Kubovy M & Gepshtein S (2000). Gestalt: From phenomena to laws. In Boyer KL and Sarkar A (Eds) Perceptual Organization for Artificial Vision Systems, 41-71. Kluwer Academic Publishers, Boston, MA, USA.

Chapter 5

GESTALT: FROM PHENOMENA TO LAWS

Michael Kubovy
University of Virginia

Sergei Gepshtein
University of Virginia

Abstract The Gestalt phenomena of grouping in space and in space-time (proximity, similarity, good continuation, common fate, apparent motion and so on) are an essential foundation of perception. Yet they have remained fairly vague, experimentally intractable, and unquantified. We describe progress we made in the quest for clarity, lawfulness and precision in the formulation of these phenomena.

Keywords: gestalt, grouping, proximity, similarity, good continuation, common fate, apparent motion, motion perception, space, time

1. INTRODUCTION

When we look at a collection of discrete entities, we often see it partitioned, or organized, into subsets, or groupings, which in turn consist of parts. Such a common observation reveals the operation of constraints the brain imposes upon processing of raw visual information. These constraints, traditionally known as principles, or laws, of perceptual organization, were first investigated by Gestalt psychologists. The first two principles of perceptual organization that Wertheimer (Wertheimer, 1923) proposed in his seminal paper are grouping by proximity (Bell and Bevan, 1968; Brunswik and Kamiya, 1953; Krechevsky, 1938; Prytulak and Brodie, 1975; Rock and Brosgole, 1964; Uttal, 1981), illustrated in Figs. 5.1A and 5.1B and grouping by similarity (Beck, 1966; Beck, 1967; Bell and Bevan, 1968; Olson and Attneave, 1970), illustrated in

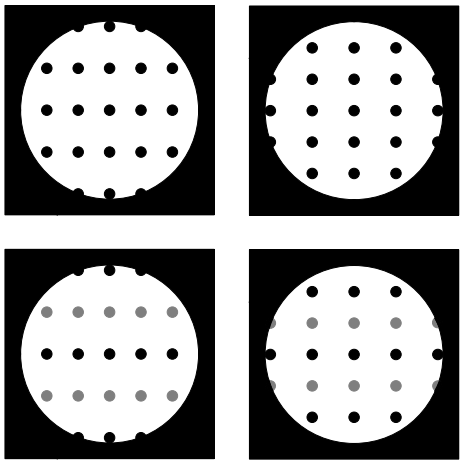

Figure 5.1 Examples

Figs. 5.1C and 5.1D. Traditional Gestalt theories of perceptual grouping have been vague and qualitative. It has long been the hope of researchers to measure the strength of a perceptual grouping. Without the ability to measure, we cannot tell by how much an organization is weakened or strengthened under different conditions, and therefore there is little hope of elucidating the underlying mechanisms.

Hochberg (Hochberg and Silverstein, 1956) proposed to measure the relative strengths of two grouping principles by pitting one against the other. We would expect grouping to be weaker if the dots in Fig. 5.1A were more evenly distributed or if the colors of the dots in Fig. 5.1C differed less. When proximity and similarity are in mutual opposition (Fig. 5.1D), one of them prevails; but this can be made to change by weakening one principle or strengthening the other. Hochberg and his associates used 6×6 rectangular lattices of squares (Hochberg and Silverstein, 1956) and 4×4 rectangular lattices of dots (Hochberg and Hardy, 1960). They determined which values of proximity and luminance are in equilibrium with respect to their grouping strength. For instance, while the spacing between columns remained constant, observers were asked to adjust the spacing between the rows of different luminance (Fig. 5.1D) until they found the spacing for which their tendency to see rows and columns was in equilibrium. Using this method, Hochberg (Hochberg and Hardy, 1960) plotted what microeconomists call an *indifference curve* (Krantz et al., 1971).[1] Hochberg reduced the luminance difference between the rows, the distance between rows for which observers reported an equilibrium between rows and columns increased (Fig. 5.2). We call this

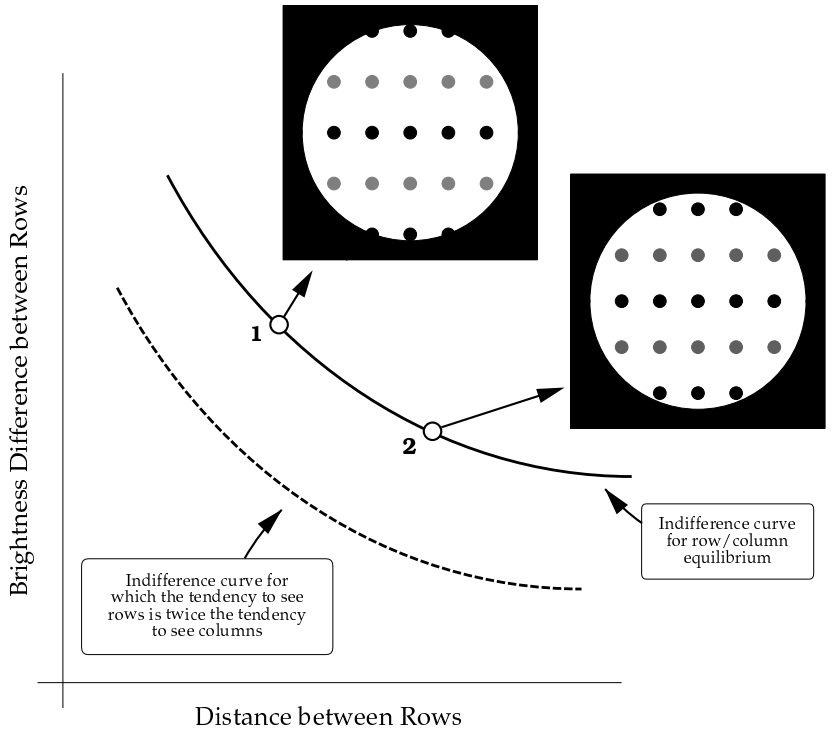

Figure 5.2 Two grouping indifference curves. Only the solid curve is achievable by methods such as Hochberg's (for the trade-off between grouping by proximity and grouping by similarity) and Burt and Sperling (for the trade-off between grouping by spatial proximity and grouping by temporal proximity). Our method allows us to plot a *family* of indifference curves.

is a *grouping indifference curve* because the observer—whose task is to find the point of equilibrium between grouping by rows and grouping by columns—is indifferent among the ⟨luminance-difference, row-distance⟩ pairs that lie on it. We developed a new quantitative approach to grouping by proximity in ambiguous patterns.

2. GROUPING BY PROXIMITY IN SPACE

Our purpose was to measure the strength of grouping by proximity, without reference to another principle of grouping. We used arrays of dots, similar to those used by the Gestalt psychologists in their classic demonstrations. On each trial we briefly presented one of sixteen dot lattices (Fig. 5.3 shows nine of them), randomly rotated, and seen through a circular aperture. (We call it an aperture because it was designed to give the impression that the lattice continues behind it. Dots

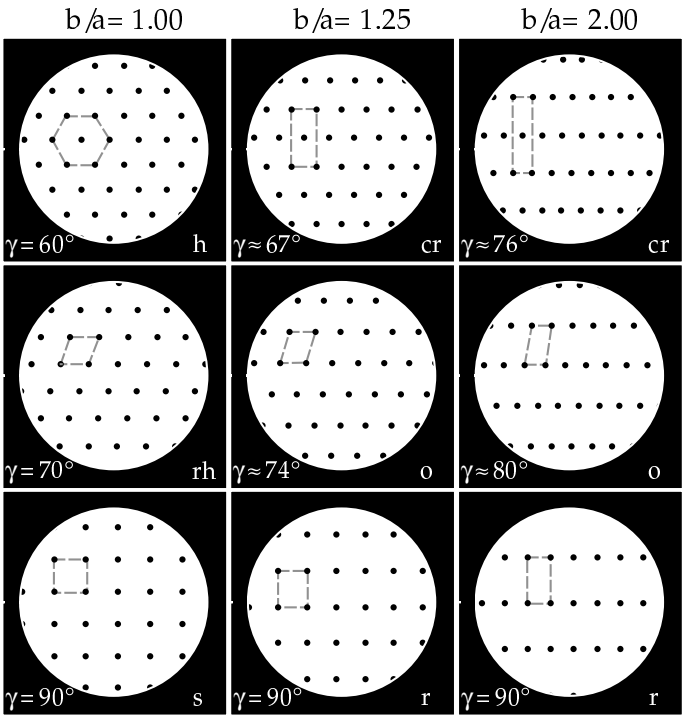

Figure 5.3 Nine examples of the 16 dot lattices used by Kubovy and Wagemans: *h*—hexagonal; *cr*—centered rectangular; *s*—square; *r* rectangular.

that fell on its boundary were partially occluded.) We offered observers the choice among four directions, and we asked them to choose the one that corresponded to the organization of the dots. Thus the task was a four-alternative forced-choice, without a correct response.

We minimized the effect of frames of reference in three ways: we minimized (a) effects of field shape by presenting the lattices as if seen through a circular aperture; (b) effects of the environmental frame of reference by randomly rotating the lattices on each trial; (c) effects of frames of reference induced by the lattice itself by restricting the number of rectangular lattices (which have the potential of creating an implicit system of coordinates) to four (out of sixteen).

