grouping hypothesis.

The influence of binocular depth perception on proximity grouping was studied experimentally by Rock and Brosgole (1964). They asked whether the distances that govern proximity grouping are defined in the 2-D image or in perceived 3-D space. Observers were shown a 2-D array of luminous beads in a dark room either in the frontal plane (perpendicular to the line of sight; see Figure 5A) or slanted in depth so that the horizontal dimension of the array was compressed (Figure 5B). The beads were actually closer together vertically than horizontally, so that when they were viewed in the frontal plane, observers always reported seeing them organized into columns rather than rows. The crucial question was how the beads would be grouped when the same lattice was viewed slanted in depth so that the beads were retinally closer together in the horizontal direction.

The answer depended importantly on whether the viewing conditions were monocular or binocular. When viewed monocularly so that the array looked like it was oriented in the frontal plane, observers reported seeing the beads grouped into rows, as predicted by retinal proximity. When the same display was viewed binocularly so that it looked like it was slanted in depth, however, observers reported seeing the beads grouped into columns. This happened because the beads now appeared to be closer in the vertical direction, as was actually the case in the 3-D world. These results thus support the hypothesis that final grouping must occur after binocular depth perception.

Analogous conclusions about the effect of binocular depth are supported for grouping by the factors of common region (Figure 6A) and element connectedness (Figure 6B). In Figure 6A, each half of the stereogram alone exhibits no differential grouping with one versus the other set of overlapping ellipses. When the two images are cross-fused binocularly, however, the resulting binocular perception shows that the black circles group strongly within the ellipses that are in the same depth plane and not with the ellipses that float above or below them. Figure 6B demonstrates the analogous effect for grouping by element connectedness. Once binocular fusion is achieved, the gray squares are seen to group into pairs according to the connecting bars that lie in the same depth plane (those on the left side), with the other bars (on the right) floating in a plane above or below them. Clearly, what matters for the final perception of grouping is the enclosure and connectedness of the elements in 3-D perceived space.

The corresponding question in the domain of lightness perception is whether the important factor in grouping by achromatic (grayscale) similarity is preconstancy retinal luminance or postconstancy perceived lightness. Rock, Nijhawan, Palmer, and Tudor (1992) answered it by using cast shadows and translucent overlays to disentangle the two possibilities. Observers were shown displays containing five columns of squares (see Figure 7A) and asked to report whether the central column grouped with those to the left or right. The critical display was carefully constructed so that the central squares were identical in reflectance to those on the left (because they were made of the same shade of gray paper), but they were seen behind a strip of translucent plastic that rendered their retinal luminance identical to the squares on the right (see Figure 7A). Thus, if grouping were based on retinal luminance, the central squares should be grouped with the luminance-matched ones on the right. If it were based on processing after transparency perception had been achieved, they would group with the reflectance-matched ones on the left. All observers reported them grouping with the reflectance-matched ones on the left. In another condition, the same luminances were achieved by casting a shadow over the central column of squares. The results for both the transparency and shadow conditions supported the postconstancy hypothesis: grouping is based on the perceived lightnesses of the squares rather than on their retinal luminances.

There is an important alternative explanation of this result, however, that must be ruled out before we can accept the claim that grouping is influenced by postconstancy lightness perception. The alternative is that grouping might be based on retinal luminance ratios, rather than absolute luminance values, because this hypothesis also predicts the obtained outcome. To discriminate between these two possibilities, Rock et al. studied the further condition shown in Figure 7B, which does not produce the perception of either shadows or transparency. In this display, observers perceive the central squares as lying in front of an opaque background (rather than in a shadow or behind a translucent strip) such that their lightnesses are now the same as the elements on the right side rather than those on the left. The grouping reported by observers actually reverses for this condition so that the central squares are seen as grouped with the ones on the right. Notice that the luminance ratios between the elements and their local surrounds are exactly the same as in Figure 7A; the only thing that differs is the edge information where the border and the central strip meet. When that information is consistent with a shadow or a transparent surface covering the central column, its elements are grouped with the lighter ones on the left; when it is consistent with them occluding an opaque strip behind them, they group with the darker ones on the right. The results are therefore consistent with

the claim that grouping is strongly influenced by lightness constancy and are not consistent with the alternative explanation that it is determined by retinal luminance ratios.

AMODAL COMPLETION

In a further experiment Palmer, Neff, and Beck (1996) examined whether grouping is influenced by amodal completion of partly occluded objects. When a simple object is partly occluded by another, its shape is completed without sensory experience ("amodally") behind the occluding object. This process is widely believed to occur during or after the perception of relative depth relations among objects based on the pictorial cue of interposition or occlusion (e.g., Rock, 1983). Is grouping by shape similarity determined by the retinal shape of uncompleted elements, as predicted by the early-only view, or is it influenced by the perceived shape of the completed elements, as predicted by any of several late views?

Grouping effects were measured using the central-column grouping task described above when the central column contained half-circles. When the straight sides of these half circles are presented abutting an opaque rectangle (see Figure 8A), they are usually perceived as whole circles completed amodally behind the rectangle. The early-only view predicts that the central elements should group with the half-circles on the right because they have the same retinal shapes, whereas a late view predicts that they will group with the full circles on the left because they have the same perceived shapes. Clearly, they group more strongly with the whole circles. Unfortunately, common region predicts the same result in this display due to the division of space by the occluding rectangle, so these two factors were decoupled in several additional conditions. The crucial manipulation was to displace the occluding strip a little farther to the side so that the half-circular shape of the central elements could be unambiguously perceived (Figure 8B). Now the central elements group more strongly with the half circles to the right. Palmer, Neff and Beck's (1996) experiment independently varied these two factors, and its results showed that both completed shape similarity and common region influence perceived grouping. This finding supports the conclusion that grouping by shape similarity is strongly influenced by the perceived shape of amodally completed objects. It is therefore incompatible with the early-only approach to grouping.

ILLUSORY FIGURES

Illusory figures are perceived where inducing elements, such as the notched ovals in Figure 9A, are positioned so that their contours align to form portions of the edges of a closed figure. The completed perception is of a figure that has the same surface characteristics as the background and that occludes parts of the inducing elements. The crucial question is whether grouping occurs only before the perception of illusory figures, as would be predicted by the early-only view, or whether grouping has a component that operates after the formation of illusory contours, as expected from any form of late view.

Palmer and Nelson (2000) investigated whether grouping can occur after the perception of illusory contours and figures. Again, the task was to decide whether a central column of elements grouped with the columns of elements on the right side or on the left side (see Figure 9). In the example shown in Figure 9, the inducing elements are horizontal ovals in the left six columns and vertical ovals in the right four columns. In their unnotched versions, the central two columns of ovals unequivocally group to the left (Figure 9B). When the ovals have been notched as in Figure 9A so that illusory rectangles are perceived, the central column of vertical illusory rectangles groups strongly to the right with the other vertical illusory rectangles, opposite to the grouping of the inducing elements themselves. To be sure that this grouping is not due simply to the nature of the individual notched elements, Figure 9C shows a control condition in which half of the elements have been rotated 180° so that the perception of illusory figures is weak or absent. In this condition about equal numbers of observers saw the central columns group to the left and right. The striking difference between the grouping evident in Figures 9A and 9C can thus be attributed to the fact that grouping is strongly affected by the perception of illusory figures.

