CHAPTER 11

Grouping

11.1 HUMAN VISION: GROUPING AND GESTALT

Remarkably, the human vision system appears to have strong opinions about when data points belong together, which is why it is easy to tell that the points in Figure 34.2 mostly lie on a line. In turn, context affects how things are perceived (e.g., see the illusion of Figure 11.1). This observation led the Gestalt school of psychologists to reject the study of responses to stimuli and to emphasize grouping as the key to understanding visual perception. To them, grouping meant the tendency of the visual system to assemble some components of a picture together and to perceive them together (this supplies a rather rough meaning to the word context used above). Grouping, for example, is what causes the Müller–Lyer illusion of Figure 11.1 — the vision system assembles the components of the two arrows, and the horizontal lines look different from one another because they are peceived as components of a whole, rather than as lines. Furthermore, many grouping effects can't be disrupted by cognitive input; for example, you can't make the lines in Figure 11.1 look equal in length by deciding not to group the arrows.

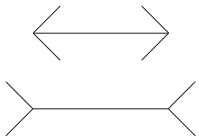

FIGURE 11.1: *The famous Muller–Lyer illusion; the horizontal lines are in fact the same length, although that belonging to the upper figure looks longer. Clearly, this effect arises from some property of the relationships that form the whole (the* gestaltqualität*), rather than from properties of each separate segment.*

A common experience of segmentation is the way that an image can resolve itself into a **figure** — typically, the significant, important object — and a **ground** — the background on which the figure lies. However, as Figure 11.2 illustrates, what is figure and what is ground can be profoundly ambiguous, meaning that a richer theory is required.

The Gestalt school used the notion of a *gestalt* — a whole or a group — and of its **gestaltqualität** — the set of internal relationships that makes it a whole (e.g., Figure 11.1) as central components in their ideas. Their work was characterized by attempts to write down a series of rules by which image elements would be associated together and interpreted as a group. There were also attempts to construct algorithms, which are of purely historical interest (see **?** for an introductory account that places their work in a broad context).

The Gestalt psychologists identified a series of factors, which they felt predis-

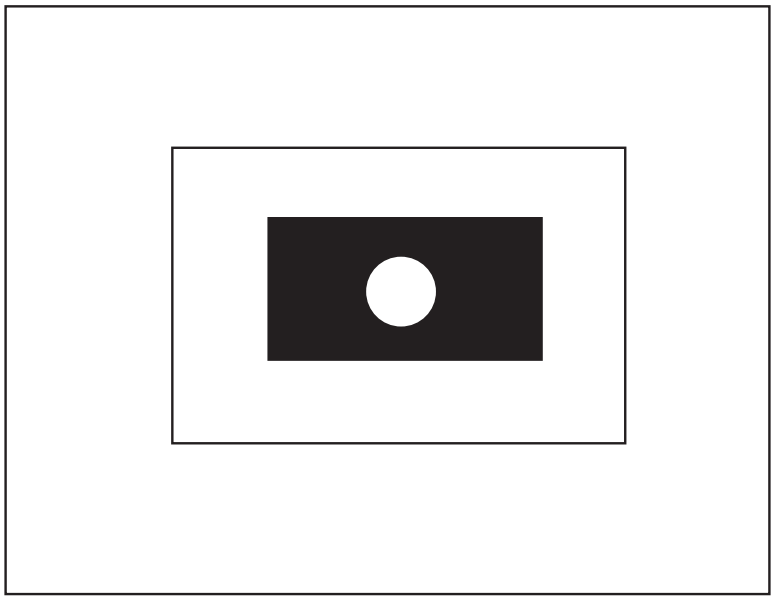

FIGURE 11.2: *One view of segmentation is that it determines which component of the image forms the figure and which the ground. The figure illustrates one form of ambiguity that results from this view; the white circle can be seen as figure on the black rectangular ground, or as ground where the figure is a black rectangle with a circular hole in it — the ground is then a white square.*

posed a set of elements to be grouped. These factors are important because it is quite clear that the human vision system uses them in some way. Furthermore, it is reasonable to expect that they represent a set of preferences about when tokens belong together that lead to a useful intermediate representation.