2.1 DOT LATTICES

Fig. 5.4 shows the main features of a dot lattice. A dot lattice is a collection of dots in the plane that is invariant under two translations. A lattice is specified by its two shortest translations in the directions AB and AC, i.e., a pair of translation vectors **a** and **b**. Bra-

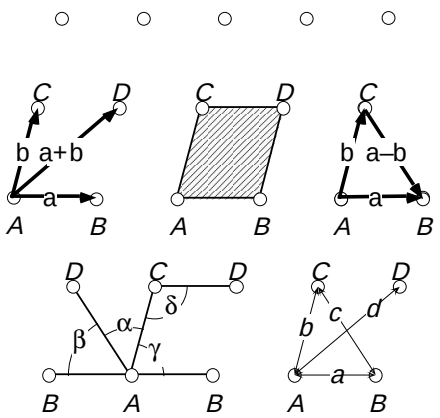

Figure 5.4 The main features of a dot lattice.

vais (Bravais, 1949), the father of mathematical crystallography, showed that the basic parallelogram of all lattices, $ABDC$, whose two sides are the vectors $\mathbf{a}$ (AB) and $\mathbf{b}$ (AC), is limited by the following conditions: $|\mathbf{a}| \leq |\mathbf{b}| \leq |\mathbf{a}-\mathbf{b}| \leq |\mathbf{a}+\mathbf{b}|$ ($AB \leq AC \leq BC \leq AD$). (We denote the magnitude—or length or norm—of vector $\mathbf{x}$ by $|\mathbf{x}|$. The symbols '+' and '−' represent vector addition and subtraction.) For all lattices, $60° \leq \angle BAC \leq 90°; 45° \leq \angle ACB \leq 90°; 0° \leq \angle ABC \leq 60°$.

The distance of any dot from its eight nearest neighbors is $|\mathbf{a}|$, $|\mathbf{b}|$, $|\mathbf{a}-\mathbf{b}|$, and $|\mathbf{a}+\mathbf{b}|$ (for convenience we will from now on denote the latter two distances $|\mathbf{c}|$ and $|\mathbf{d}|$). A lattice's basic parallelogram (and hence the lattice itself) is specified by three parameters—$|\mathbf{a}|$, $|\mathbf{b}|$ and $\gamma = \angle(\mathbf{a},\mathbf{b})$—hence if $|\mathbf{a}|$ is held constant, any lattice can be located in a two-parameter space whose coordinates are $|\mathbf{b}|$ and $\gamma = \angle(\mathbf{a},\mathbf{b})$. These coordinates are given in Fig. 5.3 to specify each lattice. Kubovy (Kubovy, 1994) showed that the lattices fall into 6 classes (one more than Bravais's classification) whose abbreviations are given in parentheses and label each of the lattices in Fig. 5.3: hexagonal (h), rhombic (rh), square (s), rectangular (r), centered rectangular (cr), and oblique (o).

Kubovy and Wagemans (Kubovy and Wagemans, 1995) presented each lattice for 300 ms on a CRT. The screen was divided into two regions, a blue disk (subtending 12.6 degrees of visual angle, dva) in the center of the screen and a black region around it. The lattices, which consisted of large number of yellow dots ($\sim$ 0.125 dva radius, no less than 1.5 dva apart), were visible in the blue region of the screen only (Fig. 5.5). After removing the lattice, they showed the observer a four-alternative response screen (Fig. 5.6). Each alternative consisted of a circle and

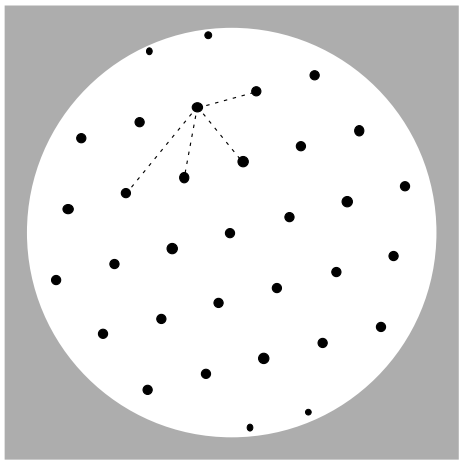

Figure 5.5 A stimulus with four vectors

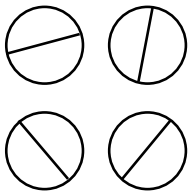

Figure 5.6 The response alternatives for the stimulus shown in Fig. 5.5

one of its diameters. The direction of the diameter corresponded to the direction of one of the four vectors of the lattice just presented.

Generalizability to other stimuli. The rationale for using dot lattices does not stand or fall on the issue of generalizability to other stimuli. Graham (Graham, 1989, p. 12) at the beginning of her book *Visual Pattern Analyzers*, observed that the results of

> experiments ... using stimuli ... at contrasts ... above ... detection thresholds ... are either sketchy or else difficult to interpret in terms of a rigorous model.

This did not deter her from writing a comprehensive review of near-threshold pattern vision. Stimuli were presented at threshold to elucidate mechanisms, not because the perception of patterns at threshold is important to the organism.

One could think of dot lattices as microscopes for the study of grouping processes. If the visual system is, *inter alia*, designed to group discrete elements into objects, then giving the system many concurrent

opportunities to perform grouping in a given orientation, may provide tractable data on the operation of local mechanisms of grouping.

2.2 THE PURE DISTANCE LAW

Procedure. Observers saw about 300 presentations of each of sixteen lattices, each randomly rotated. They were told that lattices could be perceived as collections of strips of dots and that the same lattice could have alternative organizations. They used a computer mouse to indicate the perceived organization of the lattice (i.e., the direction of the strips) by selecting one of the four circles on the response screen.

The law. The Kubovy, Holcombe and Wagemans (Kubovy et al., 1998) reanalysis of these data uses four probabilities—$p(a)$, $p(b)$, $p(c)$, and $p(d)$. (We use $\mathbf{a}$ (bold) to refer to the vector, $|\mathbf{a}|$ to refer to its length, and a to refer to the corresponding response.) Because we have only three degrees of freedom in these data we reduced them to three dependent variables by calculating $p(b)/p(a)$, $p(c)/p(a)$, and $p(d)/p(a)$. We will refer to b, c, and d collectively as v. Thus $p(v)/p(a)$ means $p(b)/p(a)$ for some data points, $p(c)/p(a)$ for others, and $p(d)/p(a)$ for the remaining ones. In addition, because in our data the range of these probability ratios is large, our dependent variable(s) are $\ln[p(v)/p(a)]$.

Fig. 5.7 shows the results of the reanalysis. This linear function, which we call the *attraction function*, whose slope is s, accounts for more than 95% of the variance. Notice the three different data symbols: they represent the data for the log odds of choosing, b, c, or d relative to a. This implies that the attraction function is a decaying exponential function with s as its parameter. This is the *Pure Distance Law.*

Let us denote the set of responses $V = \{a, b, c, d\}$, and the corresponding vectors of the lattice $\mathbf{V} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$. We find empirically that:

$$\ln \frac{p(b)}{p(a)} = -s\left(\frac{|\mathbf{b}|}{|\mathbf{a}|} - 1\right); \tag{5.1}$$

$$\ln \frac{p(c)}{p(a)} = -s\left(\frac{|\mathbf{c}|}{|\mathbf{a}|} - 1\right); \tag{5.2}$$

$$\ln \frac{p(d)}{p(a)} = -s\left(\frac{|\mathbf{d}|}{|\mathbf{a}|} - 1\right). \tag{5.3}$$

More concisely: for every $v \in V$ $(v \neq a)$ and $\mathbf{v} \in \mathbf{V}$ $(\mathbf{v} \neq \mathbf{a})$,

$$\ln \frac{p(v)}{p(a)} = -s\left(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1\right), \tag{5.4}$$

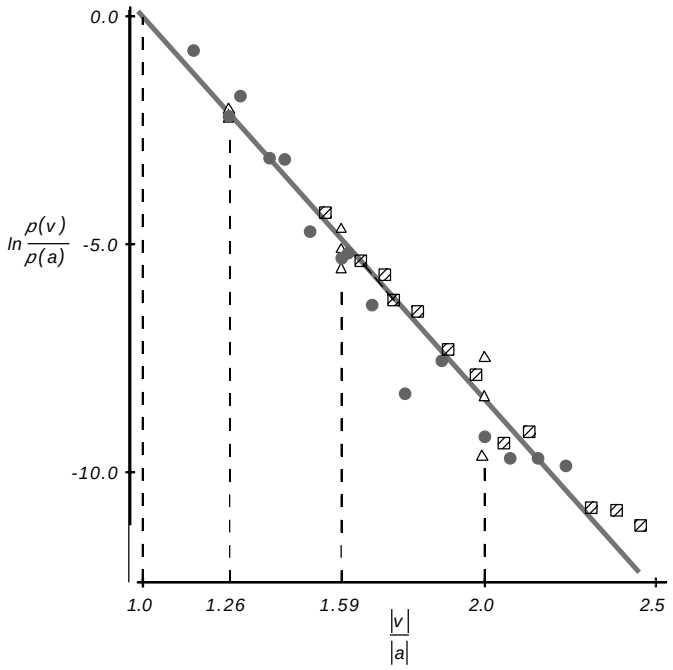

Figure 5.7 The pure distance law for dot lattices.

or

$$\frac{p(v)}{p(a)} = \exp[-s(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1)]. \tag{5.5}$$

This is a quantitative law of grouping by proximity, which states that grouping follows a decaying exponential function of inter-dot distances. We refer to this empirical relationship as a law, because it holds for a variety of patterns. Since grouping in the patterns is predicted based solely on inter-dot distances, we refer to this law as *pure distance law*.