All of these findings point to the same conclusion: Phenomenally perceived grouping – that is, the final result of underlying grouping processes -- is not governed solely by the structure of early, preconstancy retinal images, but includes influences of relatively late, postconstancy perceptions. This fact categorically rules out the early-only view in which grouping processes occur only at a 2-D, preconstancy level.

The critical unresolved problem is to determine which of the "late" theories is correct among the following three types, all of which are consistent with the findings described above.

(1) Late-only theories: Grouping processes may work only after constancy has been achieved.

(2) Early-and-late theories: Grouping processes may occur at two (or more) levels, both preceding and following the achievement of constancy.

(3) Feedback theories: Grouping processes may be part of a cascade of temporally overlapping processes that begins prior to constancy operations, but receives postconstancy feedback that alters the initial grouping results.

In the latter two cases, early grouping at the image-processing level would provide a preliminary organization that could be used to bootstrap the higher-level processes involved in constancy. The results of these constancy computations would then be used to modify the provisional two-dimensional organization that arose from image-based grouping processes, so that the final organization conforms to the perceived properties that result from constancy operations.

TESTING THE LATE-ONLY VIEW

Among the classes of "late" theories, the "late-only" version is the easiest to test, because it can be categorically eliminated if grouping processes can be shown to operate before as well as after constancy processing. There are a number of ways to try to show this. One is to try to prevent constancy from occurring by using brief, masked presentations, and then seeing whether this changes the nature of the grouping people perceive. This approach has been taken by Schulz and Sanocki (2000) who investigated the effects of presentation duration on color constancy in an experiment similar to that reported by Rock et al. (1992). The idea is that if processing can be stopped by masking a brief presentation before constancy occurs, then the observed organization would directly reflect whatever grouping is present in the preconstancy representation. Schulz and Sanocki found that under brief masked presentation conditions, grouping followed the preconstancy predictions, consistent with the idea that grouping occurs before as well as after lightness constancy. As exposure duration increased, grouping followed postconstancy predictions, as Rock et al. (1992) reported for lightness constancy under unlimited viewing conditions. Schulz and Sanocki interpreted their results as indicating that grouping occurs before as well as after color constancy processing.

In the remainder of this article, we will describe preliminary evidence that also supports the existence of preconstancy grouping, but from a different logical and methodological perspective. There we argued that if constancy affects grouping, then at least some grouping process must occur after constancy procerssing begins. The inverted logic is this: If grouping can be shown to influence constancy processing, then there must be at least some

grouping process that operates before constancy is complete. We will now describe some preliminary evidence that this is the case.

GROUPING AND SHAPE CONSTANCY

One case in which we have good evidence that grouping operates before the final perception of constancy is for the property of shape. Shape constancy refers to people's tendency to perceive the relatively constant 3-D shapes of objects rather than the highly variable 2-D shapes of their retinal projections, which change whenever the direction of gaze changes relative to the object. An ellipse on the retina, for example, is often ambiguous in shape because it can be perceived either as an ellipse in the frontal plane or as a circle slanted in depth. The crucial question is whether grouping this ambiguous stimulus with a less ambiguous contextual element in a visual display can influence whether people perceive it as a circle or as an ellipse.

Palmer and Brooks are studying this question using displays like the ones shown in Figure 10. A central ellipse is surrounded by two quadrilaterals, one of which is a square and the other of which is a trapezoid. People have a strong tendency to see the square as a square in the frontal plane, but the trapezoid as a square (or rectangle) slanted in depth. The example shown in Figure 10A is relatively ambiguous; the central ellipse can be perceived rather easily as either an ellipse in the frontal plane or as a circle slanted in depth. The idea behind the present experiments is that if the ellipse is grouped with the square, it should tend to be seen as an ellipse lying in the frontal plane with the square. If it is grouped with the trapezoid, however, it should tend to be seen as a circle slanted in depth. We manipulated grouping by varying the proximity, color similarity, and common fate relations between the ellipse and the two contextual figures.

Figures 10B and 10C show two examples that use proximity and color similarity to influence grouping. (For further examples, including dynamic displays using common fate, the reader is invited to visit our website at <http://socrates.berkeley.edu/~plab/projects.html>) or the ACTA website (http://……). In Figure 10B, the ellipse is proximal to and the same color as the square, whereas it is farther from and different in color from the trapezoid. It should therefore tend to group with the square and therefore tend to be seen as an ellipse lying in the frontal plane. In Figure 10B, however, the ellipse is proximal to and the same color as the trapezoid, whereas it is farther from and different in color from the square. It should therefore group with the trapezoid and thus be seen as a circle lying in a slanted depth plane. We also constructed displays in which the central ellipse was moving harmonically up and

down in synchrony with either the square or the ellipse (while the other element was stationary) to manipulate grouping by common fate.

We showed observers displays that employed all possible combinations of these three factors (proximity, color similarity, and common fate) and asked them to indicate on each trial whether they perceived the central figure to be a circle in depth or an ellipse in the frontal plane. From the data we have collected thus far, it is clear that all three of these grouping factors have large influences on the perceived shape of the central ellipse. When the square was more proximal than the trapezoid, for example, observers reported seeing an ellipse on 78% of the trials, but when the trapezoid was more proximal, they reported seeing a slanted circle on 74% of the trials. The proximity grouping effect thus effectively reversed the perception of shape in this context. The effects of common fate were similarly powerful; those of color similarity were somewhat less potent, but still quite clear. The results leave little doubt that grouping factors strongly affect the outcome of shape constancy processing, thus contradicting the late-only view.

GROUPING AND FIGURE-GROUND ORGANIZATION

One of the most fundamental processes in several forms of perceptual constancy is the assignment of relative depth across an edge: Which side of a given boundary is closer and which side farther relative to the observer? This is perhaps the single most important feature of figure-ground organization, in which the closer side is perceived as a "figure" against a farther "ground" as described by Rubin (1921) in his classic monograph on this phenomenon. The critical question for the present discussion is whether grouping might play an important role in determining which side of a depth edge is the closer, figural side by causing the edge to group more strongly with one side rather than with the other.

In standard figure-ground displays, there is no basis on which grouping can work because the regions are typically homogeneous. If textural elements are visible within the regions, however, then there is the possibility that the elements on one side may group more strongly with the edge than those on the other due to classical grouping factors. Consider the case of grouping by common fate, for example. Suppose that the edge between two regions moves back and forth and that the texture elements on one side of the edge move back and forth synchronously with it, whereas those on the other side are stationary (see Figure 11A). Common fate predicts that the moving elements should group with the edge. If they do, then the side with the moving elements should be perceived as closer than

the side with stationary elements, because the edge is perceived as "belonging to" the moving side. (Dynamic displays of this phenomenon can be found on our website at <http://socrates.berkeley.edu/~plab/projects.html> and the ACTA website (http://…….)

Palmer and Brooks are investigating several cases in which grouping factors should influence figure-ground perception in this way, and in all cases that we have tested thus far, they do. Common fate is the most dramatic example. When the texture on one side moves together with the edge and the texture on the other side does not move (Figure 11A), every observer thus far has reported that the moving appears to be the closer figural side, even when no texture elements are occluded by the moving edge. To be sure that this was not simply due to moving texture attracting attention or being seen as closer for some other reason, we also st udied similar displays in which the edge was stationary (Figure 11B). In this case, the grouping hypothesis predicts that the side with the stationary elements should be seen as closer because these unmoving elements now have the same motion as the unmoving edge (i.e., no motion at all) . This is just what happens. The vast majority of our observers report that the moving side now appears to be behind the stationary side, thus reversing the results of the moving edge condition.