There are a variety of factors, some of which postdate the main Gestalt movement:

- **Proximity:** Tokens that are nearby tend to be grouped.
- **Similarity:** Similar tokens tend to be grouped together.
- **Common fate:** Tokens that have coherent motion tend to be grouped together.
- **Common region:** Tokens that lie inside the same closed region tend to be grouped together.
- **Parallelism:** Parallel curves or tokens tend to be grouped together.
- **Closure:** Tokens or curves that tend to lead to closed curves tend to be grouped together.
- **Symmetry:** Curves that lead to symmetric groups are grouped together.

- **Continuity:** Tokens that lead to continuous — as in joining up nicely, rather than in the formal sense — curves tend to be grouped.
- **Familiar configuration:** Tokens that, when grouped, lead to a familiar object tend to be grouped together.

These laws are illustrated in Segmentation/Figures 11.3, 11.4, 11.6, and **??**.

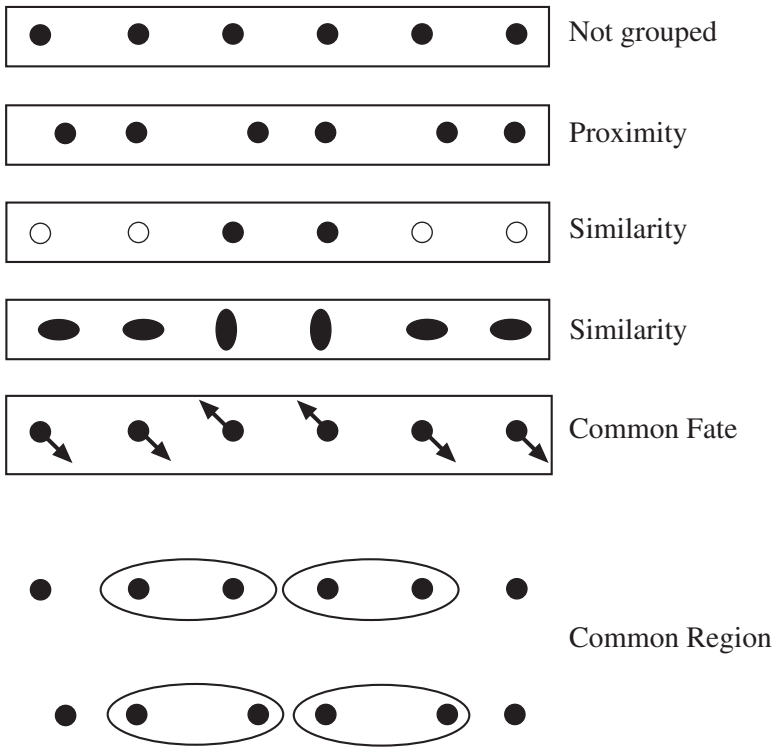

FIGURE 11.3: *Examples of Gestalt factors that lead to grouping (which are described in greater detail in the text).*

These rules can function fairly well as explanations, but they are insufficiently crisp to be regarded as forming an algorithm. The Gestalt psychologists had serious difficulty with the details, such as when one rule applied and when another. It is difficult to supply a satisfactory algorithm for using these rules — the Gestalt movement attempted to use an extremality principle.

Familiar configuration is a particular problem. The key issue is to understand just *what* familiar configuration applies in a problem and how it is selected. For example, look at Figure **??**. One might argue that the blobs are grouped because they yield a sphere. The difficulty with this view is explaining how this occurred — where did the hypothesis that a sphere is present come from? A search through all views of all objects is one explanation, but one must then explain how this search is organized. Do we check *every view* of *every* sphere with *every* pattern of spots? How can this be done efficiently?