3. GROUPING BY PROXIMITY AND SIMILARITY

3.1 THE HYPOTHESIS

Based on our findings with homogeneous dot lattices, we develop our hypothesis and articulate it as a probabilistic model for perceptual grouping and grouping variability:

1. Let us refer to our dependent variable, $p(v)/p(a)$ as *grouping strength* $\phi(v)$, where
$$\phi(v) = \frac{p(v)}{p(a)} = \exp[-s(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1)]. \tag{5.6}$$

2. **Assumption of Independence:** The probability of choosing v (note that responses are in *italics*), $p(v)$, is:
$$p(v) = \frac{\phi(\mathbf{v})}{\phi(\mathbf{a}) + \phi(\mathbf{b}) + \phi(\mathbf{c}) + \phi(\mathbf{d})} \tag{5.7}$$
which means that relative salience of the four orientations determines which orientation would be perceived.

3. The attraction function is:
$$\frac{p(v)}{p(a)} = \frac{\phi(\mathbf{v})}{\phi(\mathbf{a})} = e^{-s(\frac{|\mathbf{v}|}{|\mathbf{a}|}-1)}. \tag{5.8}$$
Or,
$$\ln\frac{p(v)}{p(a)} = \ln\frac{\phi(\mathbf{v})}{\phi(\mathbf{a})} = -s(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1), \tag{5.9}$$
which is the same as our empirically determined attraction function.

4. **Hypothesis of Additivity:** The effect of dissimilarity applied to $\mathbf{v}$ on grouping is to divide $\phi(\mathbf{v})$ by k, where k is a function of the dissimilarity. If the heterogeneity is applied to $\mathbf{v}$, then the attraction function becomes:
$$\ln\frac{p(v)}{p(a)} = \frac{\frac{\phi(\mathbf{v})}{k}}{\phi(\mathbf{a})} = -s(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1) - k; \tag{5.10}$$
but if the heterogeneity is applied to $\mathbf{a}$, then:
$$\ln\frac{p(v)}{p(a)} = \frac{\phi(\mathbf{v})}{\frac{\phi(\mathbf{a})}{k}} = -s(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1) + k. \tag{5.11}$$

3.2 EMPIRICAL TEST

To explore the hypothesis of independence and additivity in grouping by similarity we draw on the fact that there are three types of two-feature

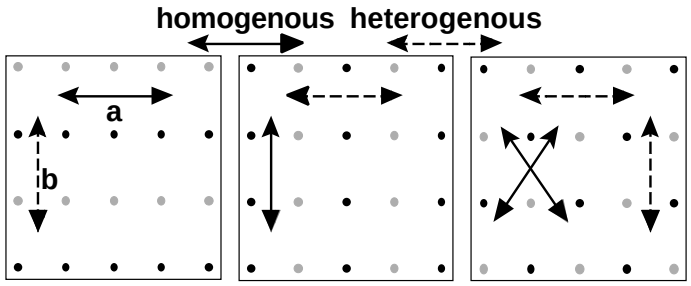

Figure 5.8 The three ways to introduce effects of similarity on grouping in a dot lattice.

dot lattices. Fig. 5.8 represents these features by different contrasts of the dots with the background. Going from left to right, they are: (i) homogeneous in the **a** direction (heterogeneous in the **b** direction), (ii) homogeneous in the **b** direction (heterogeneous in the **a** direction); and (iii) homogeneous in the **c** and **d** directions (heterogeneous in the **a** and **b** directions).

There is little literature on the joint effects of proximity and brightness similarity on grouping. We have already reviewed the work of Hochberg and his colleagues (Hochberg and Hardy, 1960; Hochberg and Silverstein, 1956). Zucker (Zucker et al., 1983) used rectangular, hexagonal, and circular random-dot moiré patterns (also known as Glass patterns), and confirmed the joint effectiveness of proximity and brightness similarity. Their most important finding was that neighboring dots showed no tendency to group if their contrast against the background was of opposite sign (see also Glass and Switkes, 1976).

3.3 PRELIMINARY DATA

We showed four subjects a variety of dot lattices. The procedure followed Kubovy and Wagemans (Kubovy et al., 1998). In this preliminary experiment we used three differences in lightness between adjacent dots: $\Delta I = 11, 14, 17$ on the computer's RGB scale. The results for each ΔI are in a separate panel of Fig. 5.9. The middle graph of each panel, labeled '$==$,' with an intercept of 0, was obtained with uniformly colored dots, i.e., it is a control, $\Delta I = 0$. It replicates the pure distance law (Fig. 5.7), and has a similar slope. The upper curve in each graph ('$=\neq$') was obtained with lattices in which **b** is homogeneous and **a** is heterogeneous. Recall that our dependent variable is $\ln[p(b)/p(a)]$ and $p(b) \geq p(a)$. Because the **a** direction is at a grouping disadvantage, the $=\neq$ curve must be higher than the $==$ curve. The lower curve in each

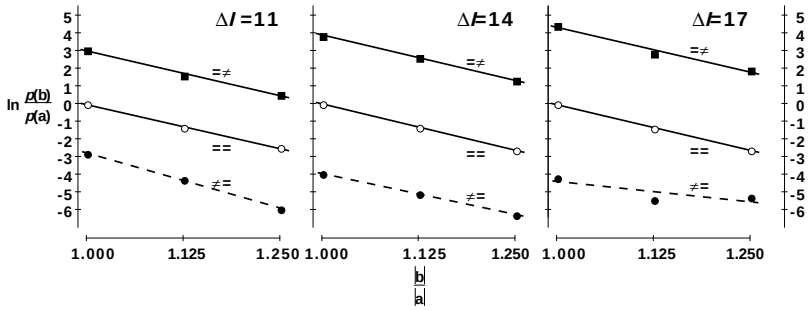

Figure 5.9 The pure distance law modulated by lightness similarity.

graph ('$\neq =$') was obtained with lattices in which **b** is heterogeneous and **a** is homogeneous. This time the **b** direction is at a grouping disadvantage, and so the $\neq =$ curve must be lower than the $==$ curve.

The first interesting fact about these data is that the effects of ΔI and $|\mathbf{b}|/|\mathbf{a}|$ data are additive: the three $= \neq$ lines in the three panels of Fig. 5.9 are parallel. Similarly the three $\neq =$ lines in the three panels are parallel (with some tendency to flattening for the $\Delta I = 17$, probably due to a floor effect). (The $==$ lines in the three panels represent the same data.)

The meaning of additivity. If further experiments support our preliminary data, i.e., if the effects of ΔI and $|\mathbf{b}|/|\mathbf{a}|$ on $\ln[p(b)/p(a)]$ are additive, we will be in a position to do three things: (a) to draw a family of indifference curves between lightness and distance; (b) to derive a common scale for the measurement of differences in lightness and relative dot distances; (c) to claim that there exists a mechanism that takes information about differences in lightness and relative dot distances and combines them multiplicatively.

Drawing a family of indifference curves. Drawing the first indifference curve is straightforward (Fig. 5.10). It is the indifference curve for the $\Delta I, |\mathbf{b}|/|\mathbf{a}|$ combinations for which **a** and **b** are in equilibrium, i.e., $\ln[p(b)/p(a)] = 0$. Find the 0-crossings of the three $\neq =$ lines. The abscissa of these 0-crossings are the three $|\mathbf{b}|/|\mathbf{a}|$ values that correspond to the three ΔI values. Similarly, find the 0-crossings of the three $= \neq$ lines, and find the corresponding abscissae. The values may be less than 1, indicating that only if $|\mathbf{b}| < |\mathbf{a}|$ can the equilibrium be achieved. The second indifference curve is the one for which $\ln[p(b)/p(a)] = 1$, i.e., $p(b) = 2.718p(a)$. The procedure is analogous to

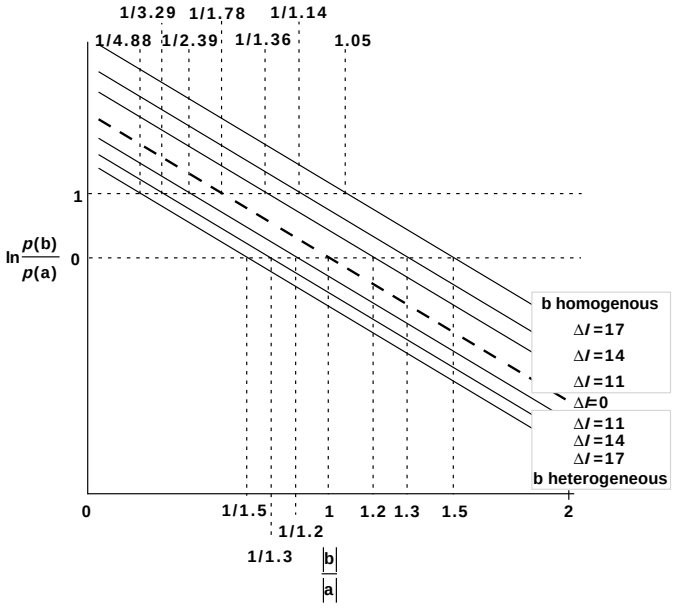

Figure 5.10 Hypothetical measurements used in the construction of two grouping indifference curves. The dashed line corresponds to the attraction function for $\Delta I = 0$, i.e., a homogeneous dot lattice.

the one we just described for the equilibrium condition, except that we find the abscissae of the intersections of the attraction functions with the line $\ln[p(b)/p(a)] = 1$.

We can now construct the indifference curves, in which we plot the two sets of $\Delta I, |\mathbf{b}|/|\mathbf{a}|$ combinations against each other (Fig. 5.11). Why don't they look like the curves in Fig. 5.2? Because the trade-off is formulated differently: we have formulated our results as if the two variables, $|\mathbf{b}|/|\mathbf{a}|$ and ΔI were applied to *competing* vectors. To keep constant the strength of one vector relative to the other (which is what we mean by indifference), we must increase one variable when we increase the other.

