What Is Object Perception?

Visual perception in general, and the visual perception of objects in particular, seems so immediate and effortless that it is difficult to comprehend its complexity. Consider, for example, the non-trivial question of what constitutes an object. Both philosophers and psychologists have occupied themselves with trying to find the necessary and sufficient properties of objects (Hirsch, 1982; Wiggins, 1980). Based on infant research, Elizabeth Spelke and her colleagues have defined objects as solid entities that (a) exhibit spatio-temporal continuity, (b) cohere within their boundaries when they move, and (c) move only when contacted by another object (Spelke, 1990; Spelke, Gurfinkel, & Van de Walle, 1995). On Spelke’s definition, animals and immaterial entities are excluded from the object category, and well they should be, at least for common usage of the term “object.” Bloom (1996) correctly excludes other entities, including puddles, shadows, holes, illusory objects, and parts of objects (e.g., fingers and cup handles). Iflison (1996) excludes pictures of objects because they are two-dimensional (2-D) rather than three-dimensional (3-D), as real objects are.

Distinctions between those entities that count as real objects and those that do not are critical if one is concerned with classifying those entities we judge as there to be real objects. However, most investigators of visual perception use the term “object perception” both more broadly and more narrowly than it is used by the authors discussed above. The term object perception is used more broadly by perception psychologists because it encompasses processes that

- integrate within and segregate between elements in the visual input;
- assign shape and 3-D structure to some of those elements;
- permit recognition of previously-seen shaped entities; and
- determine the manner in which attention is focused on the shaped entities.

Hence, investigators of visual perception typically use the term object perception to apply to both animate and inanimate objects, to pictured (2-D) as well as real (3-D) objects, and even to illusory objects.
Object Perception

The General Case

Objects are bounded by contours. Although the boundaries of physical objects are continuous, the contours extracted by visual processes are likely to be discontinuous, or segmentable. Therefore, as part of the process of segmenting contours from other contours some process must integrate the contour segments (Grossberg & Mingolla, 1985; U 1990). Ulman proposed that contour-segmentation occurs more readily for more "close" contours. According to Ulman, contour salience increases as the orientation size between neighboring contour segments increases.

Ulman’s salience contrast hypothesis implemented the Gestalt psychophysical proposal, the visual system has an inherent tendency to group segments into contours along entailing the smallest change in curvature. This tendency called "good continuation" systems to group the segments of fragmented contours, as in Figure 6.1a, and of continuous contours where they intersect other contours, as in Figure 6.1b. The view was that, by virtue of integration by good continuation, fragmented contours are "virtual contours" by Kanisza, (1975) were as real as real contours. (For recent meta evidence consistent with this hypothesis see Reitwail & Enns, 1995; Hass, Humphreys, & Chen, 1999).

The Gestalt psychologists supposed that good continuation operates over the entire course of perceptual organization. After Hubel and Wiesel (1968) showed that cells in layer V of the visual cortex (V1) were differentially sensitive to stimuli bars of different orientations, it was thought that V1 might be the neural substrate for contour integ ration and segmentation mechanisms. Recent psychophysical, computational, and neuropsychological work has blurred contour integration mechanisms and has confirmed a role for V1 cells. (For an excellent brief historical overview, see Weiskrantz, 1993.)

Field, Hayes, and Hess (1993) used Gabo patches as both contour elements and ground elements (see Figure 6.1d) and examined the conditions under which the cell could be segregated from the background. Field et al. found that, as long as the principal axes were misoriented by less than 60°, observers could accurately segregate the contours, even when the elements differed in phase, and that the element distance was up to seven times the element width. (See also Beck, Rosner, & Levy, 1990.) These effects at a distance demonstrated by Field et al. (1993) were interesting with the classic understanding of the receptive field properties of V1 cells. Here Kapadia, Ito, Gilbert, and Wurtz (1995) later showed that V1 cell response stimuli bars were enhanced substantially when a bar with the same orientation was in nearby, outside its receptive field. Strikingly, the pattern of the effect of orientation on V1 cells was very similar to the pattern obtained in psychophysical studies by Field et al. (1993). Kapadia et al. (1995). The degree to which V1 cell activity was affected by nearby bars decreased at the principal axes of the bars became increasingly misaligned and at the distance between the two bars increased. Together, these psychophysical physiological results support the hypothesis that V1 cells did indeed play a role in contour integration and segmentation.

Segmentation

In this section, we will summarize research and theory concerning processes by which the visual field is segmented, or differentiated, into contents, regions, and groups. We start with contour segregation because it is fundamental for object perception. Grouping processes and region-detecting processes are considered next.

This chapter will cover research and theory on both shape and object perception. The first section of this chapter will cover the processes involved in segmenting the visual field into contours and grouped regions. The second section will cover shape segmentation. Object perception is often considered to be the process of recovering shape from an outline, and different visual architectures for shape perception will be summarized in the third section. Finally, the relationship between attention and object perception will be covered in the fifth section.
Modal and Amodal Contour Completion

So far, the discussion has focused on real and fragmented, or "virtual" contours. In contrast, modal completion occurs when no explicit contour is present, yet an illusory, subjective, contour is perceived. An example is shown in Figure 6.1f, which appears in a white triangle resting on three black circles. The bounding contour of the white triangle is a modal contour, in that it can be perceived. Yet it is a subjective contour because, despite appearances, there are simply no white contours in the display. The black pacman shapes with two straight edges serve as the inducing elements for the subjective contour. When the subjective triangle is seen in Figure 6.1f, the black shapes appear to be circled, completing behind the subjective triangle, due to amodal completion (see below).

