M. Behrmann & R. Kimchi (Eds.),
Perceptual Organization in Vision: Behavioral and Neural Perspectives
Draft of May 20, 2001

# Grouping in space and in space-time: An exercise in phenomenological psychophysics

Michael Kubovy and Sergei Gepshtein
Department of Psychology
The University of Virginia

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.

Perceptual organization is difficult to study because it lies on the border between our experience of the world and unconscious perceptual processing. Even the term "perceptual organization" is ambiguous: it means both the *outcome* of perceptual processes—how things look—and the *mechanism* that produces it—the psychophysical processes that precede awareness. Perceptual organization is difficult to study for a second reason: because it involves both bottom-up and top-down processes. It is—like respiration—a semi-voluntary process. For example, when we first glance at a Necker cube, we are much more likely to see it as if it were seen from above than from below. Over this response we have no control; it is spontaneous and automatic. But as soon as we see the cube reverse the response appears in some measure to be under our voluntary control.

In this chapter we summarize our work on grouping in static and dynamic stimuli. In this work we make use of the fact that grouping occurs spontaneously and that it is multistable. At the same time we have taken care to minimize the effects of whatever voluntary control observers may have over what they see. We present methodologies that have allowed us to explore perceptual organization more rigorously than had hitherto been possible.

The first two sections of this chapter are devoted to empirical and theoretical studies of grouping. The third is metamethodological. Here is the reason. Our methods are *phenomenological*; they rely on the reports of observers about their phenomenal experiences. They also are *psychophysical:* they involve systematic exploration of stimulus spaces and quantitative representation of perceptual responses to variations in stimulus parameters. In short, we do *phenomenological psychophysics*. Because the observers' responses are based on phenomenal experiences, which are still in bad repute among psychologists, we fear that some may doubt the rigor of the research and seek other methods to supplant ours. So we conclude the chapter with an explication of the roots of such sceptical views, and show that they have limited validity.

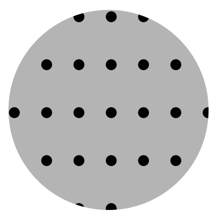
## Grouping by proximity in space

The Gestalt psychologists' accounts of grouping were vague and qualitative. This need not be the case. When one pays attention to demonstrations of grouping, one becomes aware of the differential strength of certain effects. For example, in Figure 1(a) one spontaneously sees horizontal grouping. In Figure 1(b) one can also see horizontal grouping, but with some difficulty. The tendency to see horizontal grouping is *weaker* in Figure 1(b) than in Figure 1(a). Such observations are the seed of a quantitative theory.

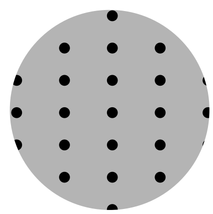
Over a quarter of a century elapsed between Wertheimer's formulation of the grouping principles and the emergence of the idea that the strength of grouping might be measurable. Hochberg and his associates—who were pioneers in this matter—thought that the only way to measure the strength of grouping by proximity was to pit it against the strength of grouping based on another principle, such as similarity. They used $6 \times 6$ rectangular lattices of squares (Hochberg & Silverstein, 1956) or $4 \times 4$ rectangular lattices of dots (Hochberg & 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 (Figure 1(d)) until they found the spacing for which their tendency to see rows and columns was in equilibrium. Using this method, Hochberg and Hardy (1960) plotted what microeconomists call an *indifference curve* (Krantz, Luce, Suppes, & Tversky, 1971).[1] When Hochberg reduced the lu-

Supported by NEI grant R01 EY 12926-06. Correspondence: Department of Psychology, PO Box 400400, University of Virginia, Charlottesville, VA 22904-4400 (e-mail: kubovy@virginia.edu). SG is presently in the School of Optometry, University of California at Berkeley (e-mail: sergeig@uclink.berkeley.edu).

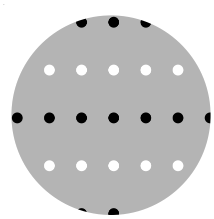
[1] Imagine a consumer who would be equally satisfied with a market basket consisting of 1 lb of meat and 4 lbs of potatoes and another consisting of 2 lbs of meat and 1 lb of potatoes. In such a case, we say that the ⟨meat, potato⟩ pairs ⟨1, 4⟩ and ⟨2, 1⟩ are said to lie on an indifference curve.