Application to scaling within and between continua. To see how we might use this technique to scale within continua, we continue the example we have been pursuing. Suppose that (a) we hold the luminance of the background constant, and (b) we hold the luminance of the dot with least contrast constant at I_1; the other dot's luminance would then be $I_1 + \Delta I$. Let us denote the equilibrium indifference curve for these stimuli $\mathcal{E}_1$. Now let us increase the luminance of the dot with least contrast to I_2, and find the equilibrium indifference curve for these

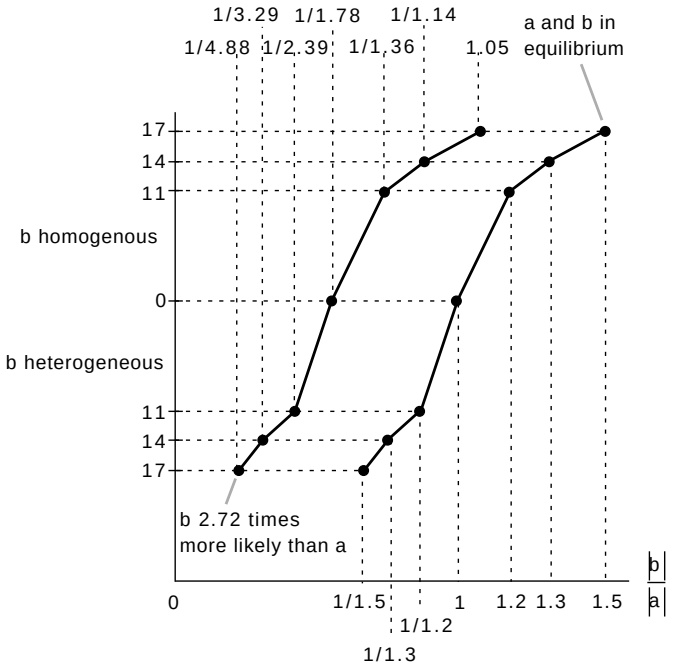

Figure 5.11 Two grouping indifference curves. Note the difference from Fig. 5.2; it is due to the fact that inhomogeneity and length increments are applied to competing vectors (see text).

stimuli, $\mathcal{E}_2$. Since by design the two indifference curves $\mathcal{E}_1$ and $\mathcal{E}_2$ are the same (i.e., they are both equilibrium indifference curves), and the $|\mathbf{b}|/|\mathbf{a}|$ axis is common to both, the ΔI scale allows us to compare luminance differences even when all the luminances in set 1 are much lower than the luminances in set 2. We can, of course, do the same with different background levels thus determining whether contrast and absolute luminance levels interact (and, if they do, how).

Research on the problem of color scaling (Derefeldt, 1991) is too extensive for us to review here, but we will point out that our results may offer a new approach to this problem. The technique we have outlined allows us to compare the similarity of colors. Hitherto, color scaling has been based either on subjective judgments of similarity or on just-noticeable differences of colors. We now have a new standard by which color differences can be scaled: we define the difference between two colors to be the $|\mathbf{b}|/|\mathbf{a}|$ ratio ($|\mathbf{b}| \geq |\mathbf{a}|$) with which they are in equilibrium when the **b** orientation is uniform in color and the two colors are applied in alternation to dots in the **a** orientation. To be sure, this method has limitations: (a) it can only be applied to the comparison of fairly small

patches (subtending at the most 1 dva); (b) data collection is laborious. Nevertheless given that each of the various color appearance systems of necessity are based on compromises, and each has its advantages and disadvantages, it is possible that the present technique may yield results that are closer to one of them.

The notions presented in the preceding paragraph generalize well to the scaling of other attributes that may affect grouping.

Comment on mechanisms. We hope to confirm that for colored (chromatic and achromatic) dot lattices,

$$\ln \frac{p(b)}{p(a)} = -s(\frac{|\mathbf{b}|}{|\mathbf{a}|} - 1) + f(B, I_0, \Delta I), \tag{5.12}$$

where B is the background luminance, I_0 is the luminance of the dot with the lower contrast, and ΔI is the luminance difference between the two dot types. Rewriting Eq. 5.12,

$$\begin{aligned} \frac{p(b)}{p(a)} &= \exp[-s(\frac{|\mathbf{b}|}{|\mathbf{a}|} - 1) + f(B, I_0, \Delta I)] \\ &= \exp[-s(\frac{|\mathbf{b}|}{|\mathbf{a}|} - 1)] \exp[f(B, I_0, \Delta I)]. \end{aligned} \tag{5.13}$$

That is to say, the b:a odds are a multiplicative function of the relative distance and dissimilarity. One way to express this result is to say that dissimilarity is a gain control on the strength of grouping by proximity. It is hard to imagine how such a multiplicative operation could be implemented unless the two variables in question provided their output to a later stage that performed the grouping.

4. GROUPING BY PROXIMITY IN SPACE–TIME

Thus far we have been discussing the perceptual organization of static displays. We were concerned with "spatial grouping," the process by which small visual elements are linked across space to form larger and more complex entities, culminating in the perception of objects and surfaces. In a natural ecology, however, perceptual organization must be dynamic, because in general the observer—and often objects—move. So we must add the dimension of time to the domain of our research.

The limiting case of perceptual organization in space-time is the perception of a stationary scene, when the observer views a static display from a fixed vantage point. This is the case we explored in our studies of spatial grouping: successive elements of the visual input are separated by null spatial displacement. If we introduce spatial displacement

between the elements at successive instants, motion may be perceived. When motion is seen, each visual element may either have a separate identity and move separately, or it may merge with other elements to form a larger–scale object, so that the element's motion will be the same as the object's.

4.1 SPATIAL AND TEMPORAL GROUPING

To perceive motion, vision must link successive elements across space and across time, a process we call "temporal grouping." In this process the visual system establishes a *correspondence* between elements visible at successive instants, which is why temporal grouping is also known as *matching*. Matching determines which elements at successive instants will retain their identity, i.e., will appear to belong to the same object. The visual entities that undergo temporal grouping are called *matching units* or *correspondence tokens* (Ullman, 1979).

What is the relation between spatial and temporal grouping? We will consider two models: (a) According to the *sequential model* (SM; Fig. 5.12*a*) spatial and temporal grouping are separable and serial. (b) According to the *interactive model* (IM; Fig. 5.12*b*), spatial and temporal grouping are *in*separable, because the propensity of elements to group across space and across time is interdependent.

The SM does not imply an ordering of spatial and temporal grouping. According to one version of the SM (hinted at by Neisser, 1967, who viewed motion perception as an integration of successive snapshots of the scene) spatial grouping alone determines the visual entities that will undergo temporal grouping. But temporal grouping may precede spatial grouping. This happens, for example, in random-dot cinematograms, which are dynamic displays in which each frame contains a different random texture designed so that no shape can be seen in any individual frame. If the frames are correlated, so that some elements within a region coherently change their location across the frames while the remaining dots are either stationary or change their positions at random, the correlated region segregates from the rest of the display due to spatial grouping of similar moving elements. Gestaltists referred to such organization as grouping by *common fate* (Wertheimer, 1923).

Spatial and temporal grouping in the theories of motion perception. Evidence from a variety of experiments converges to show that vision derives several parallel spatial visual representations: local luminances, spatially–segregated features, or more complex entities. Currently the dominant low-level approach to the perception of dynamic scenes distinguishes between three systems, or mechanisms, mediating

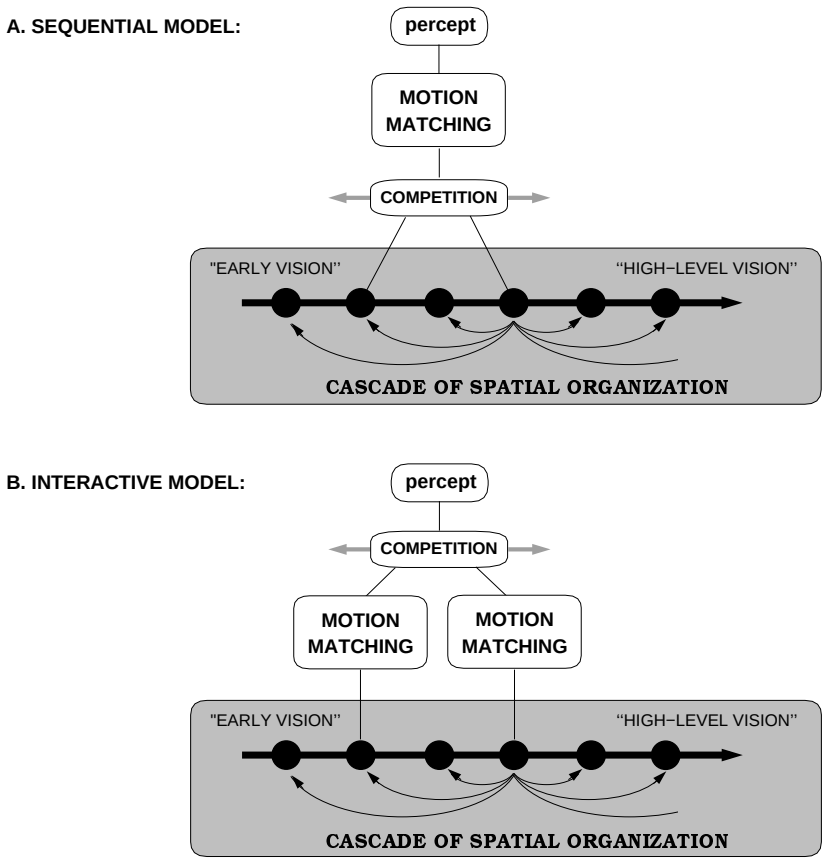

Figure 5.12 **A.** According to the sequential model (SM), the operations of spatial and temporal grouping are separable, i.e., perception of motion can be explained by sequential application of the two kinds of grouping. For example, spatial grouping may precede temporal grouping, as shown in the diagram: Visual representations of different complexity (or scale) compete before temporal grouping as to which of the representations will derive the matching units for temporal grouping. The more salient spatial organization (see text) will determine *what* is seen to move. The notion of "cascade of spatial organization" refers to multiple spatial representation of the visual scene, where (a) more complex representations may contain entities from less complex ones (in this sense the cascade is hierarchical), (b) and each representation may interact with both more and less complex representations (shown for one level for simplicity). In the model, the mechanisms of temporal grouping can access the alternative representations in parallel, as soon as they become available (McClelland, 1979). **B.** According to the interactive model (IM), spatial and temporal grouping are inseparable in the sense that both spatial and temporal grouping determine the level of spatial organization at which matching units arise. In the IM, the competition occurs between the outputs of parallel motion matching operations (Wilson et al., 1992) applied to different levels in the cascade of spatial organization. Thus, the salience of both spatial and temporal grouping contributes into the formation of matching units.