Both physiological and behavioral evidence suggests that subjective contours are generated by early visual processes. Physiological investigations have identified cells in V1 and V2 that respond to both real and subjective contours shortly after stimulus onset (Gross & Heeger, 1993; Perrett & von der Heydt, 1989; von der Heydt & Perrett, 1989). Behavioral evidence indicates that the time required to find subjective contour targets does not increase as the number of locations to be searched increases. Such results suggest that subjective contours are generated in parallel across the visual field; focal attention is unnecessary (Davis & Driver, 1994; Gurnsey, Humphrey, & Kapitan, 1992).

In addition, psychophysical investigations demonstrate that, similar to real and virtual contours, modal contours are perceived between contours reversed elements (Prinzmetal, 1998). Thus, although the perceived outcomes are very different, there are clear similarities in the early processes that produce real and subjective contours.

The only straight contours present in Figure 6.1f are those of the three black inducers. Nevertheless, the straight edges are not perceived as belonging to the black shape. Rather, the black shapes are completed as circles lying behind the subjective triangle. This is a case of amodal completion. Amodal completion occurs when contour lines (or edges) are perceived to connect behind occluding surfaces. This implicit contour completion is considered amodal because a connecting edge is not seen; in contrast, modal completion, where contours that are not present in the physical display are too visible. (For review, see Kanizsa, 1987.)

Psychophysical investigations indicate that amodal contours (and the amodal surfaces bounded by those contours) are complete sufficiently early in processing that observers cannot ignore them even when doing so would improve their performance on experimental tasks (He & Nakayama, 1992, 1993).

Figure 6.1. (a) & (b) The number 4, visible in the drawing in (b), is hidden in the drawing in (a). It is reproduced from Perspectives, 21(6) by Hochberg, L. © 1964. Reproduced by permission of Penton Hall, Inc., Upper Saddle River, NJ. (c) Fragmented contours grouped by good continuation. Connected contours grouped by good continuation into continuous contours ABC and EFG. The left and right fields show a sample target and comparison display used by Field et al. (1999) reproduced with permission from Elsevier Science. (d) A subjective contour triangle. (e) A pacman occluded by a black rectangle. According to Kellman and Shipley's (1995) reliability is the edges of the gray rectangle do not complete amodally.
Kellman and Shipley (1991) articulated a relatability rule that predicts when amodal completion will occur. The relatability rule states that amodal contour completion will occur only when smoothly curving contours of intersecting common roots meet at an angle less than 90°. Hence, the black inducement elements complete amodally as circles in Figure 6.1f because the smoothly curving intersections of the convex contours of the inducing elements meet each other. The edges of the gray shape in Figure 6.1g would not complete amodally, however, because the smoothly curving extensions of the inducing elements do not meet. The relatability rule captures local connections as contour continuity.

I end this section by raising the possibility that both modual and amodal completions are generated by the same processes that integrate real and virtual lines; hence, neither may be special cases after all. Consistent with this possibility, Kellman, Yin, and Shipley (1998) showed that these amodal contours that satisfy the relatability rule have some of the same properties as modual contours. Despine and Bonnet (1993) showed that the properties of real and modual contours overlap. Moreover, real and subjective contours function similarly as substrates for certain higher-level processes (Peronnin & Chauvin, 1994b).

**Looking Beyond V1 in Explaining Contour Segregation, Integration, and Completion**

A number of investigators, including Binoculars (1987), Rock, and Walch and Slaugher (1998), showed that familiarity affects modual completion. C. Mouze and Caramagh (1998) demonstrated familiarity effects on visual contour completion. Furthermore, Hoebpberg and Peronnin (1993) and Zemel, Bihlmaier, Mussa, and Ravelet (under review) demonstrated that familiar shapes are more likely than unfamiliar shapes to be completed amodally. And, Seiiter (1994) showed that in addition to local processes, more global processes, such as the symmetry of the completed figure, play a role in early modual completion processes. These results suggest that one must look beyond V1 to gain a full understanding of contour integration and segregation processes.

**Grouping**

In addition to good continuation, the Gestalt psychologists identified a number of factors that increase the likelihood that a set of entities will be grouped together and segregated from other entities. For instance, entities that are similar are likely to be grouped together. Similarity can be determined over any number of dimensions such as shape, color, or size. An example of grouping by similarity can be seen in Figure 6.2a. In addition, elements that are close to one another are likely to group together. See the display in Figure 6.2b. Proximity appears to be determined by the perceived distance separating the elements rather than by the physical distance, which suggests that a Gestalt percept is expected to be formed regardless of whether the elements are close to one another. As well, elements that move together are likely to be grouped together. If the elements in columns 1, 3, and 6 of Figure 6.2b were to move upward, the elements in columns 2 and 4 would move as stationary entities, the moving elements would group together in response to the movement of a common fate and would segregate from the stationary elements. Although common fate was traditionally defined for moving versus stationary elements, or for elements moving in opposite directions, Leonard, Singer, and Fehl (1996) recently found that temporal modulatory brightness outputs can help separate the visual field as well.

**Level at Which Grouping Occurs**

Evidence obtained from a variety of sources suggests that grouping processes are in a task in which observers are asked to categorize a target item appearing at fixed left and right sides (B. A. Erkson and K. E. Erkson, 1974). Distances located in a group with the target (by virtue of similarity or common fate) than when they are not.

![Figure 6.2](image_url)
whether the fundamental units for object perception are global, bounded regions, or whether they are smaller units (see Boofle, 1994; Boofle and Lessenwedge, 1986; Hochberg, 1961; Kimchi, 1998; Peterson & Hochberg, 1983, 1989). A second issue is whether an object constitutes the fundamental, or dominant, segmentation factor, or whether U and the Gestalt grouping and configurational factors constitute a subset of a larger set of factors that cooperate to organize the visual field (Peterson, 1994b, 1999).