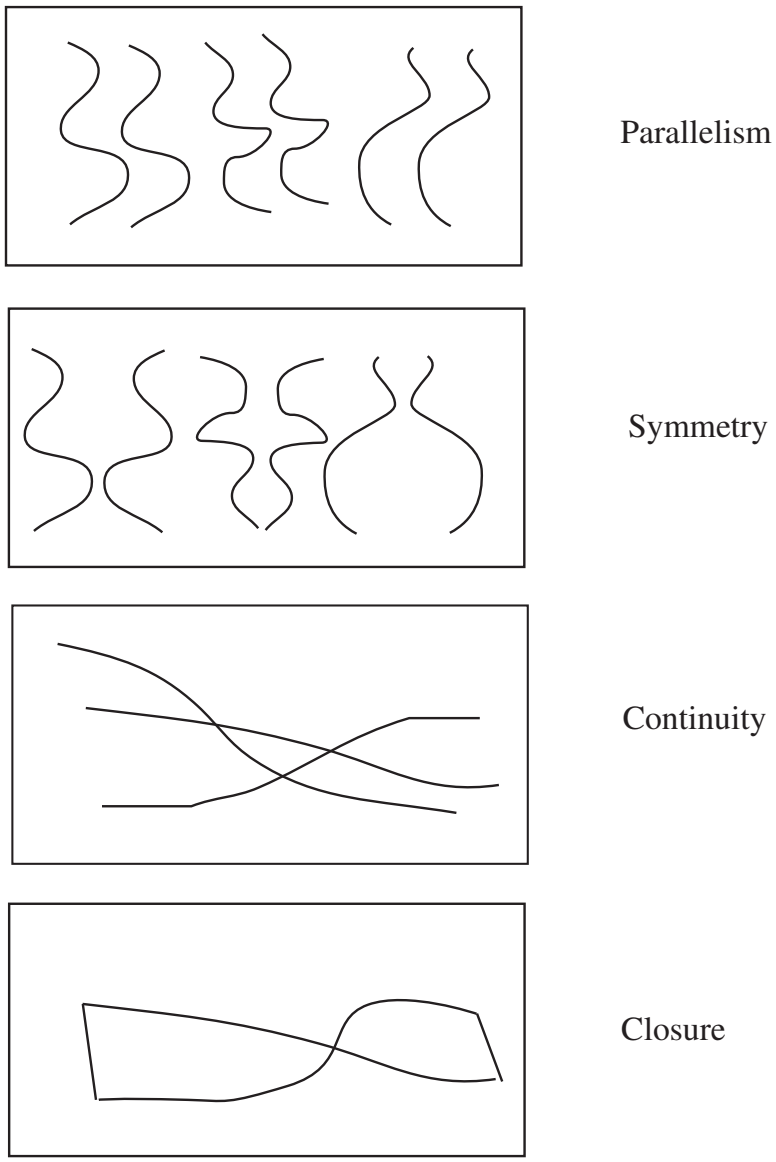

FIGURE 11.4: *Examples of Gestalt factors that lead to grouping (which are described in greater detail in the text).*

The Gestalt rules do offer some insight because they explain what happens in various examples. These explanations seem to be sensible because they suggest that the rules help solve problems posed by visual effects that arise commonly in the real world — that is, they are *ecologically valid*. For example, continuity may represent a solution to problems posed by occlusion — sections of the contour of an occluded object could be joined up by continuity (see Figure 11.5).