the perception of motion (Lu and Sperling, 1995), based on these alternative spatial representations: (a) The *first–order* system (also called the Fourier motion system) is only sensitive to the modulations of *local luminances*. This system is insensitive to spatial visual configuration and thus constitutes a temporal grouping operation applied directly to the raw visual input. (b) A second–order system (also called non–Fourier motion system) has been proposed to account for motion that is seen in texture stimuli whose motion is invisible to the first–order system. This system matches spatially–segregated *features* across time and thus requires some preliminary spatial organization. The perception of second–order motion is achieved by subsequent applications of spatial and temporal grouping. The first– and second–order systems are fast and sensitive to the eye of origin, i.e., motion is not perceived if the successive frames are shown in alternation to different eyes. (c) A *third–order* system has been proposed to accommodate the perception of much slower motion, indifferent to the eye of origin. This system is able to derive motion of visual entities constructed by elaborate spatial analysis (such as figure–ground segregation), based on a variety of visual attributes (such as texture and color).

Although spatial and temporal organizations seem to operate sequentially within the systems (see, e.g., Chubb and Sperling, 1989) the three–system conception of motion perception as a whole is mute on separability of space and time in perceptual organization and therefore can be adapted to agree with either the SM or the IM. As we show in Fig. 5.12, the alternative representations could either (a) compete as to which of the spatial representations will input into the motion computation and thus will determine the identity of the moving objects, or (b) the competition could occur after motion computation within each spatial scale. In other words, the parallel representations of the scene could compete before or after the matching operation. The first scheme agrees with the SM, where spatial organization precedes matching. The second scheme agrees with the IM, because it implies that the outputs of temporal grouping will determine which level of spatial organization will dominate perception. Both alternatives have been considered in the literature. Consider the two schemes in turn:

- The issue of separability of spatial and temporal grouping pertains to the long–standing question of whether form information affects the perception of motion. The debate about form-motion interactions has often been cast in terms of the SM, perhaps because apparent motion has been commonly studied using shapes spatially well–segregated from the rest of the scene. The assump-

tion of the SM has been carried over into the more general studies of the interaction of form and motion. For example, Croner and Albright (Croner and Albright, 1997; Croner and Albright, 1999) found that spatial segmentation by form cues, such as color or luminance differences, affects the discrimination of motion masked by visual noise. They considered two alternative models of the form–motion interactions; both models assumed that spatial organization happened before motion processing, but the competition between the motions from spatially segregated regions occurred either before or after motion matching, i.e., matching of individual elements has not been considered at all.

- Wilson, Ferrera and Yo (Wilson et al., 1992) proposed a model where the matching operation is applied both directly to the raw visual input (simulating the first–order motion system) and after some preprocessing (simulating the second–order motion system). The outputs of the two parallel motion computations interact in the model to simulate the competition between motion directions, so that the most salient motion is perceived (winner–take–all). This is a truly interactive model (Fig. 5.12*b*), but it was inspired by evidence which is also consistent with the SM: (a) Physiological evidence that showed that the cortical areas responsible for motion perception receive both first– and second–order spatial information (e.g., Maunsell and Newsome, 1987). The two inputs could compete before motion matching operation (Fig. 5.12*a*). (b) Psychophysical evidence of motion integration that was launched by a seminal study of Adelson and Movshon Adelson and Movshon, 1982. These authors argued that perception of moving visual patterns could be explained by spatial integration of the motion signals of the pattern components (see Fig. 5.13*b* and the section on matching of groupings on page 22). Several later studies found that the integration of component motions depends on a variety of factors that could affect spatial grouping between the components: similarity of the components, the relative direction of their motion, and the salience of their intersections (Wilson, 1994; Smith, 1994; but see Stoner and Albright, 1994 for a different view).

In the rest of this chapter we will (a) illustrate how the SM might work, (b) show that evidence thought to support the IM is actually consistent with the SM, (c) show that *all* the evidence which supports the IM comes from a special class of visual displays, (d) present preliminary data from two experiments that suggest how to tackle the question of separability of spatial and temporal grouping.

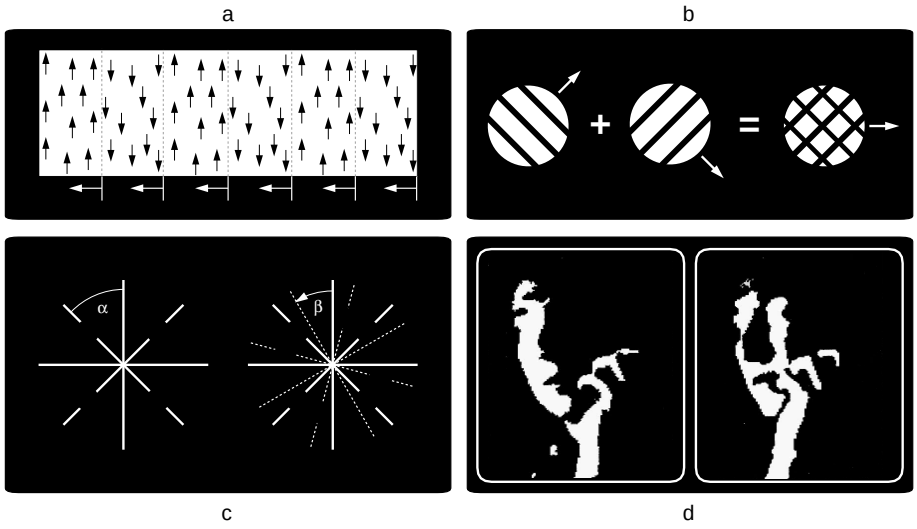

Figure 5.13 **Sequential model** Illustrations of sequential application of spatial and temporal grouping. *a.* In the "second-order" stimulus of Cavanagh and Mather (1990), random texture elements moving upward alternate with strips of random texture elements moving downward. The elements moving in common direction group by common fate (see text), which allows the mechanism of spatial grouping to derive (virtual) boundaries between the different moving strips. Temporal grouping then matches the strip boundaries between the frames to yield the percept of leftward drift. *b.* When identical drifting gratings are superimposed, they form a checkerboard pattern, a "plaid," which appears to move in the direction different from the directions of the component gratings (Adelson & Movshon, 1982). This effect can be explained by first applying spatial grouping to individual gratings, to form a plaid, which becomes a complex matching unit. *c.* The "broken wagon wheel" demonstration of Ullman (1979). LEFT: Each other spoke of the "wagon wheel" is interrupted in the middle. The angle between the neighboring spokes is α. RIGHT: When the "wheel" is rotated anticlockwise by angle $\beta > \frac{\alpha}{2}$ one sees three rotating objects: Matching between the outmost and the innermost segments of the broken and intact spokes yields two distinct clockwise motions; matching between the spoke interruptions yields an anticlockwise motion. Ullman explained this effect in the spirit of the SM, by assuming that matching units are always small-scale entities, such as line segments, also when the segments are connected. *d.* The two images of "Mooney faces" are shown to observers in rapid alternation. When the observers see a face, they perceive it rotating in depth. When they do not see a face, they perceive an incoherent motion in the picture plane (Ramachandran et al., 1998). In keeping with the SM, spatial grouping forms faces which become matching units.

4.2 THE SEQUENTIAL MODEL AND MATCHING UNITS

Ullman (Ullman, 1979) proposed a very specific version of the SM. He held that matching units are *always* small-scale visual elements. As we will see, there is evidence to suggest that this version of the SM is too restrictive because there are many examples of more complex organizations that serve as matching units. We will now describe four examples that are compatible with the SM for which the matching units are spatial organizations of increasing complexity. In the first example we will show how one can apply Ullman's model to displays that do not consist of small-scale elements. In the remaining three examples we will show that the assumption of small-scale units will not work.

The SM and Ullman's theory of matching units Our first illustration concerns displays in which the matching units are one or more continuous lines or connected regions, rather than a collection of small-scale discrete elements (Fig. 5.13*c*). We are referring to the *aperture problem* (Wallach, 1935; Wallach and O'Connell, 1953; Hildreth, 1983): a moving line is seen through an aperture that occludes the line's endpoints. We see motion orthogonal to the line. This display raises two problems: (a) Any segment on the line at time t_i may match any segment on the lsine at time t_{i+1}; this is the *correspondence problem.* (b) The size of these segments is unknown; this is the *matching unit problem.* If we follow Ullman (Ullman, 1979) and assume (a) that the visual system considers the line to be a collection of short line segments or dots, which it uses as matching units, and (b) that the visual system chooses the shortest path between successive matching units to solve the correspondence problem, then we can correctly predict the visual system's solution to the aperture problem.