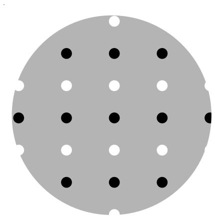
(a) Horizontal grouping by proximity



(b) Vertical grouping by proximity



(c) Proximity and similarity in concert



(d) Proximity and similarity in opposition

*Figure 1.* Examples of grouping by proximity and of the interaction of grouping by proximity and similarity

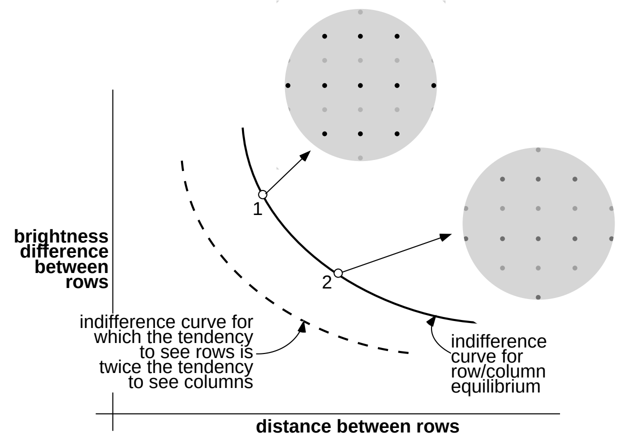



*Figure 2.* Two grouping indifference curves. Only the solid curve is achievable by methods such as Hochberg's (illustrated here for the trade-off between grouping by proximity and grouping by similarity) and Burt and Sperling (1981) (for the trade-off between grouping by spatial proximity and grouping by temporal proximity). Our method allows us to plot a *family* of indifference curves.

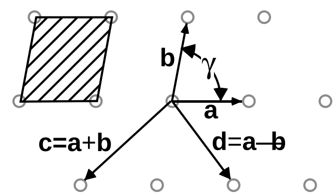



*Figure 3.* The features of a dot lattice (see text).

minance difference between the rows, the distance between rows for which observers reported an equilibrium between rows and columns increased (Figure 2). We call this is a *grouping indifference curve* because the observer is indifferent among the ⟨luminance-difference, row-distance⟩ pairs that lie on it: they are all in equilibrium.

Unfortunately, this method can give us only one indifference curve: the equilibrium indifference curve. We cannot tell where to place a grouping indifference curve for which all ⟨luminance difference,distance pairs⟩ are such that the tendency to see rows is twice as strong as the tendency to see columns (dashed curve in Figure 2).

Can we measure the strength of grouping by proximity without reference to another principle of grouping? We have found that if we use a suitable class of stimuli, we can.

## *Generalizing the Gestalt lattice*

The suitable class of stimuli is *dot lattices* (Figures 1(a) and 1(b)). These are arrays of dots similar to those used by the Gestalt psychologists in their classic demonstrations. In most previous demonstrations and experiments, such arrays have been rectangular, with one direction vertical. Our dot lattices differ in two ways: (1) The two principal directions of grouping are not always perpendicular, and (2) neither principal orientation of the lattice need be vertical or horizontal.

A dot lattice is an infinite collection of dots in the plane. It is characterized by two (nonparallel) translations, represented by vectors **a** and **b** (Figure ). The idea of the two translations can be understood as follows. Suppose you copied the lattice onto a transparent sheet, which was overlaid on top of the original lattice, so that the dots of the overlay were in register with the dots of the original lattice. You could pick up the overlay and shift it in either the direction **a** by any multiple of the length of **a**, $|\mathbf{a}|$, and the dots of the overlay would once again be in register with the dots of the original lattice. The same is true of **b**. In other words, translating the lattice by **a** or **b** leave it unchanged, invariant. Operations that leave a figure unchanged are called *symmetries* of the figure. Therefore these two translations are symmetries of the lattice.

The two translation vectors **a** and **b** are not the only ones that leave the lattice invariant. In addition, the vector difference of **a** and **b**, $\mathbf{a}-\mathbf{b}$ (which we denote **c**, Figure ) and the vector sum of **a** and **b**, $\mathbf{a}+\mathbf{b}$ (which we denote **d**) are also symmetries of the lattice.[2] Any dot in the lattice has eight

[2] There is an infinity of others, but they need not concern us here.