Similar findings about motion effects on perceived depth were reported by Yonas, Craton, and Thompson (1987), but not in the context of grouping. Our grouping hypothesis suggests that grouping factors other than common fate should produce similar results, and preliminary data suggest that they do. We are examining grouping by color similarity using a red or blue boundary line between two white regions, one of which contains red texture elements and the other of which contains blue texture elements. When the border is red, the side with red texture elements tends to be perceived as figural, and when the border is blue, the side with blue texture elements tends to be perceived as figural, consistent with the grouping hypothesis. These effects seem to be less potent than the motion effects described above, but they are statistically reliable. We are also finding effects due to proximity (the side with texture elements closer to the border tends to be seen as figural), orientation (the side with texture elements parallel to the orientation of the border tends to be seen as figural), and synchrony (the side with texture elements that change luminance synchronously with the border tends to be seen as figural).

The most obvious conclusion is that the grouping of an edge with regional texture does indeed affect depth perception across an edge. Because depth edge assignment is an important aspect of many different forms of constancy, this finding supports the further conclusion that grouping operates prior to the completion of depth and constancy processing.

It is perhaps worth mentioning that the conception of grouping that we advocate here -- namely, grouping between edges and texture elements of a region -- is somewhat unorthodox. Grouping is a relation that normally holds between two or more perceptual objects, but edges are not usually considered perceptual objects. They have been important theoretical constructs in visual processing ever since Hubel and Wiesel (1959) first introduced "edge detectors" into the vocabulary of visual theory, but they are not usually considered the sort of independent elements that could enter into grouping relations. We do not see why not, and the results we are finding suggest that it may be useful to think of them in this way.

GROUPING AND LIGHTNESS CONSTANCY

The analogous early grouping claim in the domain of lightness perception would be that grouping affects the achievement of lightness constancy. There are a number of previously reported results in the lightness perception literature that support a closely related claim: namely, that grouping affects lightness contrast. (Lightness contrast is the tendency for people to see an element of a given lightness value as lighter in a context of a dark surround and as darker in a context of a light surround.) Agostini and Proffitt (1993), for example, showed that when a medium gray dot moves in common fate with a field of black dots, it looks lighter than an identical gray dot that moves in common fate with a field of white dots. The conclusion that the gray dots' lightnesses are perceived relative to the dots within their own group is consistent with an early grouping hypothesis, and it is sufficient to explain the observed contrast effect. There are quite a few such reports of strong and consistent grouping effects producing contrast phenomena that are otherwise difficult to explain (e.g., Adelson, 1993; Gilchrist, Kossyfidis, Bonato, & Agostini, 1999; Todorovic, 1997; White, 1979).

Such effects are not definitive evidence that grouping occurs before constancy processing, however, because lightness contrast and lightness constancy are not the same process. Rather, it seems possible that contrast effects occur after constancy -- effectively operating on the postconstancy representation -- in which case grouping might conceivably occur after constancy is achieved, but before contrast mechanisms come into play. It is therefore important to show direct influences of grouping on lightness constancy.

As anecdotal evidence for this claim, consider an experience that one of us had suggesting that grouping does, in fact, influence lightness constancy rather directly. After exercising at the gym one day, Palmer looked upward into his gym locker at his shirt, which was hanging from a hook at the top. At first, the shirt looked like it

had a dark spot spreading down from where it was suspended by the hook, as though dirt or rust from the hook had stained the shirt. When he grabbed the shirt below the spot and lifted it, however, the edge of the dark spot didn't move upward with the shirt as the edge of a stain would have, but stayed fixed relative to the locker as a shadow would have. Palmer reports that he immediately perceived (correctly and without conscious thought) that the initially perceived "stain" was actually just a shadow cast by the top of the locker. The fact that the dark spot didn't group with the shirt by common fate indicated that it was unlikely to be a reflectance edge caused by a stain, and the fact that it did group with the locker by static common fate suggested that it was an illumination edge (shadow) cast by the locker.

As in the case of grouping effects on figure-ground and depth across an edge, this analysis is unconventional in that it requires treating edges as elements that can be grouped with other edges. But why not? We currently believe that such edge-grouping processes may play a significant role in lightness constancy processing by helping to disambiguate luminance edges as either reflectance or illumination edges. Although common fate is probably the most powerful grouping factor in disambiguating the interpretation of luminance edges, other grouping factors should operate similarly. The critical edge should look more like a reflectance edge if it is grouped withnearby edges that are unambiguously due to reflectance, and it should look more like an illumination edge if it is grouped with nearby edges that are unambiguously due to illumination. We have not yet performed the relevant experiments and so do not have any hard data to support these predictions, but we expect to test these predictions in the near future.

THEORETICAL IMPLICATIONS

We take these grouping effects on shape constancy, figure-ground edge assignment, and lightness constancy as preliminary evidence supproting some version of an early grouping hypothesis: namely, that at least some grouping operations occur before depth and constancy processing are complete. Our previous findings indicate the opposite conclusion: namely, that at least some depth and constancy processing occurs before grouping is complete. How can we reconcile this apparent conflict?

Perhaps the simplest possibility would be that grouping occurs at least twice, once before and again after depth and constancy processing is complete in a feedforward architecture. The idea is that there might be two discrete representations -- a preconstancy representation of 2-D image-based features and a postconstancy

representation of 3-D object-based features -- and grouping operations might take place after each representation is constructed. There might also be further representations, such as a category-based one, that would, in turn, induce further grouping changes based on internal knowledge of the relevant categorical type.

Another possibility is that grouping may be an integral part of depth and constancy processing, within which it may work iteratively as part of a feedback loop. The idea here is that there may be a single representation that is continuously updated as depth and constancy processes revise perceptual estimates of environmental properties. Grouping would thus initially work on image-based features and later on object-based features as depth and constancy processing complete their work. In this conception there is only one visual representation and only one grouping process, but they both work together to alter the content of the representation to reflect environmental structure more faithfully as processing progresses.

This sort of iterative architecture would be representationally efficient, but it might be a difficult one within which to compute effectively. The problem is that unless the updating is temporally coherent, such that all parts of the representation are updated at the same time, there is the possibility that grouping and/or constancy processing would be trying to take account simultaneously of 2-D image-based information in some parts of the representations and of 3-D object-based information in others. Such confusions would not occur within a feedforward architecture because it contains separate representations and processing of image-based and object-based features, but at an added cost.

Before closing, it is important to note a limitation of the conclusions that can be drawn about when grouping happens relative to constancy processing. Although perceptual constancy is a relatively coherent theoretical concept, it is surely not a single process that occurs at a discrete time or place in visual analysis. Let us consider depth perception as an example because it is a crucial part of almost all constancy processing. The problem is that depth perception relies on many diverse components, as any introductory textbook worth its salt will demonstrate. The processing of binocular disparity information appears to be largely an early data-driven process that occurs in cortical areas V1 and V2, probably without much high-level feedback. In contrast, the analysis of depth that comes from at least some of the so-called pictorial cues is likely to place much later in processing with substantial high-level feedback. Depth from familiar size, for example, requires assigning objects to basic level categories, a process currently thought to happen somewhere in inferotemporal (IT) cortex. Depth processing therefore appears to occur over a wide range of the ventral pathway, from V1 to IT. The question of when grouping

happens relative to depth perception (in general) must therefore be considered only a relatively crude indication of when perceptual grouping happens and any conclusion is necessarily conditioned by the imprecision of the landmark that is being used. At the very least, one must consider the kind of depth information that is being used to test inferences about when grouping happens relative to depth perception (e.g., binocular disparity versus familiar size).