The same analysis works for other displays as well. For example, Wallach, Weisz and Adams (Wallach et al., 1956) observed that if one rotates an ellipse about its center, under some circumstances it is seen as a rigid rotating object, and under others it is seen as an object undergoing elastic (non-rigid) transformation. The closer the ellipse aspect ratio is to one (i.e., the more closely it approximates a circle) the more likely we are to see an elastic transformation. In keeping with Ullman's view, Hildreth (Hildreth, 1983) assumed that the matching units are fragments of ellipse contour. She then showed that the observed effect of aspect ratio would be predicted by a system that found the smoothest velocity field that maps successive contour fragments onto each other.

The analyses of both Ullman and Hildreth fit the SM by default; for all intents and purposes these authors do not assume a spatial grouping process other than the processes that locate edges and lines. Hence on their view temporal grouping can have no influence on spatial grouping.

Recursive spatial and temporal grouping In the second illustration, we consider a hierarchical perceptual organization, i.e., a display in which elements move within objects that are themselves moving. To understand the nature of temporal grouping, consider *grouping by common fate* (Wertheimer, 1923), which occurs when elements extracted by temporal grouping are segregated from the background by spatial grouping and form a moving figure. Temporal grouping occurs for both translation (Wertheimer, 1923) and rotation (Julesz and Hesse, 1970). The elements extracted by temporal grouping may undergo more complex spatial organization than in grouping by common fate to yield, for example, a three-dimensional object (*shape-from-motion*; Ullman, 1979). This evidence is consistent with the SM because the matching units are derived by spatial grouping alone which is followed by temporal grouping. This temporal grouping determines the directions and the velocities of the elements which is used by the subsequent spatial organization to derive the objects' shape.

Cavanagh and Mather (Cavanagh and Mather, 1990) created a stimulus consisting of a collection of vertical strips consisting of random elements moving upward alternating with strips of random elements moving downward (Fig. 5.13*a*). The boundaries between these strips are easily visible; when they were made to drift to the left, observers saw a compelling motion to the left. According to the SM, the short-lived random elements are output by the earliest stage of spatial grouping, $\mathcal{S}_1$, which cannot do much grouping because the elements in each frame are random. The elements in each frame are matched by temporal grouping ($\mathcal{T}_1$) and identified as dots moving in up or down. Dots moving in the same direction undergo spatial grouping ($\mathcal{S}_2$) by common fate to generate the different-moving strips, as a result of which we see boundaries between them. These boundaries, which from frame to frame are translated to the left, serve as input to $\mathcal{T}_2$. $\mathcal{T}_2$ compares successive boundaries and detects their leftward motion (called *second-order motion* by Cavanagh and Mather (Cavanagh and Mather, 1990). Note, however, that the output of $\mathcal{T}_2$ does not depend on the fact that the boundaries between the strips are derived using temporal grouping. These boundaries could have been produced by spatial grouping based on a common property other than common motion, such as luminance or color. Thus, $\mathcal{T}_2$ is independent from $\mathcal{T}_1$, as the SM requires.

Matching of groupings In our third illustration, spatial grouping organizes visual primitives into groupings that become matching units. Adelson and Movshon (Adelson and Movshon, 1982) showed observers two superimposed moving gratings through a circular aperture (Fig. 5.13*b*). When either moving grating was presented alone, it was seen to move at right angles to the orientation of its bars. When the superimposed gratings were *identical*, the gratings were fused and were seen as a single plaid moving in an orientation different from the motion of the individual gratings. However, when the superimposed gratings were *different*, they were not fused; they were seen as two overlaid gratings, each moving at a right angle to the orientation of its bars, as if each had been displayed alone (motion transparency). We have good reason to believe that the outcome of this experiment was determined by the propensity of the gratings to fuse when overlaid as *static* gratings. From informal observations made in our lab we have concluded that when two identical static gratings are superimposed, they are likely to be seen as a plaid, but when two superimposed gratings differ in the lightness or the thickness of their bars, they are likely to be seen as two overlaid gratings. Thus from the appearance of the static displays we infer the output of the spatial grouping by similarity that derives the matching units—the gratings or the plaids—independent of temporal grouping.

Matching of high-level units In our fourth illustration we show how complex organizations become matching units. Ramachandran (Ramachandran et al., 1998) presented observers with an alternating pair of fragmented patterns, called "Mooney faces," that are sometimes seen as a face, and sometimes as a random pattern (Fig. 5.13*d*). When the observers saw the pattern as a face, they experienced motion in a direction different from the direction specified by matching of the individual fragments. Thus, the familiarity of an object can affect spatial grouping, which affects perceived motion, independent of temporal grouping, in keeping with the SM.

4.3 EVIDENCE FROM THE TERNUS DISPLAY

We will now consider the evidence that seems to support the IM, but is actually consistent with the SM.

Consider the Ternus (Ternus, 1936) display, in which dots can occupy four equally-spaced collinear positions *abcd*. These displays consist of two rapidly alternating frames—f_1 and f_2. The dots in f_1 are at *abc*; the dots in f_2 are at *bcd*. This display can give rise to two percepts:

(a) *Element motion* (*e*-motion), which occurs when the dots in positions *b* and *c* appear immobile, and a single dot appears to move between the positions *a* and *d*; (b) *Group motion* (*g*-motion), which occurs when three dots appear to move back-and-forth as a group, from *abc* to *bcd*.

The longer the inter-stimulus interval (ISI; inter-frame-interval in this context), the higher the likelihood of *g*-motion (Pantle and Picciano, 1976; Kramer and Yantis, 1997). This phenomenon is called the *ISI effect*. According to Kramer and Yantis (Kramer and Yantis, 1997) the ISI effect implies that temporal grouping between successive elements affects the spatial grouping between concurrent elements, thus supporting the IM. They assumed that the shorter the ISI, the stronger the temporal grouping. Thus, when ISI is short, temporal grouping overrides the spatial grouping of the concurrent dots, and *e*-motion is likely. As ISI grows, the strength of temporal grouping drops and allows concurrent dots to group within the frames, thus increasing the likelihood of *g*-motion.

We hold that the ISI effect is not inconsistent with the SM for two reasons: (a) Longer ISIs could have two effects: (i) they could weaken temporal grouping, as Kramer and Yantis assumed, or (ii) they could allow more time for spatial grouping to consolidate the organization of concurrent dots. If the latter is true, then we could attribute the ISI effect to spatial rather than temporal grouping, and thus is consistent with the SM. (b) If an observer sees *g*-motion, one cannot tell whether the matching units were dots or dot groupings, because in both cases matching yields motion in the same direction. Therefore, the group motion percept may actually be based on matching between individual dots, just as it is in *e*–motion, and different spatial distances would favor different kinds of *e*–motion (Korte, 1915; Braddick, 1974; Burt and Sperling, 1981).

4.4 EVIDENCE AGAINST THE SEQUENTIAL MODEL

Of all types of motion perception, the one that truly undermines the generality of the SM is the perception of overlapping objects and surfaces whose relation is changing dynamically (henceforth, *dynamic superposition*). The perception of dynamic superposition poses a challenge to the SM because spatial grouping alone cannot derive matching units when objects and surfaces are revealed gradually. Take, for example, the perception of kinetic occlusion (Michotte et al., 1964; Kaplan, 1969), where a hitherto visible (or invisible) part of the scene is perceived to become occluded by (or revealed from behind) an opaque object or surface (

Sigman and Rock, 1974; Kellman and Cohen, 1984; Tse et al., 1998). In such a case a simple correspondence between the successive views is impossible because one frame has a different number of elements than the next frame, or because the elements in successive frames are markedly different. Likewise, if the moving object or surface is transparent (Shipley and Kellman, 1993; Cicerone et al., 1995), finding correspondence is hampered because the appearance of the covered region changes as it becomes covered. Perhaps the most dramatic demonstration of perception under dynamic superposition is anorthoscopic form perception (Rock, 1981), where observers can perceive the form of an object revealed successively through a narrow slit in the occluder. In all these cases, the visual system must accumulate information over time to produce a percept which is the most likely cause of the observed optical transformations.

Although the evidence of perception under dynamic superposition undermines the SM, we think that it is too specific to carry the burden of refuting the SM in favor of the IM. Displays of dynamic superposition contain characteristic clues, which may trigger specialized mechanisms. For example, two clues present in kinetic occlusion are the accretion of texture (as the textured object emerges from behind the occluder; Kaplan, 1969), and the presence of "T-junctions" between the contours of the occluder and of the occluded object. These cues may trigger a mechanism specialized in dealing with dynamic superposition, or a high-level inferential mechanism designed to construct a plausible interpretation of the scene in a process of thought-like problem-solving (Helmholtz, 1962; Kanizsa, 1979; Rock, 1983).