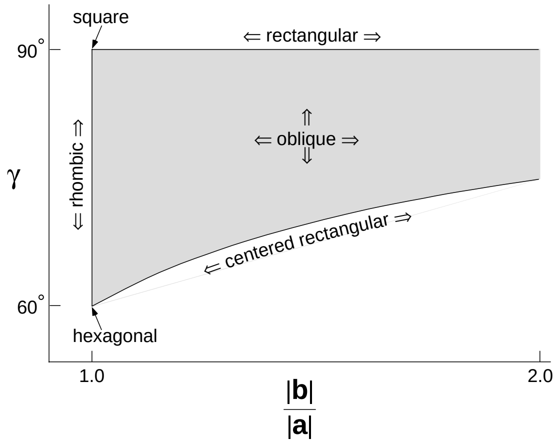



*Figure 4*. The space of dot lattices. See Figure 5.

neighbors. Its distance from a neighbor is either **a**, **b**, **c**, or **d**.

Another way to look at dot lattices is to consider its *basic parallelogram*, *ABCD* in Figure . It is the building block of the dot lattice. Its sides are **a** and **b**; its diagonals are **c** and **d**.

Any lattice can be defined by specifying three parameters: two distances, $|\mathbf{a}|$ and $|\mathbf{b}|$, and the angle between the two vectors, $\gamma$. If we do not care about the scale of a lattice, and are concerned only with its shape, only two parameters are needed: $^{|\mathbf{b}|}/_{|\mathbf{a}|}$ and $\gamma$.[3]

Before we proceed, we mention three constraints on dot lattices:

1. The following relation between vector lengths holds: $|\mathbf{a}| \leq |\mathbf{b}| \leq |\mathbf{c}| \leq |\mathbf{d}|$,

2. For geometric reasons, $60° \leq \gamma \leq 90°$ (Kubovy, 1994).

The two-parameter space of lattices is depicted in Figure 4. This space can be partitioned into six classes, whose names appear in Figure 4. The differences among these classes are portrayed in Figure 5, where each class occupies a column.[4] In the top row of each column is the name of the lattice class. In the second row we show a sample lattice. In third row we show the basic parallelogram of the lattice. In the fourth row we compare the lengths of the four vectors. A dashed line connecting two bars means that they are of the same length. In the fifth row we depict the important properties of the lattice which determine its symmetries. We spell out these properties symbolically in the lines of text at the bottom of each column. Two of these classes consist of just one lattice: the hexagonal lattice and the square lattice. The left to right order of the lattice classes in Figure 5 is determined by their expected degree of ambiguity:[5]

**Oblique:** No two vectors are of equal length. Therefore these lattices have only two symmetries (disregarding the identity): the two translations.

**Rectangular:** Because $|\mathbf{c}| = |\mathbf{d}|$, lattices in this class have three more symmetries than oblique lattices: two mirrors (one that bisects **a** and one that bisects **b**) and a rotation of 180° (also known as *twofold rotational symmetry*, or a *half-turn*).

**Centered rectangular:** Because $|\mathbf{b}| = |\mathbf{c}|$, you can always draw a rectangle that “skips” a row and a column (such as *BDEF*). This means that lattices in this class have two additional symmetries, called *glide reflections*. Imagine a horizontal axis between two adjacent rows of the lattice. Now reflect the entire lattice around this axis, *while translating* it over a distance $|\mathbf{a}|/2$. This transformation is similar to the relation between the right and left footprints made by a person walking on wet sand. There is also a vertical glide reflection in these lattices.

**Rhombic:** The symmetries of this class of lattices are the same as those of centered rectangular lattices. Nevertheless these symmetries are more salient because $|\mathbf{a}| = |\mathbf{b}|$.

**Square:** Here we have two equalities: $|\mathbf{a}| = |\mathbf{b}|$ and $|\mathbf{c}| = |\mathbf{d}|$. These add two mirrors along the diagonals of the basic parallelogram, and fourfold rotational symmetry instead of the twofold rotational symmetry that the preceding classes inherited from the rectangular lattice.

**Hexagonal:** Here we have a triple equality: $|\mathbf{a}| = |\mathbf{b}| = |\mathbf{c}|$. This means that this lattice has six mirrors and sixfold rotational symmetry.

The phenomenology of each lattice is determined by the symmetries we have just described. We might expect that the more symmetries a lattice has, the more unstable it is. We discuss this idea in the next section.