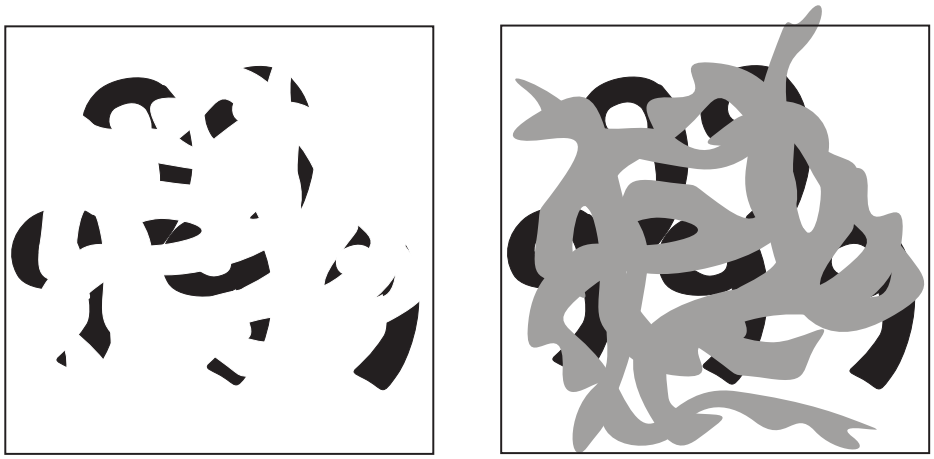

FIGURE 11.5: *Occlusion appears to be an important cue in grouping. It may be possible to see the pattern on the* **left** *as a collection of digits; that on the* **right** *is quite clearly some occluded digits. The black regions on the left and right are the same. The visual system appears to be helped by evidence that separated tokens are separated for a reason, rather than just scattered.*

This tendency to prefer interpretations that are explained by occlusion leads to interesting effects. One is the *illusory contour*, illustrated in Figure 11.7. Here a set of tokens suggests the presence of an object most of whose contour has no contrast. The tokens appear to be grouped together because they provide a cue to the presence of an occluding object, which is so strongly suggested by these tokens that one could fill in the no-contrast regions of contour.

This ecological argument has some force because it is possible to interpret most grouping factors using it. Common fate can be seen as a consequence of the fact that components of objects tend to move together. Equally, symmetry is a useful grouping cue because there are a lot of real objects that have symmetric or close to symmetric contours. Essentially, the ecological argument says that tokens are grouped because doing so produces representations that are helpful for the visual world that people encounter. The ecological argument has an appealing, although vague, statistical flavor. From our perspective, Gestalt factors provide interesting hints, but should be seen as the *consequences* of a larger grouping process, rather than the process itself.

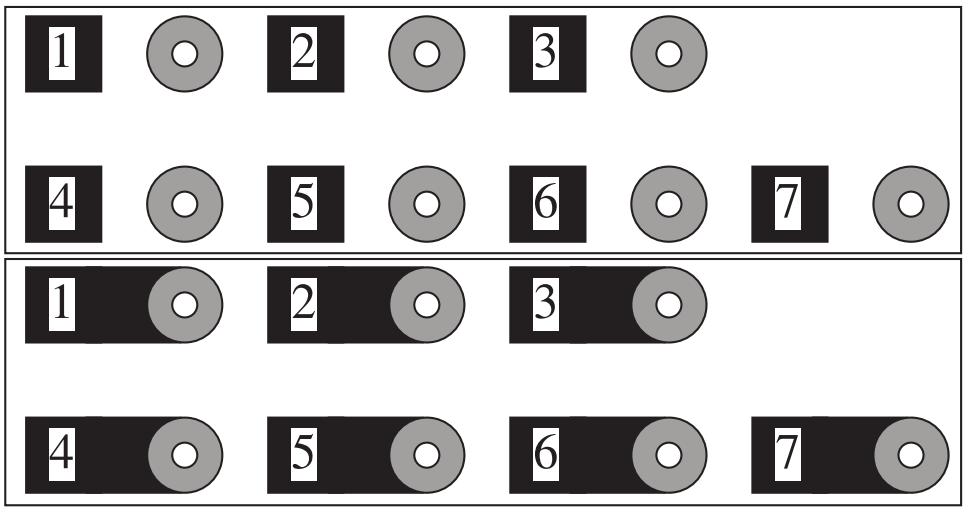

FIGURE 11.6: *An example of grouping phenomena in real life. The buttons on an elevator in the computer science building at U.C. Berkeley used to be laid out as in the* **top** *figure. It was common to arrive at the wrong floor and discover that this was because you'd pressed the wrong button — the buttons are difficult to group unambiguously with the correct label, and it is easy to get the wrong grouping at a quick glance. A public-spirited individual filled in the gap between the numbers and the buttons, as in the* **bottom** *figure, and the confusion stopped because the proximity cue had been disambiguated.*

FIGURE 11.7: *The tokens in these images suggest the presence of occluding objects whose boundaries don't contrast with much of the image. Notice that one has a clear impression of the position of the entire contour of the occluding figures. These contours are known as* illusory contours.