4.5 MOTION LATTICES

To refute the SM, we must demonstrate that spatial and temporal grouping interact even when a simple correspondence between the successive frames is possible and no specialized, or inferential, mechanism is required. We used spatio-temporal dot lattices, *motion lattices*, in which we could independently vary the strength of spatial and temporal grouping by manipulating spatial proximity between concurrent and successive dots. As we observed earlier with regard to the Ternus displays, the duration of the ISI does not necessarily determine the strength of temporal grouping, because (a) spatial grouping may consolidate as the ISI grows, and (b) longer ISIs may favor matching over a different spatial range. Therefore in our motion lattices we held ISI constant, and varied the strength of temporal grouping by manipulating the *spatial proximity between successive dots*. In our displays, as in the Ternus displays,

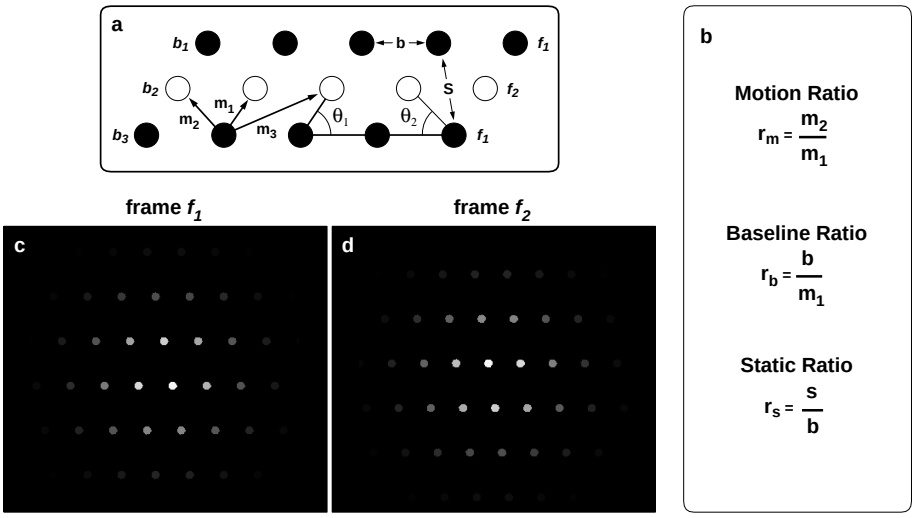

Figure 5.14 *a* Three rows of a motion lattice. Each row of dots is called a "baseline," labeled b_i. The black and white circles stand for dots that appear in different frames. The parameters of motion lattices are explained in the text (page 26). *b* Three ratios that govern the perception of motion lattices: r_s controls grouping of dots within the frames; r_m controls grouping of dots across the frames (dot-to-dot matching); r_b, controls relative salience of the two types of grouping: spatial (within the frames) and temporal (across the frames). *c* and *d* show snapshots of two successive frames of a motion lattice (not to scale).

either element motion (*e*-motion) or group motion (*g*-motion) can be seen. The advantage of motion lattices over the Ternus display is that in lattices the directions of *e*-motion and *g*-motion differ. The direction of *e*-motion is determined by matching individual dots in the successive frames of the display. The direction of *g*-motion is determined by matching of dot groupings (the strips of dots, or *virtual objects*) in successive frames; the direction of *g*-motion is orthogonal to the orientation of the objects.

Motion lattices generalize displays introduced by Burt and Sperling (Burt and Sperling, 1981), who presented observers with a succession of brief flashes of a horizontal row of dots. Between the flashes, the row was displaced both horizontally and downward, so that under appropriate conditions observers saw the row moving downward and two the right or left. Burt and Sperling studied the trade–off between space and time in motion matching, and the effect of element similarity on matching. However, they did not explore the effect of relative proximity between concurrent and successive dots, and did not report on what we call *g*–motion. In motion lattices each frame contains a two–dimensional

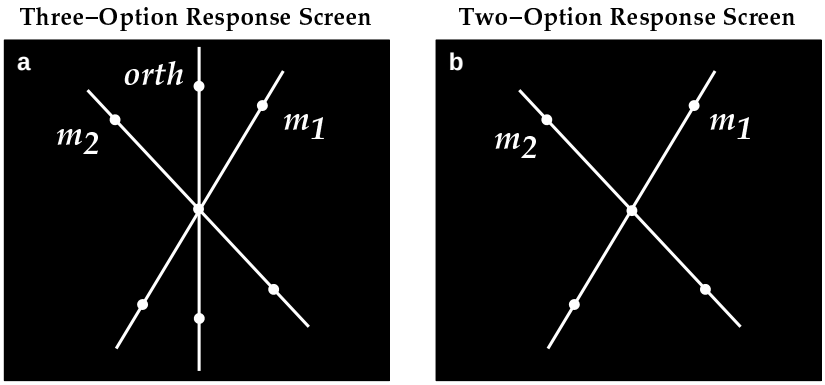

Figure 5.15 Two versions of the response screen—with two and three response options. Observers' task was two click on a circle attached to the radial line which is parallel to the perceived direction of apparent motion.

pattern of dots, which allowed us to set up a competition between alternative spatial organizations within a frame and ask whether temporal grouping affects spatial grouping.

As we saw in the first part of this chapter, spatial grouping in motion lattices is determined by the relative proximity between *concurrent* dots. If we hold temporal parameters constant, then temporal grouping depends on proximities between *successive* dots. According to the SM the propensity of dots to form virtual objects within frames, and thus yield g-motion, is independent of the determinants of temporal grouping. To test the SM, we ask whether the frequency of g-motion changes when we hold constant the relative proximity between concurrent dots and vary the proximity between successive dots.

The geometry of motion lattices is explained in Fig. 5.14. Fig. 5.14*a* shows three rows (baselines) of a motion lattice. The solid and open circles stand for dots that appear in frames f_1 and f_2, respectively. The three spatial parameters of motion lattices are: $|\mathbf{b}|$, the distance between adjacent dots in a baseline; $|\mathbf{m}_1|$, the shortest distance between successive dots; and θ_1, the acute angle between the orientations $\mathbf{b}$ and $\mathbf{m}_1$. The acute angle between $\mathbf{b}$ and $\mathbf{m}_2$ is θ_2. $\mathbf{S}$ is the second (after $\mathbf{b}$) shortest distance within a frame. To minimize edge effects we modulated the luminance of lattice dots according to a Gaussian distribution. We held $\mathbf{m}_1$ constant. Figs. 5.14*c–d,* show two successive frames of a motion lattice captured from the computer screen (not to scale).

4.6 GROUP MOTION VS. ELEMENT MOTION

Two kinds of AM are observed in motion lattices: *e*-motion, derived from element-to-element matching, and *g*-motion, derived from group-to-group matching.

In *e*-motion, each b_i dot is matched with either a b_{i-1} or a b_{i+1} dot. This yields motion in the direction parallel to the vector connecting the matched dots. The three shortest vectors between the successive dots are $\mathbf{m}_1$, $\mathbf{m}_2$ and $\mathbf{m}_3$, so that $|\mathbf{m}_1| \leq |\mathbf{m}_2| \leq |\mathbf{m}_3|$. The shorter distance between the dots, the more often they are linked in AM (von Schiller, 1933; Ramachandran and Anstis, 1983). Thus in motion lattices, when $|\mathbf{m}_1| < |\mathbf{m}_2| < |\mathbf{m}_3|$, motion parallel to $\mathbf{m}_1$ (m_1 motion) is perceived more frequently than parallel to $\mathbf{m}_2$ (m_2 motion), and never parallel to $\mathbf{m}_3$. (The strength of grouping between the successive elements is called *affinity* [Ullman, 1979]: The shorter the inter-dot distance, the higher affinity, the higher is the probability of the dots to be linked in AM.)

In *g*-motion, the perceived direction of motion is orthogonal to the strips of dots, the *virtual objects*. This type of motion cannot be explained by element-to-element matching, because it is clearly seen even when neither $\mathbf{m}_1$ nor $\mathbf{m}_2$ are orthogonal to the virtual objects. For *g*-motion to arise, the concurrent dots within each frame must group to form virtual objects. These virtual objects do not necessarily correspond to the baselines. When this happens, temporal grouping matches the virtual objects as wholes and yields AM in the direction orthogonal to their orientation.

In our experiments observers viewed a motion sequence and then they indicated the perceived orientation of apparent motion by clicking on circles attached to the lines on the response screen (Fig. 5.15, labels were not present in the display). The response screen consisted of two radial lines parallel to the most probable orientations of element-to-element apparent motion, $\mathbf{m}_1$ and $\mathbf{m}_2$, and (in some experiments) a third radial line orthogonal to the baselines, *orth*.

First we had to ascertain that changes in the frequency of *g*-motion are not attributable to observers' confusion between response options, because response confusion might mimic an interaction between the two types of grouping. This possibility arises because the "leakage" of responses between the *orth* responses and the responses (m_1, m_2) might increase, as the orientations of $\mathbf{m}_1$ and $\mathbf{m}_2$ vectors approach the orientation of *orth* response option. We did a control experiment with both two and three response options (Fig. 5.15), i.e., either with only m_1 and m_2 response options, or with m_1, m_2 and *orth* response options. Because we

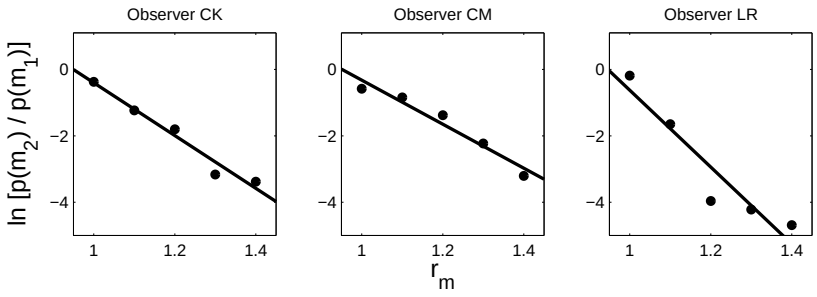

Figure 5.16 Affinity functions for three observers (each observer did 160 trials per condition). Each function shows a trade-off between two alternative e-motions, m_1 and m_2.

found only slight differences between the results of the two experiments, we will not refer further to the two-alternative experiment.

4.7 AFFINITY AND OBJECTHOOD

In Fig. 5.16 we plot the relative frequencies of the m_1 and m_2 responses as a function of the motion ratio, r_m for three observers. We call such functions *affinity functions*, by analogy to our attraction functions. Affinity is a measure of the propensity of the successive elements to group in apparent motion. It is analogous attraction, which is a measure of the propensity of concurrent elements to group and form virtual objects. In the left panel of Fig. 5.17 we plot the average data from Fig.