Consistent with Peterson's view that UC operates as one cue among many, Han et al. (1999) found that grouping by a cue known to operate quickly—proximity—was accentuated as fast as grouping by UC and was not enhanced when combined with UC. However, they found that grouping by a cue known to operate more slowly—similarity—was accomplished more slowly than grouping by UC and was enhanced when combined with UC. Furthermore, developmental research suggests that UC is not a dominant factor in infants' organization of the visual world (Spelke, 1988). However, consonant with Palmen and Rock's view that UC defines the entry level elements for perception, Wason and Kamin (1999) found that, in adults, other things being equal, attention may select regions defined by UC, even when the selection of larger units would speed task performance (Wason & Kamin, 1999). Additional research is required to determine whether UC has the priority position of defining the first fundamental units for perceptual organization or whether it is simply one of many cases, each of which has different strengths and more contours.

Shape Assignment

The integration and segmentation of contours, groups, and regions is not sufficient for shape perception because not all regions in the visual field are perceived to have shape, some as perceived as shapeless backgrounds. Contours can be described as shared by two regions one lying on each side. Whenever two regions share a contour, two perceptual outcomes are possible. One outcome is that the contour is assigned to one region only; whereas the adjacent region is left contour-less. In this case, the region to which the contour is assigned is the "figure"; the adjacent region is the "ground." By virtue of contour ownership, the figure appears to have a definite shape, whereas the adjacent ground does not, at least as the contour is shared with the figure. When this outcome, termed figure-ground segregation, is perceived the shared contour is seen as an occluding contour, in that it appears to occlude parts of the ground (i.e., the ground appears continuous behind the figure). As example is shown in Figure 6.3a.

A second outcome that can be perceived when two adjacent regions share a contour is that the shared contour is assigned to both regions rather than to just one region. Kennedy 1973, 1974). When this outcome, called figure-figure segregation, is perceived, the shared contour signifies the meeting of two surfaces or objects, both of which appear to be shape by the contour. The two surfaces can appear to lie on the same depth plane, in a flat pattern (Figure 6.3b), or to jut in depth, as in the two surfaces of a cube that meet at a common edge (Figure 6.3c). Examples such as Figures 6.3b & 6.3c demonstrate that one-sided contour assignment is not "exiguous," as some have claimed (Boofle & Driver, 1995). In some situations, such as the one depicted in Figure 6.3c, figure-figure segregation it
clearly the preferred organisation. That may be because the Y- and arrow junctions in figure are themselves early cues to 2-D structure (Enns & Rensink, 1991; Hummel & Biederman, 1992). In other situations, such as the one depicted in Figure 6.3a, how figure-ground segregation seems to be preferred. The likelihood of seeing figure-grounds in figure-ground segregation can be attributed to the cross-region balance of (a) configural factors identified by the Gestalt psychologists and others, (b) contour recognition cues and (c) monocular and binocular depth cues. These factors are discussed next.

Gestalt Configural Factors

The Gestalt psychologists identified a number of factors that affect the likelihood of a region will be attributed a figure status while its adjacent region will be attributed a ground status; these factors are called the Gestalt configural factors. Regions that are (a) small area than their surrounds, (b) symmetric (especially around a vertical axis), (c) contain and/or (d) enclosure are likely to be seen as figures, whereas their adjacent regions are likely to be seen as grounds. Demonstrations devised by Gestalt psychologists in the first half of the twentieth century supported these claims (for reviews, see Hochberg, 1971; Pomer, 1986).

Recently, research replicated and extended the demonstrations of the Gestalt psychologists. For instance, Kanizsa & Gerbino (1976) tested the importance of global cues and found that it is a stronger cue to figure status than symmetry. The important cue is closure as a Gestalt's configural cue was recently confirmed by Kovacs & Julesz (1994) who used a detection task rather than the phenomenological reports favored by the Gestalt psychologists. Kovacs and Julesz (1994) obtained lower detection thresholds for the presented near the center of a region bounded by a closed curve than for targets nearer an equivalent distance from the contour outside the bounded region. Given that these regions tend to be seen as figures, these results replicate and extend previous conductors (Wong & Weisstein, 1983) who reported that detection of high spatial frequency are superior when targets fall on the figure rather than the ground. Similarly, using a court-matching paradigm as an indirect measure of perceived organization, to avoid it of the demand character of phenomenological report, Driver, Badeley, and Rabbitt (1988) recently confirmed the importance of closure of relative area as a segregation cue.

Modern research has revealed new factors that can be added to the list of configural cues. Brown and Weisstein (1988) showed that when different spatial frequency pairs cover two adjacent regions, the region covered with the higher spatial frequency is likely to be seen as the figure. O’Shea, Blackburn and Oto (1994) showed that the region that contrasts most with the background is likely to be seen as the figure. And, Hoffman and Singh (1997) showed that regions with distinctive patterns (defined as large area, configural factors that can be computed locally. For instance, Hoffman and Sing

Do “Configural” Implies “Global”?

Although configural cues are often considered global, holistic cues, recent research in cues that configural factors can be computed locally. For instance, Hoffman and Singh...
(1997) post-distinctiveness is measured locally. Further, Stevens and Brooks (1988) showed that convexity can operate locally. In addition, Han et al. (1999b) and Kimchi (1994) have shown that configural factors are not necessarily mediated by global, object-wide mechanisms. These findings are important because they are consistent with the evidence indicating that figural status need not be assigned to an entirely bounded region. A region can be figure along one portion of its contour and ground along another portion (Hochberg, 1980, 1998; Hoffman & Singh, 1997; Petterson & Hecox, 1996), as illustrated in Figure 6.3d.

Level at Which Configural Cues Operate

Recent research confirms the Gestalt claim that configural factors are computed early in processing. Petterson and Gibson (1994a) showed that symmetry can determine figure-ground assignment in masked exposure as short as 28 ms (but not in 14 ms mask exposure).