## *The instability of lattices*

Kubovy and Wagemans (1995) conducted an experiment to measure the instability of grouping in dot lattices, which amounts to measuring their ambiguity. On each trial they presented one of sixteen dot lattices (Figure 6), sampled systematically from the space shown in Figure 4 (Figure ). The screen was divided into two regions, the aperture and the black mask around it (Figure 8(a)). The lattices, which consisted of a large number of yellow dots, were visible in the blue region of the screen only. Observers saw each lattice in a random orientation, for 300 ms. They were told that each lattice could be perceived as a collection of parallel 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

[3] Let $|\mathbf{a}| = 1.0$. Then $|\mathbf{c}| = \sqrt{1+|\mathbf{b}|^2 - 2|\mathbf{b}|\cos\gamma}$ and $|\mathbf{d}| = \sqrt{1+|\mathbf{b}|^2+2|\mathbf{b}|\cos\gamma}$.

[4] Bravais (1866/1949), the father of mathematical crystallography, found five classes: according to his scheme, centered rectangular and rhombic lattices belong to the same class. The taxonomy proposed by Kubovy (1994) has six classes because he did not only consider the symmetries of dot lattices (as did Bravais), but also their metric properties.

[5] The reader who wishes to know more about the mathematics of patterns would do well to consult Martin (1982) and Grünbaum and Shepard (1987).

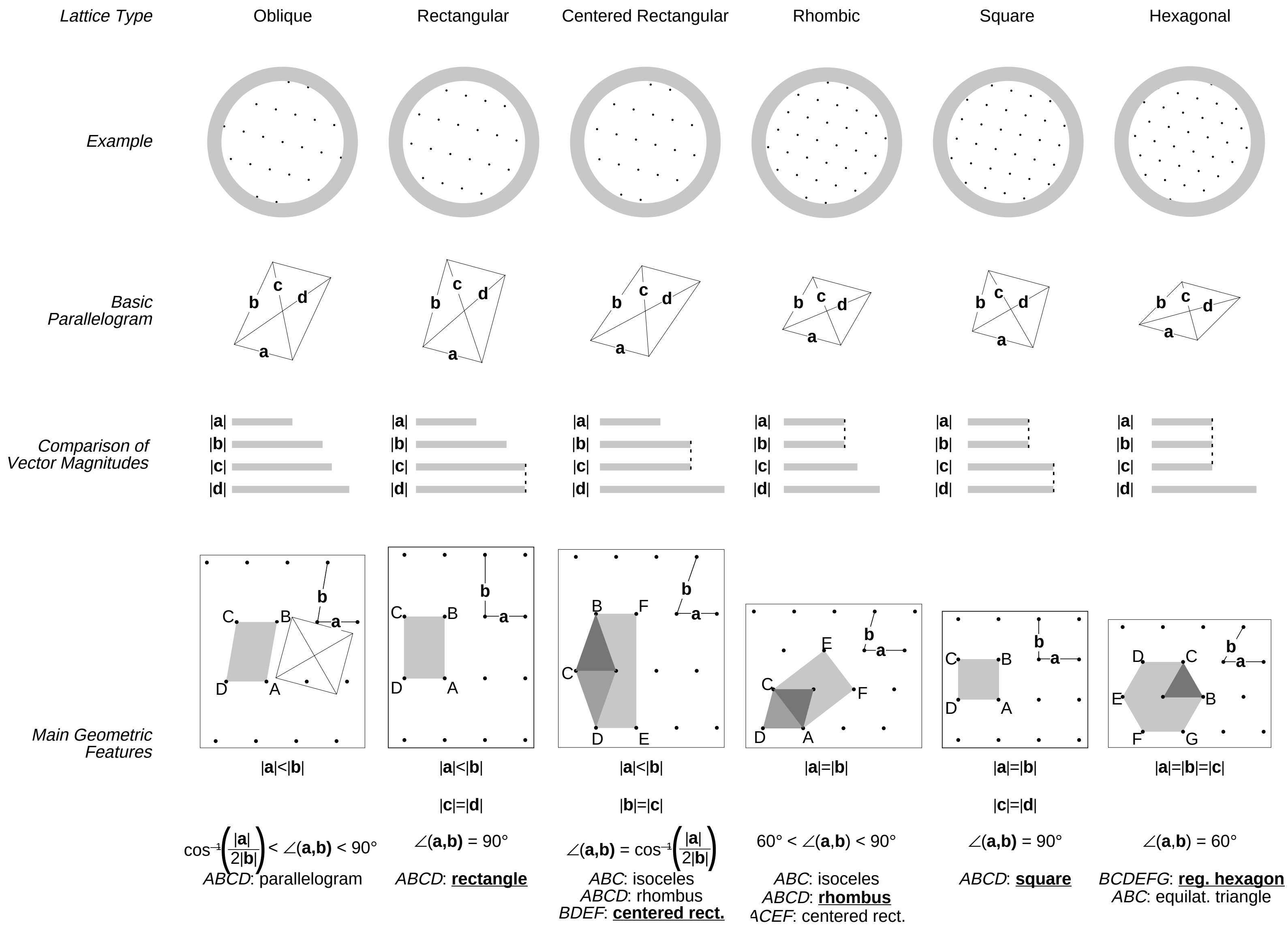



*Figure 5.* The six classes of dot lattice according to Kubovy (1994).