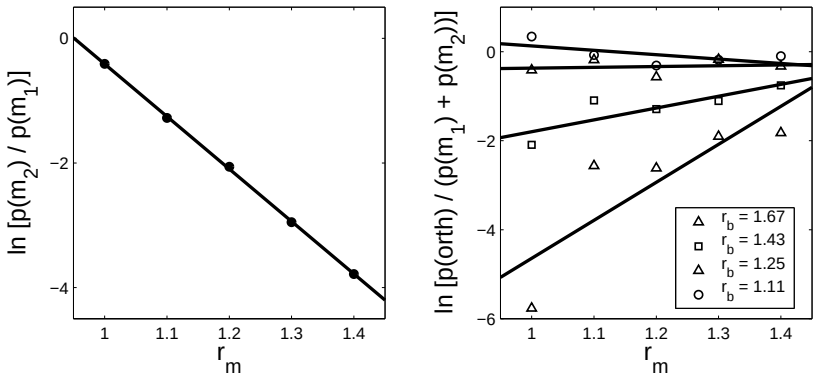

Figure 5.17 Affinity and objecthood functions (left and right plots, respectively) averaged across three observers from Fig. 5.16. The objecthood functions show the trade-off between two kinds of motion: e-motion, based on dot-to-dot matching (m_1 and m_2 responses), and g-motion, based on group-to-group matching (*orth* responses).

5.16. The affinity function is linear, much as our attraction functions are, suggesting that we are on the right track in our generalization of our work on static organization to dynamic organization.

In the right panel of Fig. 5.17 we plot the relative frequency of g-motion vs. e-motion. We refer to the functions on this plot as *object-hood functions* because they describe the trade-off between two kinds of matching—between virtual objects and between dots. The higher r_b and r_m, the more frequently observers reported e-motion at the expense of g-motion.

These data show that we can systematically manipulate the relative frequency of g-motion vs. e-motion. We have yet to answer a crucial question. Suppose we hold constant r_s, the parameter that controls static spatial grouping within each frame, and we vary other parameters of our display. Will we observe a variation in the tendency to see g-motion? If we do then we will have unequivocally demonstrated the validity of the IM using stimuli that do not trigger specialized processing. We will then be in a position to argue that the default model for *all* motion perception is an interactive one.

Acknowledgments

This research was supported by NEI grant No. R01 EY 12926–06. Correspondence regarding this article may be addressed to either of us, Department of Psychology, 102 Gilmer Hall, POBox 400400, The University of Virginia, Charlottesville, VA 22904–4400.

This chapter appears in: *Perceptual Organization for Artificial Vision Systems.* Eds. K. L. Boyer and S. Sarkar. pp. 41-71. Boston: Kluwer Academic Publishers. This is a pre–publication draft that differs in minor ways from the published version.

Notes

1. Imagine a consumer who would be equally satisfied with a market basket consisting of 4 lbs of meat and 2 lbs of potatoes and another consisting of 2 lbs of meat and 3 lbs of potatoes. In such a case, we say that the ⟨meat, potato⟩ pairs ⟨4, 2⟩ and ⟨2, 3⟩ are said to lie on an indifference curve.

References

Adelson, E. H. and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300:523–5.

Beck, J. (1966). Effect of orientation and shape similarity on perceptual grouping. *Perception & Psychophysics*, 1:300–302.

Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics*, 2:491–495.

Bell, R. and Bevan, W. (1968). Influence of anchors and the operation of certain gestalt organizing principles. *Journal of Experimental Psychology*, 78:670–678.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research*, 14:519–27.

Bravais, A. (1949). *On the systems formed by points regularly distributed on a plane or in space*. Crystallographic Society of America, N.p.

Brunswik, E. and Kamiya, J. (1953). Ecological cue-validity of "proximity" and of other gestalt factors. *American Journal of Psychology*, 66:20–32.

Burt, P. and Sperling, G. (1981). Time, distance, and feature tradeoffs in visual apparent motion. *Psychological Review*, 88:171–95.

Cavanagh, P. and Mather, G. (1990). Motion: The long and short of it. *Spatial Vision*, 4:103–29.

Chubb, C. and Sperling, G. (1989). Second-order motion perception: Space/time separable mechanisms. In *Proceedings: 1989 IEEE Workshop on Motion*, pages 126–38. IEEE Computer Society Press, Washington, DC.

Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., and Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics*, 57:761–77.

Croner, L. J. and Albright, T. D. (1997). Image segmentation enhances discrimination of motion in visual noise. *Vision Research*, 37:1415–27.

Croner, L. J. and Albright, T. D. (1999). Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. *Journal of Neuroscience*, 19:3935–51.

Derefeldt, G. (1991). Colour appearance systems. In Gouras, P., editor, *The perception of colour*, Vision and visual dysfunction, Vol. 6, pages 218–261. CRC Press, Boca Raton, FL.

Glass, L. and Switkes, E. (1976). Pattern perception in humans: Correlations which cannot be perceived. *Perception*, 5:67–72.

Graham, N. V. S. (1989). *Visual pattern analyzers*. Oxford University Press, New York.

Helmholtz, H. v. (1962). *Treatise on Physiological Optics*, volume III (Originally published in 1867). Dover Publications, New York.

Hildreth, E. C. (1983). *The Measurement of Visual Motion*. The MIT Press, Cambridge, MA.

Hochberg, J. and Hardy, D. (1960). Brightness and proximity factors in grouping. *Perceptual and Motor Skills*, 10:22.

Hochberg, J. and Silverstein, A. (1956). A quantitative index of stimulus-similarity: Proximity versus differences in brightness. *American Journal of Psychology*, 69:456–458.

Julesz, B. and Hesse, R. I. (1970). Inability to perceive the direction of rotation movement of line segments. *Nature*, 225:243–4.

Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. Praeger, New York.

Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics*, 6:193–8.

Kellman, P. J. and Cohen, M. H. (1984). Kinetic subjective contours. *Perception & Psychophysics*, 35:237–44.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, 72:194–296.

Kramer, P. and Yantis, S. (1997). Perceptual grouping in space and time: Evidence from the Ternus display. *Perception & Psychophysics*, 59:87–99.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement*, volume I: Additive and polynomial representations. Academic Press, New York.

Krechevsky, I. (1938). An experimental investigation of the principle of proximity in the visual perception of the rat. *Journal of Experimental Psychology*, 22:497–523.

Kubovy, M. (1994). The perceptual organization of dot lattices. *Psychonomic Bulletin & Review*, 1(2):182–190.

Kubovy, M., Holcombe, A. O., and Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, 35(1):71–98.

Kubovy, M. and Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative gestalt theory. *Psychological Science*, 6(4):225–234.

Lu, Z.-L. and Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, 35:2697–722.

Maunsell, J. H. R. and Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10:363–401.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86:287–330.

Michotte, A., Thinès, G., and Grabbé, G. (1964). Les compléments amodaux des structures perceptives. In *Studia Psychologica*. Publications Universitaires de Louvain, Louvain, Belgium.

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

Olson, R. K. and Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology*, 83:1–21.

Pantle, A. J. and Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science*, 193:500–2.

Prytulak, L. S. and Brodie, D. A. (1975). Effect of length, density, and angle between arms on gestalt grouping. *British Journal of Psychology*, 66:91–99.

Ramachandran, V. S. and Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature*, 304:529–31.

Ramachandran, V. S., Armel, C., and Foster, C. (1998). Object recognition can drive motion perception. *Nature*, 395:852–3.

Rock, I. (1981). Anorthoscopic perception. *Scientific American*, 244:145–53.

Rock, I. (1983). *The logic of perception*. The MIT Press, Cambridge, MA.

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

Shipley, T. F. and Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision*, 7:323–39.

Sigman, E. and Rock, I. (1974). Stroboscopic movement based on perceptual intelligence. *Perception*, 3:9–28.

Smith, A. T. (1994). The detection of second–order motion. In Smith, A. T. and Snowden, R. J., editors, *Visual Detection of Motion*, pages 145–76. Academic Press.

Stoner, G. R. and Albright, T. D. (1994). Visual motion integration: A neurophysiological and psychophysical perspective. In Smith, A. T. and Snowden, R. J., editors, *Visual Detection of Motion*, pages 253–90. Academic Press.

Ternus, J. (1936). The problem of phenomenal identity (originally published in 1926). In Ellis, W. D., editor, *A source book of Gestalt psychology*, pages 149–60. London: Routledge & Kegan Paul.

Tse, P., Cavanagh, P., and Nakayama, K. (1998). The role of parsing in high-level motion processing. In Watanabe, T., editor, *High-Level Motion Processing*, pages 249–66. The MIT Press, Cambridge, MA.

Ullman, S. (1979). *The Interpretation of Visual Motion*. The MIT Press, Cambridge, MA.

Uttal, W. R. (1981). *A taxonomy of visual processes*. Lawrence Erlbaum, Hillsdale, NJ.

von Schiller, P. (1933). Stroboskopische Alternativversuche. *Psychologische Forschung*, 17:179–214.

Wallach, H. (1935). Uber visuell wahrgenommene Bewegungsrichtung. *Psychologische Forschung*, 20:325–80.

Wallach, H. and O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–7.

Wallach, H., Weisz, A., and Adams, P. A. (1956). Circles and derived figures in rotation. *American Journal of Psychology*, 69:48–59.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung*, 4:301–350.

Wilson, H. R. (1994). Models of two-dimensional motion perception. In Smith, A. T. and Snowden, R. J., editors, *Visual Detection of Motion*, pages 219–51. Academic Press.

Wilson, H. R., Ferrera, V. P., and Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9:79–97.

Zucker, S. W., Stevens, K. A., and Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics*, 34:513–522.