An approach used recently to partition perception into early and late-acting processes has been to use brain-damaged individuals who distribute attention preferentially toward the side of space or the side of an object contrasted to the born damage (contralateral spaces or contralateral sides of objects; Heijm & Van Den, 1980; Kindlmann, 1970; see Chapter 7 for more details). Processing that occurs in unmasked contralateral space can be considered "preattentive"—that is, can be considered to occur before the intentional allocation of attention. Driver et al. (1992) showed that the configural cut of smallness of relative area operates effectively to determine figure-ground segregation in contralateral space. Similarly, Driver et al. (1992) found that the configural cut of symmetry affected figure-ground segregation normally in a patient who was unable to consciously attend to the contralateral sides of the figures he saw, and hence, unable to judge accurately whether or not the figures he perceived were symmetric.

None of the factors described above determined which of two adjacent regions will appear to be shaped (i.e., will be seen as the figure) at the time, especially when other competing configural cues are present. Furthermore, the likelihood of assigning shape to one region or the other is affected by contrast-orientation cues and depth cues as well as by configural cues. The perceived segregation depends upon the balance of cues across regions competing for figural status (Petterson, 1994a, 1999).

Quick Access to Memories of Object Structure

It was traditionally assumed that access to memories of objects occurs only after grouping and segregation processes have produced the figures or objects in the visual array (e.g., Koffler, 1929; Nisbett, 1967; Biederman, 1987). Consistent with this assumption, Petterson and her colleagues found evidence that object memories activated by contours can serve as one more shape-assignment cue. Their results were obtained using stimuli (the shapes shown in Figure 6.3e and 6.3f in which adjacent regions sharing a contour differed in the degree to which they resembled known objects. One, "high distinctive," region was a good depiction of an upright known object (such as a face). The other, "low distinctive," region was not (i.e., the black regions in Figures 6.3e and 6.3f). (The density of each region was determined by between-subjects agreement in a pre-test in which observers listed all the known objects each region represented when it was seen as a figure.) Other relevant cues, such as configural cues and the size and binocular depth cues, were sometimes present in their displays, and when these cues favored the interpretation that the low-distinctive region was the figure (e.g., Petterson, Harvey, & Weidenhagen, 1991; Petterson & Gibson, 1995), subjects assigned the configural cut of smallness to the high-distinctive region in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present.

Petterson et al. (1991) compared the likelihood that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e and versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.

Petterson et al. (1991) proposed the hypothesis that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e or versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.

Petterson et al. (1991) proposed the hypothesis that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e and versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.

Petterson et al. (1991) proposed the hypothesis that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e and versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.

Petterson et al. (1991) proposed the hypothesis that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e and versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.

Petterson et al. (1991) proposed the hypothesis that the high-distinctive region was the shaded figure when the displays were upright (as shown in Figures 6.3e and versus inverted (i.e., as seen when you turn the book upside down). Such rotations in the picture plane do not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f. However, rotation in the picture plane does not change the configural or depth cues present in the displays in Figures 6.3e and 6.3f.
In addition, Peterson, Gerhardstein, Mosermeier, and Rapcsak (1998) tested individuals with unilateral brain damage whose attention was biased away from the contralateral contour of the regions of the experimental display. Nevertheless, contour recognition processes seemed to operate normally on the unintended (contralateral) contours, suggesting that contour recognition processes proceed without the benefit of focused attention.

Physiological evidence also supports the claim that shape assignment is accomplished early in visual processing, Zipser, Lamme, and Schiller (1990) measured a response in V2 cells 80-100 ms after stimulus onset that was evident when line, shaded, figures, but not grounds, fell on the cells' receptive fields. If this differential activation is indeed indicative that shape assignment or figure-ground segregation has been accomplished, these data support the view that these processes occur early in visual processing. Because the inferior temporal cortex, located downstream from V2, and important for object recognition, can be activated 60 ms after stimulus onset, however, these data are also consistent with the proposal that object memories can affect figure-ground segregation.

Just as for the other segregation-relevant cues, the likelihood that the region providing a good fit to object memories will be seen as figure depends on the balance of other cues. In other words, the cue originating in quick access to object memories does not always dominate the configural and depth cues (Povinelli & Gibson, 1992, 1994). Indeed, just as most of the other configural cues or depth cues is a necessary component of the segregation process, neither is a good fit to an object memory. Therefore, segregation can proceed without subcortical contributions from object memories for novel objects, as it can proceed without contributions from the configural cue of symmetry for asymmetric objects (or without congruency for concave objects). However, when known objects are present, the segregation process can benefit from prior experience, as it can benefit from convexity when convex objects are present.

Depth Cues

Many depth cues, including binocular disparity (stereopsis), contour, motion parallax, texture, and shading, affect the likelihood that shape will be assigned to the region lying on one or the other side of a contour. Sustained displays. Nevertheless, contour recognition processes seem to operate normally on the unintended (contralateral) contours, suggesting that contour recognition processes proceed without the benefit of focused attention.

Physiological evidence also supports the claim that shape assignment is accomplished early in visual processing, Zipser, Lamme, and Schiller (1990) measured a response in V2 cells 80-100 ms after stimulus onset that was evident when line, shaded, figures, but not grounds, fell on the cells' receptive fields. If this differential activation is indeed indicative that shape assignment or figure-ground segregation has been accomplished, these data support the view that these processes occur early in visual processing. Because the inferior temporal cortex, located downstream from V2, and important for object recognition, can be activated 60 ms after stimulus onset, however, these data are also consistent with the proposal that object memories can affect figure-ground segregation.

Just as for the other segregation-relevant cues, the likelihood that the region providing a good fit to object memories will be seen as figure depends on the balance of other cues. In other words, the cue originating in quick access to object memories does not always dominate the configural and depth cues (Povinelli & Gibson, 1992, 1994). Indeed, just as most of the other configural cues or depth cues is a necessary component of the segregation process, neither is a good fit to an object memory. Therefore, segregation can proceed without subcortical contributions from object memories for novel objects, as it can proceed without contributions from the configural cue of symmetry for asymmetric objects (or without congruency for concave objects). However, when known objects are present, the segregation process can benefit from prior experience, as it can benefit from convexity when convex objects are present.