The previous paragraph raises an important challenge that must be faced in understanding the issues that we have been discussing: namely, the relation between psychophysical findings (such as we have outlined here) and underlying physiological mechanisms. Given the well-known ordering of ascending visual projections (e.g., retina, LGN, V1, V2, ... IT) and the discovery of the locus of cells that are relevant to depth and constancy processing (e.g., the cells in V2 that von der Heydt and Peterhans (1989) identified that respond to illusory contours), it is tempting to try to translate the terms "early" and "late" directly into physiological descriptions, such as "before V2" and "after V2." The problem is that the well documented, massive, backward connections from higher levels to lower levels throughout the visual system make such translations difficult, if not impossible. Processing that goes on in V2 might be functionally either early or late, depending on whether it happens without or with the benefit of feedback from higher levels and from which higher levels it might receive feedback (see, e.g., Hochstein & Ahissar, 2002). The precise relation between the burgeoning literature on the physiology of the visual system and the kind of functional analysis we have discussed here thus constitutes a difficult, but important, area for future research.

Author Note

Requests for reprints and other correspondence should be sent to Stephen E. Palmer, Psychology Department, University of California, Berkeley, CA 94720-1650. The research reported in the first half of this article was supported in part by Grant 1-R01-MH46141 from the National Institute of Mental Health to the first author.

References

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. Science, 262, 2042-2044.

Agostini, T., & Proffitt, D. R. (1993). Perceptual organization evokes simultaneous lightness contrast. Perception, 22(3), 263-272.

Buffart, H., Leeuwenberg, E., & Restle, F. (1981). Coding theory of visual pattern completion. Journal of Experimental Psychology: Human Perception and Performance. 7, 241-274.

Gilchrist, A., Kossyfidis, C., Bonato, F., & Agostini, T. (1999). An anchoring theory of lightness perception. Psychological Review, 106, 95-834.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron, 36, 791-804.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. Cognition, 18, 65-96.

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

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. (pp. 181-211) Hillsdale, NJ: Erlbaum.

Lee, S., & Blake, R. (1999). Detection of temporal structure depends on spatial structure. Vision Research. 18, 3033-3048.

Leeuwenberg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. American Journal of Psychology, 84, 307-349.

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

Neisser, U. (1967). Cognitive Psychology. New York: Appleton-Century-Crofts.

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

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

Palmer, S. E., & Levitin, D. J. (1998). Synchrony: A new principle of perceptual grouping. Paper presented at the 39th Annual Meeting of the Psychonomic Society, Dallas, TX.

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

Palmer, S. E., & Nelson, R. (2000). Late influences on perceptual grouping: Illusory figures. Perception and Psychophysics, 62 (7), 1321-1331.

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

Rock , I. (1983). The logic of Perception. Cambridge MA: MIT-Press.

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

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

Rubin (1921). Visuell Wahrgenommene Figuren [visually perceived patterns]. Kobenhaven: Glydenalske boghandel.

Schulz, M. F., & Sanocki, T. (2003). Time course of perceptual grouping by color. Psychological Science, 14, 26-30.

Todorovic, D. (1997). Lightness and junctions. Perception, 26(4), 379-394.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1991). Accessibility: A criterion for regularity and hierarchy on visual pattern codes. Journal of Mathematical Psychology, 35, 151-213.

van Lier, R.J., van der Helm, P.A., & Leeuwenberg, E.L.J. (1994). Integrating global and local aspects of visual occlusion. Perception, 23, 883-903.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. Journal of Neuroscience, 9(5), 1731-48.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt [On Gestalt Theory]. Psychologische Forschung, 4, 301-350. Partial translation in W. D. Ellis (Ed.), 1950, A sourcebook of Gestalt psychology (pp. 71-81). New York: The Humanities Press.

White, M. (1979). A new effect of pattern on perceived lightness. Perception, 8, 413-416.

Yonas, A., Craton, L.G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. Perception and Psychophysics, 41, 53-59.

Figure Captions

Figure 1. Demonstrations of classical grouping principles. In each case, the elements that are related by the named factor tend to be grouped together perceptually.

Figure 2. Two newly described factors that produce perceptual grouping: common region (Palmer, 1992) and element connectedness (Palmer & Rock, 1994).

Figure 3. Palmer and Rock's (1994) theory of perceptual organization. After initial registration of the image and detection of edges, regions are formed using uniform connectedness and figures are distinguished from grounds. This provides the initial entry level units into a perceptual hierarchy, from which superordinate units are achieved by grouping and subordinate units are achieved by parsing.

Figure 4. A stereogram that demonstrates the influence of binocular depth on grouping. After binocular fusion, the central column of squares groups with those on the right, whereas before fusion they do not. (To fuse the images binocularly, look between at the middle image and try to cross your eyes. Fixating on your finger or a pencil point held above the page may help. Cross your eyes to a degree that produces four distinct versions of the original three. Moving your finger or pencil closer to your eyes or to the page until you see four images should help to achieve this. The central pair of images are binocular and should appear to separate into two distinct depth planes with the central column in the same depth plane as those on the right in both cases, in the near plane in one case and the far plane in the other.)

Figure 5. Depth effects in perceptual grouping. Parts A and B show stimuli used by Rock and Brosgole (1964) to investigate whether proximity grouping is governed by retinal or perceived distances. (See text for further information.)

Figure 6. Stereoscopic depth effects on grouping by common region (A) and element connectedness (B). Once the displays are fused stereoscopically (see the caption for Figure 4 for instructions on how to fuse these images), the circles in A and the squares in B group according to the inducing elements that are in the same depth plane and not with the ones in the different depth plane.

Figure 7. Stimulus displays used by Rock, Nijhawan, Palmer, and Tudor (1992) to show that grouping is influenced by lightness constancy. Part A shows that when the central column of squares is seen behind a translucent strip of plastic or a shadow, it groups with the reflectance matched elements on the left rather than the

luminance matched ones on the right. Part B shows that when the central squares are seen as in front of an opaque strip of paper, they are grouped with the reflectance matched ones on the right rather than the luminance-ratio matched ones on the left.

Figure 8. Stimulus displays used by Palmer, Neff, and Beck (1996) to show that grouping is influenced by visual completion. Part A shows that amodally completed half-circles group with the full circles on the left. Because this effect is confounded by common region, part B shows that moving the occluder slightly further to the side reduces this effect. The results showed that both completed shape and common region have effects in this kind of display.

Figure 9. Stimulus displays used by Palmer and Nelson (2000) to show that grouping is influenced by illusory figures. Part A shows that the central columns group to the right with the other vertical illusory rectangles rather than according to the orientation of the inducing elements. Part B shows that the inducing elements alone are strongly grouped to the left according to the orientation of the ellipses. Part C shows a control condition in which the same inducing elements have been rearranged, in which case no clear grouping is evident.

Figure 10. Stimulus displays used by Palmer and Brooks to show that grouping by proximity and color similarity affects shape constancy. Part A shows that the central element can be seen either as an ellipse in the frontal plane or as a circle slanted in depth. Part B shows that when the central element is closer to and the same color as the square, it tends to be seen as an ellipse in the frontal plane. Part C shows that when the central element is closer to and the same color as the trapezoid, it tends to be seen as a circle slanted in depth.