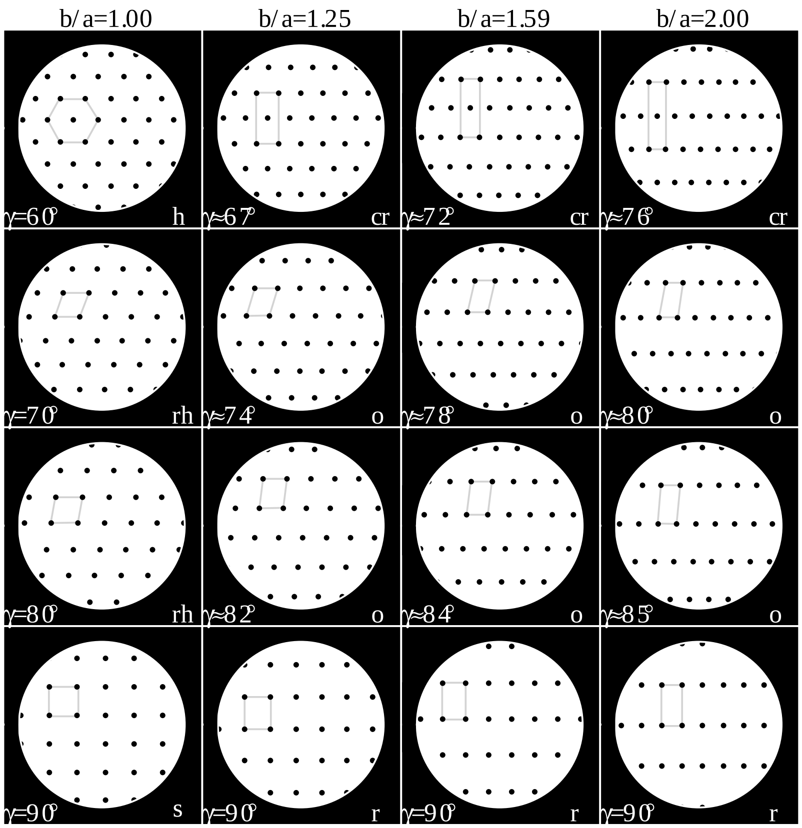



*Figure 6*. The sixteen dot lattices used by Kubovy & Wagemans (1995): *h*—hexagonal; *cr*—centered rectangular; *s*—square; *r*—rectangular; *o*—oblique. At the top of the figure, the $|\mathbf{b}|/|\mathbf{a}|$ ratio. In the lower left-hand corner of each panel, the value of $\gamma$.

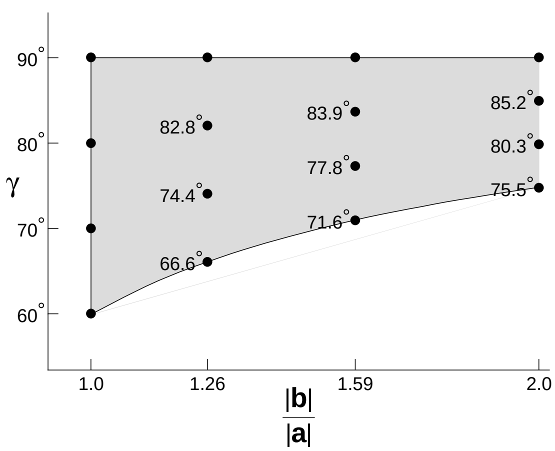
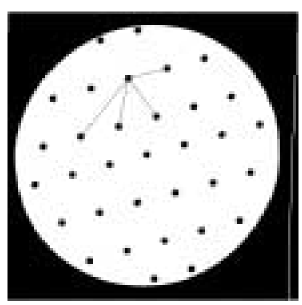
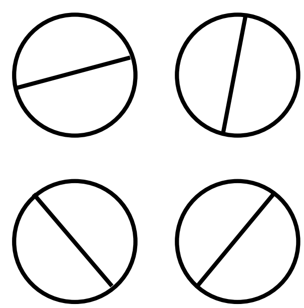



*Figure 7*. How the sixteen stimuli were sampled from the space of dot lattices. See Figure 4.