Depth Cues

Many depth cues, including binocular disparity (stereopsis), contour, motion parallax, texture, and shading, affect the likelihood that shape will be assigned to the region lying on one or the other side of a contour. Sustained displays. Nevertheless, contour recognition processes seem to operate normally on the unintended (contralateral) contours, suggesting that contour recognition processes proceed without the benefit of focused attention.

Physiological evidence also supports the claim that shape assignment is accomplished early in visual processing, Zipser, Lamme, and Schiller (1990) measured a response in V2 cells 80-100 ms after stimulus onset that was evident when line, shaded, figures, but not grounds, fell on the cells' receptive fields. If this differential activation is indeed indicative that shape assignment or figure-ground segregation has been accomplished, these data support the view that these processes occur early in visual processing. Because the inferior temporal cortex, located downstream from V2, and important for object recognition, can be activated 60 ms after stimulus onset, however, these data are also consistent with the proposal that object memories can affect figure-ground segregation.

Just as for the other segregation-relevant cues, the likelihood that the region providing a good fit to object memories will be seen as figure depends on the balance of other cues. In other words, the cue originating in quick access to object memories does not always dominate the configural and depth cues (Povinelli & Gibson, 1992, 1994). Indeed, just as most of the other configural cues or depth cues is a necessary component of the segregation process, neither is a good fit to an object memory. Therefore, segregation can proceed without subcortical contributions from object memories for novel objects, as it can proceed without contributions from the configural cue of symmetry for asymmetric objects (or without congruency for concave objects). However, when known objects are present, the segregation process can benefit from prior experience, as it can benefit from convexity when convex objects are present.

Theories of Object Recognition

An adequate theory of object recognition must account for:

- the accuracy of object recognition over changes in object size, location, and orientation (and it would be preferable if this account did not pose a different memory record each view of every object ever seen);
- the means by which the spatial relationships between the parts or features of an object are represented; and
- the attributes of both basic-level and subordinate-level recognizers (e.g., recognizability of a fish as both a bird and as a specific kind of bird).

Current competing object recognition theories differ in their approach to each of these attributes (see Biederman, 1987, 1995; Tarr, 1995; Tarr & Bulthoff, 1998).

Recognition by Components Theory

According to the Recognition by Components (RBC) theory, proposed by Biederman (1987), objects are parsed into parts at precise portions of their bounding contours, and these parts are represented in memory by a set of abstract 3-D components, called "primitives." In RBC theory, proposed, other theorists have stressed the importance of recognition of both major regions of the bounding contour of objects (e.g., Hoffman & Richards, 1985; Noda, 1982; Marr & Nishihara, 1978) and 3-D representational components (i.e., cylinders, boxs,
1981; Marr, 1982; Marr & Nishihara, 1978). Biederman expanded the set of components from cylinders to generalized forms (e.g., convex-section swept out in depth along an axis). Further, Biederman (1979a) showed that a false set of 3-D geons (N = 24) can be defined by combining a small set of binary or trinary constraints that can easily be extracted from two-dimensional images. Thus, in BGC, a representation of an object’s 5-F structure was derived from constraints extracted from a single 2-D view. The constraints specify the shape of the cross-section of the geon, the shape of the axis of the geon, and changes at the tip of the cross-section. The constraints are then transferred to a scene model of the object (Figure 6.4). Examples include the following: for edges, whether they are straight or curved, parallel or non-parallel, converging or diverging; and for cross-sections, whether they shrink, expand, or remain constant in size as they travel along the geon axis. Sample geons and scene objects constructed from them are shown in Figure 6.4a.

The constraints from which the geons are constructed are viewpoint-invariant properties or “non-accidental properties,” in that they are unlikely to occur in the image as an accident of viewing position (Lowe, 1965; Wilkins & Tremblay, 1983). For instance, under most viewing conditions, except for accidental views, curved lines do not look straight, nor do converging lines appear parallel. Biederman and his colleagues (Biederman, 1987, 1995; Biederman & Gerhardt, 1993, 1999) argued that geon extraction is viewpoint invariant because the constraints that specify the geons are viewpoint invariant. The production of the constraints that object recognition should be viewpoint invariant follows, provided that the same geons and geon relations are extracted from the image, in different views. That, according to RBC theory, only a small number of views of each object are needed to be represented in memory.

BGC specified the spatial relations between the geons comprising an object are specified in terms of categorical relationships such as “top-of,” “below,” or “face-to,” rather than in metric terms (Biederman, 1987; Hummel & Biederman, 1992). It is known that object recognition fails when the parts are rearranged (Cave & Kosslyn, 1983; Tarr, et al., 1991). Nevertheless, prior to RBC, little consideration had been given to the question of how the spatial relationships between the parts of an object are coded.

Evidence: Pro and Con

As all good theories should be, BGC makes testable predictions and, consequently, is falsifiable. After the publication of Biederman’s (1987) article, research and theorizing on object recognition flourished, and continues to flourish today. From the 1980s to today, Petrich investigated questions such as whether the (a) front-end features and constraints are as important as claimed by BGC, (b) object recognition is viewpoint invariant, (c) RBC can account for both subsymbolic and basic-level recognition, and (d) RBC’s proposals concerning the coding of spatial relationships. The fact that many objects can be recognized from their bounding contours alone indicates that bounding contours are highly invariant for object recognition.