Figure 11. Stimulus displays used by Palmer and Brooks to show that grouping affects figure-ground perception and relative depth across an edge. When the edge moves (as indicated by the arrow at the top of part A) in the same direction as the elements on one side, observers see the moving side as closer and figural. When the edge does not move (as indicated by the circle at the top of part B) and thus is related by static common fate to the unmoving side, the moving side is seen as farther away and background. Common fate of edge and texture thus determines which side is seen as closer and figural.

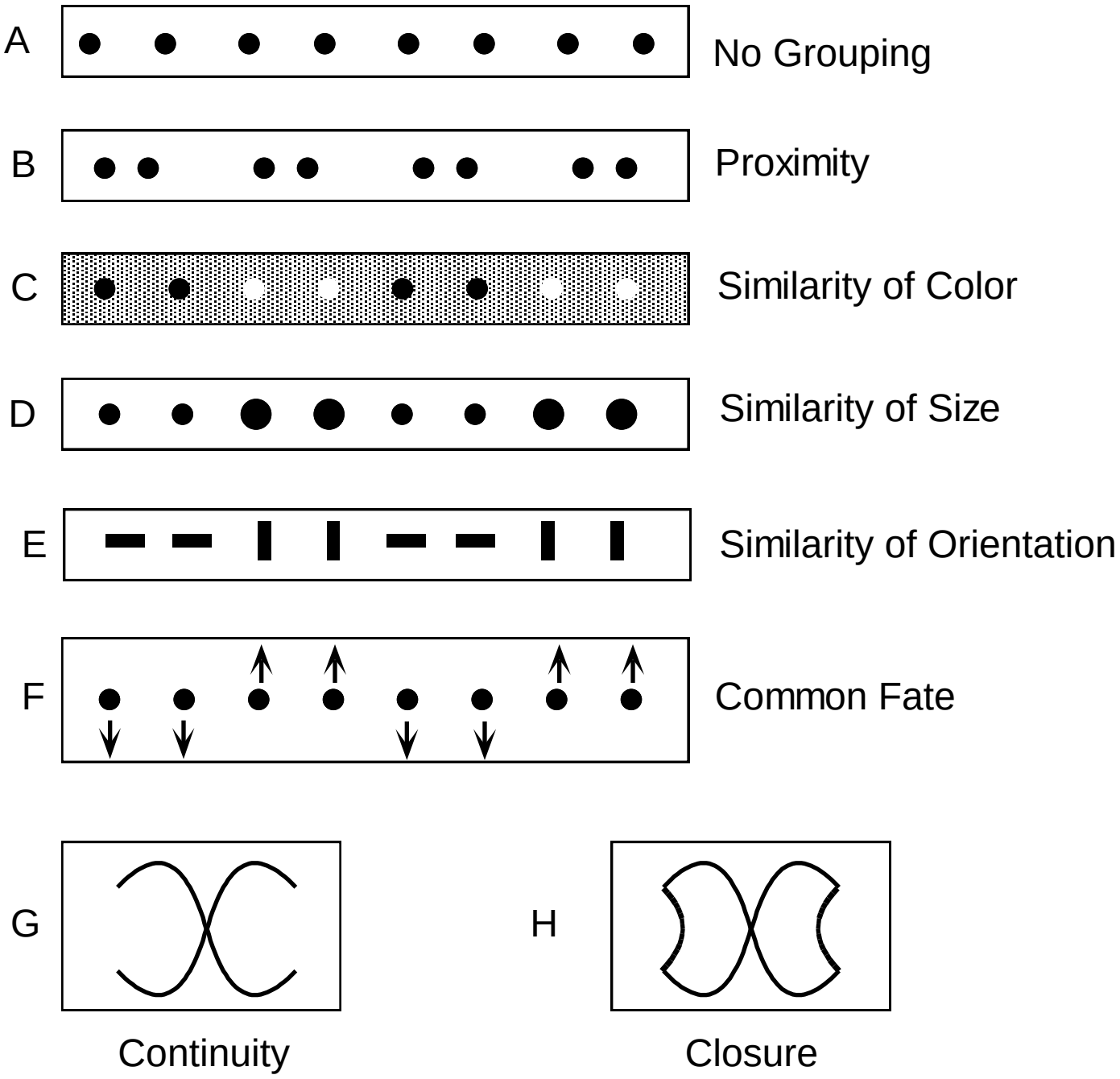

A
No Grouping
B
Proximity
C
Similarity of Color
D
Similarity of Size
E
Similarity of Orientation
F
Common Fate
G
Continuity
H
Closure