Figure 6.4. (a) Sample geons and objects constructed from them. Reproduced from Biederman (1995), with permission of MIT Press. (b) Perceptual objects (top row) and schematic objects (bottom row) used in tests of Multiple Views Theory (reproduced from Lotz et al., 1996, with permission from Elsevier Science). (c) Examples of the data that caused cells at various levels in the ventral processing stream.
Another criticism is that the object representations in Multiple Views Theory are too much like two-dimensional templates. It is well known that template-like representations leave perception susceptible to distortion by slight changes in any object features (Newell, 1967), regardless of whether they lie on, or internal to, the object's bounding contours. Yet human perception is not nearly so robust to such changes. It was just this robustness of object perception that led early theorists to propose that object contours were viewpoint-independent, size-independent, location-independent, etc. (i.e., in Marr's 1982 terminology, they were "object-centered" representations). It is feared that Multiple Views Theory might require an unreasonably large number of representations for each object. Moreover, it is not clear how different views of objects are determined to be the same object, rather than similar but different objects. Another criticism is that in Multiple Views Theory there is no provision, save for that implied in template matching, for representing the spatial relations between parts of objects. This is a drawback, given the importance of the spatial relations between features, and the behavioral distinctions between the spatial relations between parts of objects and between parts of objects (see above).

Three criticisms point to research that must be done to elaborate Multiple Views Theory. Current attempts to resolve these problems include (a) using the bounding contour of the object to inter-relate different views, and (b) exploring the feasibility of viewpoint coding for spatial relations between features (for a summary of recent research see Tarr and Bulthoff, 1998). Recall that both bounding contours and categorically coded spatial relations play important roles in RBC. It may turn out that a complete theory of object recognition must incorporate principles of both RBC and Multiple Views Theory (Suzuki, Petecon, Mowen, & Behrmann, under review; Tarr & Bulthoff, 1998). There also remains the possibility that, in addition to orientation-dependent representations (such as those identified by Logothetis et al., 1995), there exist object-centered representations (i.e., representations that permit orientation-independent object recognition) (Corballis, 1986; Solms, Turnbull, Kaplan, Solms, & Milner, P., 1998; Turnbull & Mecarsky, 1996).

Open issues

In addition to the issues discussed at the end of the preceding section, two other issues must be considered in order to understand object recognition. The first concerns the role of local features in object recognition. For the part, theories assume that object recognition is a global or holistic process. However, both behavioral and computational evidence (Moser, Zemel, Buhmann, & Williams, 1992; Petreman & Heerey, 1990; Ullman, 1991) suggests that object recognition is mediated by local cues. Those local cues that are necessary and sufficient for object recognition have yet to be determined. Furthermore, exposing median evidence suggests that the local component of representation is affected by the experience (Lou & Murphy, 1997; Mount & et al., 1992; Schlos, Goldstone, & Tullibrew, 1998; Zemel et al., under review). Future research is exploring the nature of the local cues mediating object recognition and the degree to which they are learned, and the interactions between local and global cues will be important for object recognition theory.

The second open issue concerns the nature of the representational primitives. In both Multiple Views Theory and RBC Theory, there is a clear resemblance between the object and the representational components. Indeed, it is easier to think about the components of
Models of the Relationship Between Segmentation, Shape Assignment, and Object Recognition

In order to understand how object perception proceeds, it is important to understand how the components processes of integration and segmentation, shape assignment, and object recognition are ordered. Which precede the others? Which serve as substrates for others? In what follows, I first discuss traditional hierarchical models. Next, I summarize a parallel model my colleagues and I have proposed. Finally, I point out the open questions that must be addressed to adjudicate between these models.

Hierarchical Models

The Gestalt psychologists proposed that segmentation, shape assignment and recognition were ordered serially and hierarchically, with grouping and segmentation completed first and forming the substrate for shape assignment, and shaped regions in turn providing the substrate for, and necessarily being determined prior to, access to object memory. (For some evidence consistent with this proposal that segmentation is completed before shape assignment see Sekuler & Palmer, 1992; for contradistinctive evidence see Brunn, Beramini, & Demantis, 1997; Peterson et al., 1991; Kellman et al., 1998.)

An influential model of vision proposed by David Marr (1982) was also serial and hierarchical. Unlike the Gestalt psychologists, Marr concentrated on the traditional depth cues at the expense of the configurational cues, arguing that the sphere of influence of the latter was restricted to 2-D displays, which represent only a small subset of the conditions under which the visual system operates. (The current dissertation between those who study the perception of shape based upon configurational cues versus depth cues can be traced to Marr's position.) According to Marr, visual input proceeds through a number of stages, illustrated in Figure 6.5(a). The first stage is a process which, in which edges are made explicit. The second stage entails the construction of the 2-D sketch, in which surface and viewer-relative orientations emerge. The third stage is the construction of the 3-D model, and as a final step, the 3-D model is matched to 3-D object models stored in memory. In Marr's theory, there is a clear sequence from edge extraction through 3-D shape assignment before object memories are accessed. Marr's theory was proposed before either the BBC Theorems or the Multiple View Theory of object recognition. Indeed, the BBC Theory owes much to Marr and Nishihara's (1978) work.

More recent interactive hierarchical models of the relationship among segmentation, shape assignment, and object recognition allow feedback from higher levels to influence processing at lower levels. However, these models maintain a hierarchical structure in that lower-level processes must at least be initiated before higher-level processes are initiated, illustrated in Figure 6.5(b). McClelland (1979, 1985; McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1982, 1986; Vecera, & O'Reilly, 1998).

In hierarchical views of perceptual organization, configurational and depth cues are construed as lower-level, or bottom-up, cues - cues that do not require access to higher-level mnemonic representations, and shape assignment based upon these cues is considered a lower-level process than object recognition. Consequently, according to these views, an object memory cannot be accessed before shape assignment and perception is at least partially completed. On the basis of the evidence that contour recognition proceeds infinitesimal shape assignment, my colleagues and I proposed a parallel model, described next.