Figure 1

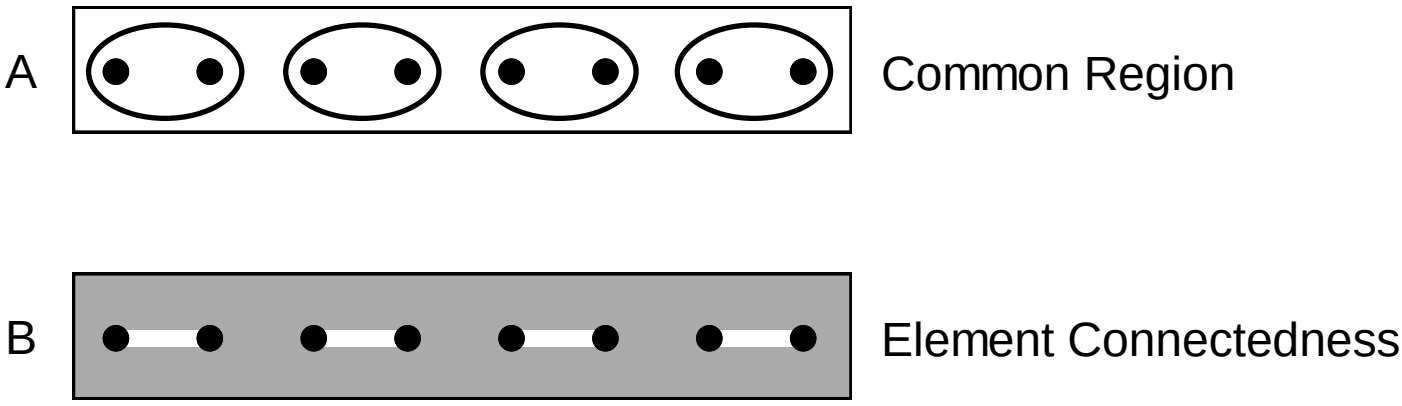


Figure 2

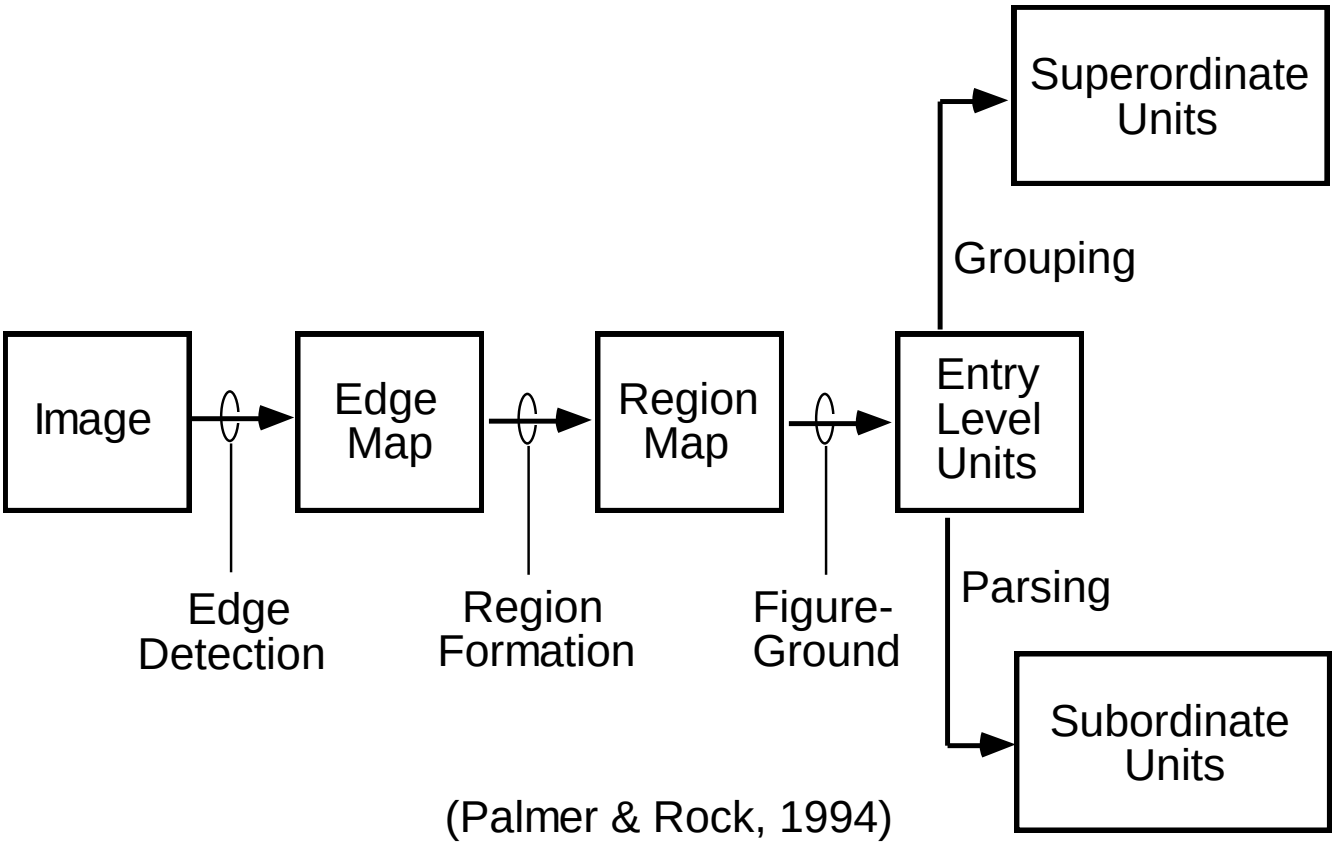

Superordinate Units
Grouping
Image
Edge Map
Region Map
Entry Level Units
Parsing
Edge Detection
Region Formation
Figure-Ground
Subordinate Units
(Palmer & Rock, 1994)

Figure 3

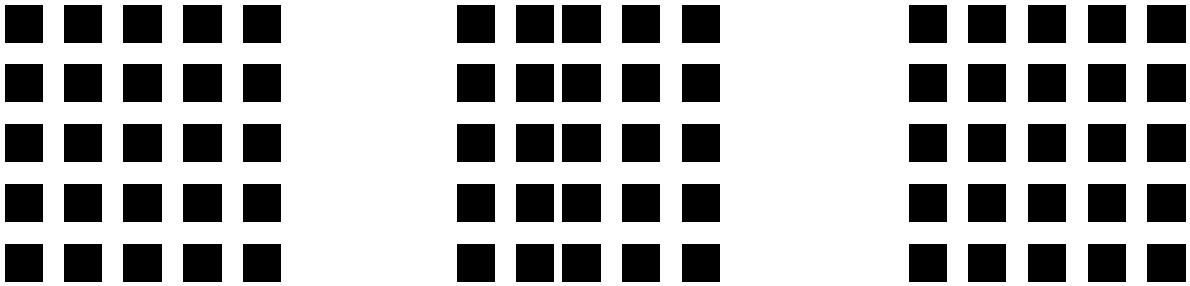
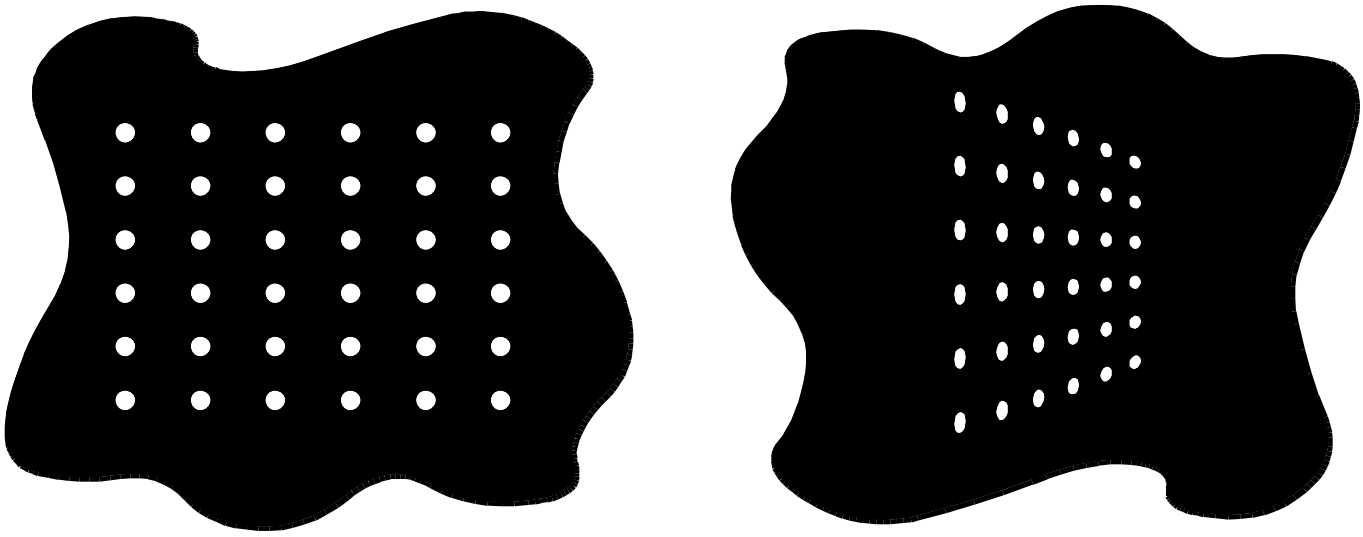