A Parallel Model

Recall that investigations with figure-ground displays indicated that object memories access quickly in the course of processing affect the shape assignment. The cues arising from the activated object memories did not dominate the other configurational cues or depth cues. Neither the configurational and depth cues constrain access to object memories. Rather, active object memories seem to serve as a more cue among the many cues that contribute to the likelihood that a region will be seen as a shaped figure rather than a shapeless ground. Critically, object memories affect shape assignment only when they were accessed quickly. Influence from object memories could be removed either by inverting the visual (as in delaying the access to object memories), or by using contours detected later than those in processing (e.g., random-dot stereo edges versus luminance edges).

On the basis of this evidence, my colleagues and I proposed that, as soon as contours are segmented in the visual input, quick access to object memories via contours initiate The model is a parallel model because object memories are accessed via contour-based mechanisms at the same time that other processes assess the Gestalt configurational cues and depth cues, and all of these processes interact to affect shape assignment. See Figure 6.5(b). We do not suppose that time course of all of these processes is the same. We suppose that shape assignment based upon configurational cues and/or depth cues does not precede access to object memories, either partially or wholly, as would be assumed on hierarchical models (Peterson, 1994a, 1999; Peterson & Gibson, 1994a, 1994b).
A Continuing Debate

Hierarchical interactive models have been adapted to account for the evidence indicating that shape assignment is affected by access to object memories (e.g., Vecera & O'Reilly 1998, 2000). In Vecera and O'Reilly's (2000) model, lower-level processes do not constrain the operation of higher-order recognition processes and effects of recognition processes on the processing at lower levels are evident at the earliest times. It will be difficult to distinguish this version of an interactive hierarchical model from a parallel model (e.g., Peterson, 1999; Peterson et al., 2000).

Many theorists prefer hierarchical models (serial vs. interactive) to parallel models because they believe that hierarchical models are better able than parallel models to account for the perception of novel or unrecognized objects (e.g., Marr, 1982; Warrington, 1982). However, this belief is based on the incorrect assumptions that (a) in parallel accounts inputs from higher-level object memories are necessary for shape assignment and perception, and (b) high-level influences must dominate low-level factors. Neither of these assumptions is necessary and neither is held in the parallel models proposed by Peterson (1999; Peterson et al., 2000; also see above).

A hierarchical model implicitly underlies the notion of object files proposed by Kahneman and Treisman (1984) to account for preserved object continuity over changes in perceived identity, shape, color, or location. They proposed that temporary representations of objects are created at an intermediate hierarchical level, before object identity is established. These "object files" mediate object continuity over changes in object features such as location, color, and shape, provided that the changes are not too extreme. Priming experiments support the existence of object files (e.g., Kahneman, Treisman, & Gibbs, 1992; Treisman, Kahneman, & Burkull, 1983), but experimental evidence suggests that object files may code some aspects of object identity as well (Gooluck & Irwin, 1996; Hender & Attes, 1994). Thus, object files must be understood within a hierarchical model in which one assumes that spatio-temporal continuity is less likely to be maintained over changes in object identity than over changes in other object features. Although this prediction might be generated by a serial hierarchical model, it is not necessary on a parallel account.

Another reason underlying a preference for both serial and interactive hierarchical approaches to perceptual organization is that brain structures are thought to be arranged hierarchically. For instance, occipital cortex is initially activated via cortical connections earlier in time than temporal and parietal cortices, which in turn, are activated (indeed, Vecera and O'Reilly 2000) argue that their hierarchical interactive model of figure-ground assignment is an architectural model rather than a processing model. It is tempting to associate functional stages such as those proposed by MacKay (1982) with these sequentially activated brain regions. However, it must be remembered that there are massive feedback connections between brain regions as well as feed-forward connections (Feldman & Van Essen, 1991; Zeki, 1993). These feedback connections from brain regions activated later in time via cortical connections can alter the activity in brain regions activated earlier in time. When these feedback connections are taken into consideration, it becomes very difficult to pinpoint the stage at which various aspects of perceptual organization are accomplished (Bridgick, 1992; Peterson, 1999).
Attention and Object Perception

In this section, we consider research concerning the relationship between object perception and attention. We begin by considering whether attention makes something an object. Imagine trying to orient in order to count them, for example. Wolfe and Bennett (1997) define an object as "a ncasurable thing as distinct from a collection of nterable things and as distinct from unnumerable stuff." We can certainly count objects, but we can also count other things like the spaces between the words on this line of text. For example, enumerability does not make something an object. Similarly, one can count spaces as well as objects, but attending does not necessarily make a space an object (Rubin, 1915/1958; Peterson & Gerardinien, under review; Peterson & Gerardinien, Menemets, & Raynor, 1998; Peterson & Gibson, 1994). Thus, neither attention nor enumerability is sufficient for object perception. The related question of whether attention, or intention, on the viewer's part is necessary for object perception is currently being explored, as a consequence of pioneering work by Mack and Rock (1998). This topic is covered next. Then, in the following section, we consider the evidence indicating that there exists an object-based form of attention that is distinct from spatial attention.

It's Attention Necessary for Object Perception?

Inattentional Blunders

Mack and Rock (1998; Mack, Tang, Tuma, Kahn, & Rock, 1992; Rock, Linnet, Grant, & Mack, 1992) found that a large percentage of observers were effectively blind to the unattended object of an object when they were performing a difficult discrimination task. The discrimination task entailed judging which of the two arms of a cross was longer, when the difference between the arms was quite small. The cross was exposed briefly (e.g., 200 ms) and followed by a masking stimulus. On the third trial on which observers performed this task, a simple geometric object was presented in one of the four quadrants sketched by the cross at the same time that the cross was presented. When questioned shortly after this critical trial, many observers reported that they had not seen any unusual object. Some observers did report that something unusual had happened on the critical trial, but they were unable to report the simple geometric shape of the object that had been shown (e.g., a square or a triangle). Mack and Rock called this phenomenon "inattentional blindness." They argued that if the observer's attention or intention is not directed to perceiving an object, then object perception does not occur. Note that inattentional blindness is necessarily inferred from performance on a memory task rather than from performance on an online perception task. (Presumably, if observers knew they would have to occasionally detect objects, their perceptual intentions would change to accommodate this object detection task.) The phenomenon of inattentional blindness raises the possibility that one may need attention or intention to perceive objects consciously. This in turn raises a question about terminology. Should the term "perception" (and the term "object perception" in particular) be reserved only for conditions in which observers can report being consciously aware of what they perceived? I argue that it should not. Research summarized in this chapter indicates that many of the components processes involved in object perception can be computed outside the observer's attentional focus, outside of awareness. Other evidence comes from work by C. M. Moore and Egger (1997), who adapted the Mack and Rock paradigm and presented convincing evidence that grouping occurs without attention or intention. Regardless of whether or not the term "object perception" is ultimately reserved for conscious object perception, the important question of how attention or intention contributes to conscious object perception remains.

Stimulus Selection

A related debate in the search literature concerns whether or not it is possible for a stimulus to draw attention automatically if an observer is not intending to search for that stimulus in the first place. Pop-out effects have often been taken as evidence that unusual stimulus features or abrupt stimulus onsets can attract attention (e.g., Triesman, 1988; Yamaguchi, 1993, 1996). "Top-out" occurs when a single target stimulus differing in some basic feature from other "distractors" is detected quickly, and target detection latency does not increase as the number of distractors increases (e.g., the time to detect a red dot among green dots does not increase appreciably as the number of green dots increases). The sp-k detection responses were originally attributed to "stimulus selection" because the automatic attraction of attention to the distinct stimulus feature in the display. But Mack, Rock, and others pointed out that, in experiments demonstrating pop-out effects, observers are typically given advance information. This about the identity of the target feature or stimulus. Therefore, pop-out effects cannot serve as evidence that targets automatically attract attention in virtue of being different from the other display items. In experiments in which attention and the set were carefully controlled, Folk, Remington, and Johnston (1992) and Gibson and Keber (1998) failed to find evidence for stimulus selection, consistent with the view
that task determines what observers perceive. These results are consistent with the hypothesis that attention/intention is necessary for conscious perception, although it must be remembered that no "common everyday" task as single task veridical in the experiment--that expertise provides a universally accepted standard.

Binding

Attention may be required to bind together the various properties of an object, such as color, form, and movement (Treisman, 1988; Treisman & Gelade, 1980) as well. Treisman and her colleagues argued that, without attention, such features can be combined incorrectly, and illusionary conjunctions can occur (e.g., illusionary conjunctions of color and form, or form and motion). Poonam (1981, 1995) showed that, when grouped displays are not attended, illusory conjunctions are more likely to occur within grouped entities than across grouped entities. (Note that Poonam's results, like Moore and Egbert's, suggest that grouping itself can occur without attention.) Wolfe and Bennett (1997) recently demonstrated that attention is necessary to contain the features of an object, at least for conscious report. They argue that, prior to the allocation of attention, objects are nothing more than loose collections of basic features organized on the basis of spatio-temporal properties (i.e., Kanizsa & Treisman's object file). However, it is important to remember that incompatibility to conscious report does not necessarily imply that perceptual organization has not occurred. There is some evidence that binding has occurred, even when it cannot be measured via conscious reports (Robinson, 1998; Wojciulik & Kanwisher, 1998).

Experiments such as these indicate that attention must be considered if we are to understand object perception, but does the relationship between perception and conscious report. I turn next to consider the evidence indicating that a specialized form of object-based attention exists.

Object-Based Attention

It has long been known that attention can be allocated to locations in space that are different from the location where the eyes are directed (Posner, 1980). More recently, it has been shown that attention can be allocated to objects independently of the space they occupy, for instance, Driver & Halligan, 1990; Gilmore & Egbert, 1995; Treisman, Kahanov & Buxbaum, 1980). Evidence that attention can be "object-based" and not just "space-based" takes various forms. A form of evidence for "object-based" attention is that it takes longer to move attention a given distance between two objects than the same distance within the same object (Egly, Driver, & Rafal, 1994; Egly, Rafal, Driver, & Starmovels, 1994). Another manipulation of object-based attention is that observers require less time to report the same measure of a simple object than about twice as much for different objects. This second effect is obtained even when the two objects overlap each other and occupy essentially the same location (Duncan, 1984; Golomb, 1998), and even when portions of the single object are occluded by another object (Oehlmann, Zemel, & Moore, 1998). A third demonstration of object-based attention entails moving objects. When a visual object moves to a new location, attention moves with the object, rather than in addition to

staying in the cued location (Kahneman, Treisman, & Gibbs, 1992; Tipper, Arthas, Driver, 1990; Tipper, Driver, & Weaver, 1991). Thus, attention seems to spread readily within an attended object (i.e., between two objects, 51) to encompass the visible features of an attended object, and to move with an object.

In summary, the study of object perception and the study of attention are two areas that are integrated. Objects may be perceived consciously only if they are attended by the observer's attention. Once objects are perceived consciously, they form a unique unit for the spatial and attentional attention that is distinct from a purely spatial stimulus. Questions regarding the relationship between space and objects have been raised throughout this chapter and will continue to be raised in the future. Questions regarding the relationship between consciousness and attention are important in all (see Chapters 7 and 10) for all levels of these processes.

Suggested Readings


Additional Topics

Colour and Surface Detail

To what extent do shape and surface detail influence object recognition? (Laderman & Jo, Price & Hucsh, 1999; J. W. Tania & Freehend, 1999)

Are Indecid Objects?

For discussion of holes, see Cusati & Vare (1993); Bloom (1996); Bloom & Gifford (Hochberg, 1996).

Tactile and Auditory Object Perception

What general principles can we draw from earlier studies in which object perception occurs, for example, are we able to use the same principles to understand object perception in a new study (Dubin, 1996; Freehend, 1999; Freehend, 1999).

References

The perception of linearity (pp. 462–70). New York: Springer-Verlag.