Figure 4



A B

Figure 5

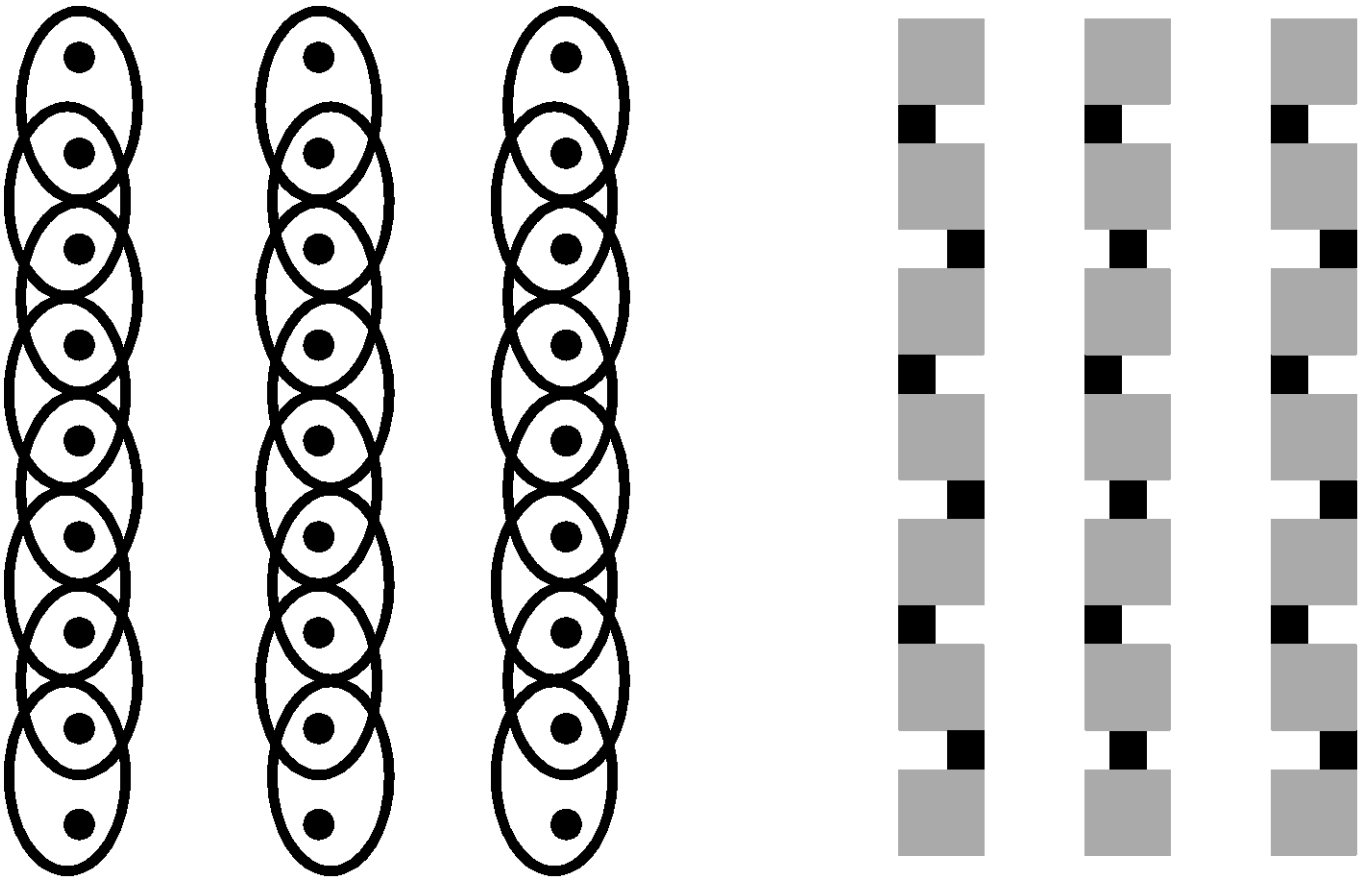


Figure 6

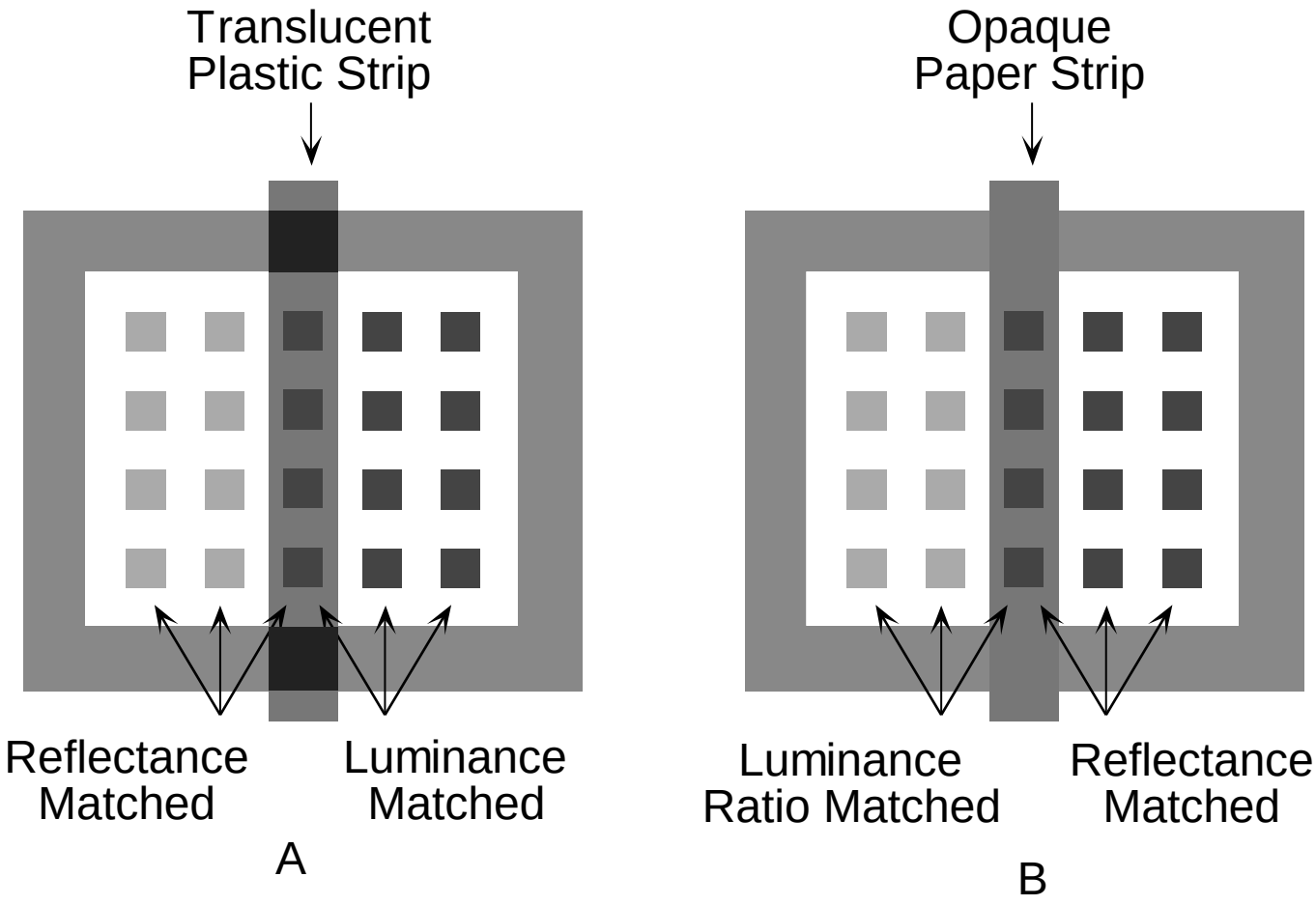
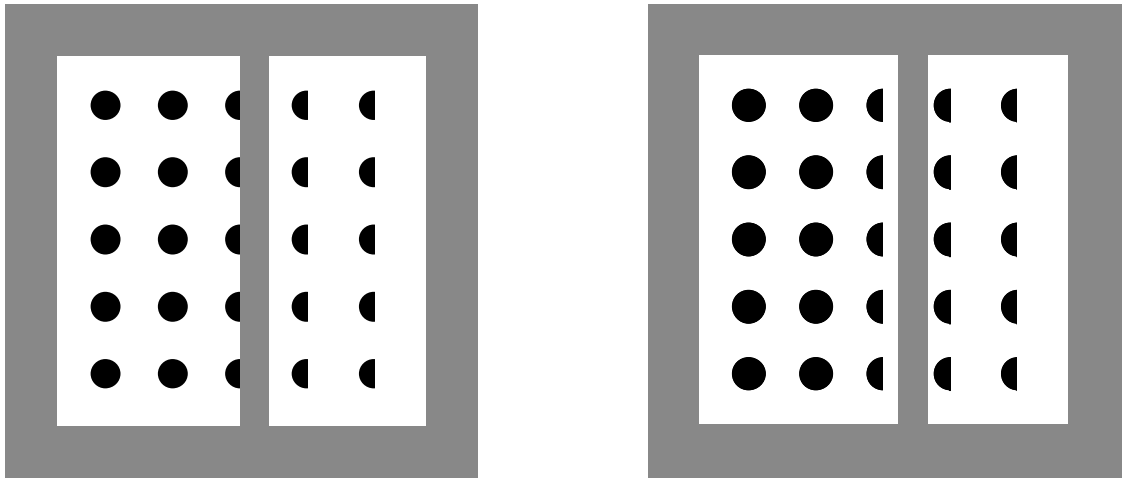

Translucent
Plastic Strip
Reflectance
Matched
Luminance
Matched
A
Opaque
Paper Strip
Luminance
Ratio Matched
Reflectance
Matched
B

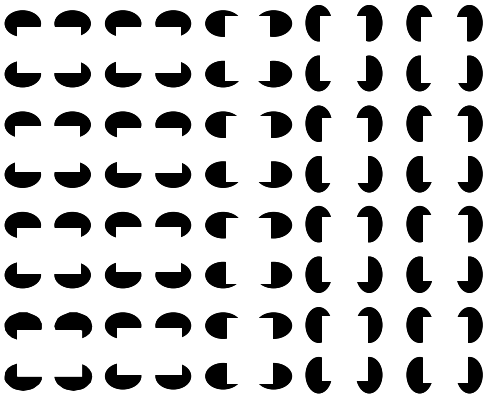
Figure 7



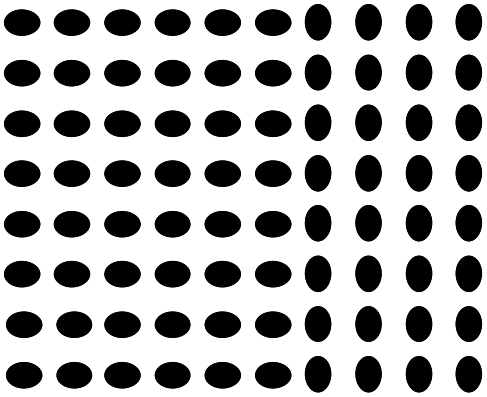
A B

Figure 8

A



B



C

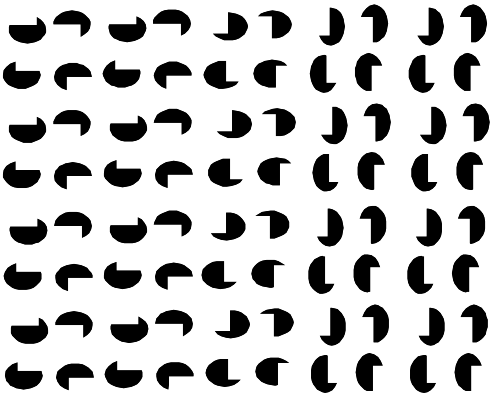


Figure 9

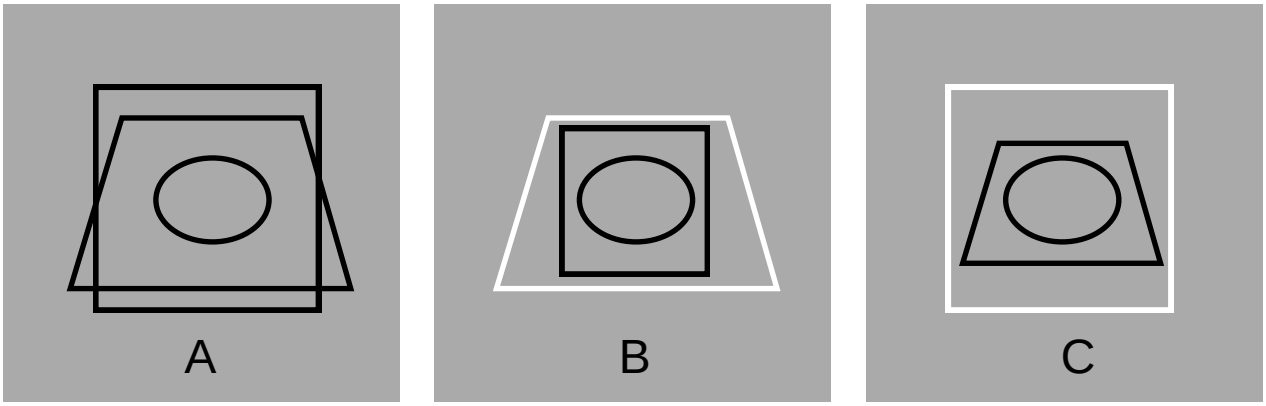


Figure 10

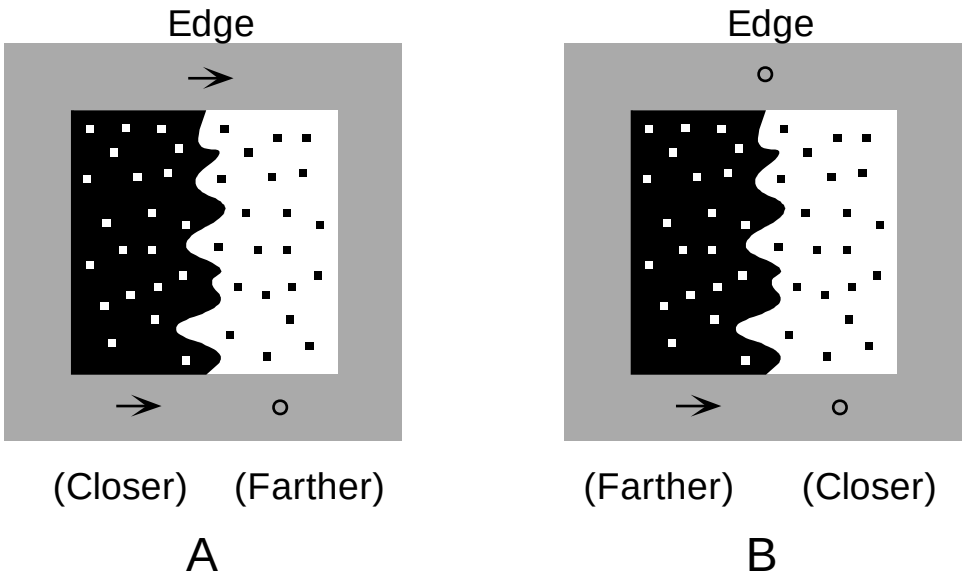


Figure 11