(a) A stimulus with four vectors



(b) The four response alternatives. Clockwise, from upper left: $a, b, d, c$.

*Figure 8*. The Kubovy & Wagemans (1995) experiment.

selecting one of four icons on the response screen (Figure 8(b)). Each icon consisted of a circle and one of its diameters. The orientation of the diameter corresponded to the orientation of one of the four vectors of the lattice just presented. Because the task involved a four-alternative forced-choice (4AFC) but *no incorrect response*, it is an example of experimental phenomenology (see p. 18).

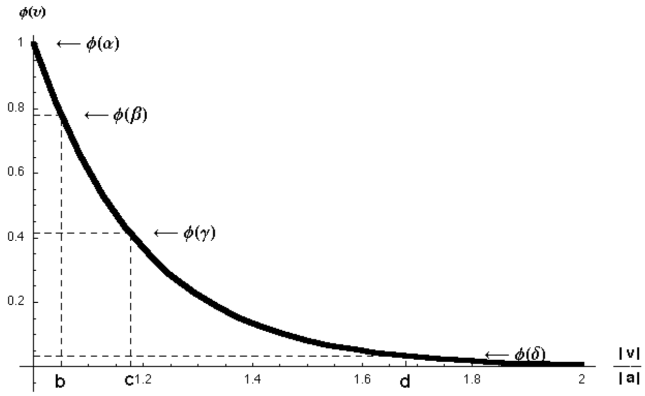



*Figure 9.* Grouping strength in the theory proposed by Kubovy & Wagemans (1995). Here we have set the slope of the grouping strength function to $s = 5$. We also illustrate how one calculates the four grouping strengths for an oblique lattice in which $^{|\mathbf{b}|}/_{|\mathbf{a}|} = 1.05$ and $\gamma = 70°$. To indicate the inter-dot distances in this lattice we have placed the letters 'b' , 'c', and 'd' at $^{|\mathbf{b}|}/_{|\mathbf{a}|} = 1.05$, $^{|\mathbf{c}|}/_{|\mathbf{a}|} \approx 1.18$ and $^{|\mathbf{c}|}/_{|\mathbf{a}|} \approx 1.68$. The corresponding values of $\phi$ are $\phi(\beta) \approx 0.78$, $\phi(\gamma) \approx 0.41$, and $\phi(\delta) \approx 0.034$ ($\phi(\alpha)$ is always 1.0).

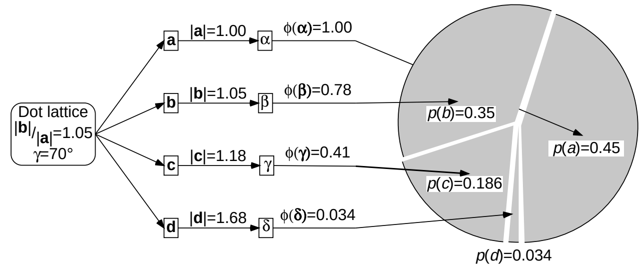



*Figure 10.* How $p(v)$ is computed, according to the model of Kubovy & Wagemans (1995). The parameters are the same as those in Figure 9.

*Some theory*.

Kubovy and Wagemans wanted to better understand the nature of Gestalts. They chose to formulate the least Gestalt-like model they could and see where the data deviated from their predictions. We will develop these ideas in a way that differs somewhat from the presentation in Kubovy and Wagemans (1995).

Suppose that grouping is the product of interpolation mechanisms, and suppose that the visual system provides a number of independent orientation-tuned interpolation devices (OTID). Let us suppose that the $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and $\mathbf{d}$ vectors in the lattice excite four of these devices—$\alpha, \beta, \gamma$, and $\delta$—and that the others remain quiescent. The activated OTIDs will produce outputs that depend on the distance between dots in the four directions. We we call these outputs *grouping strengths*, and we label them $\phi(\alpha), \phi(\beta), \phi(\gamma), \phi(\delta)$. To make this function independent of scale, we use *relative* rather than absolute inter-dot distances, e.g., $^{|\mathbf{b}|}/_{|\mathbf{a}|}$ (where $|\mathbf{a}|$ is the shortest distance between dots), rather than $|\mathbf{b}|$.

**Grouping strength:** If $\mathbf{v}$ is a general element of the set of lattice vectors, $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$, and $\upsilon$ is a general element of the set of OTIDs, $\{\alpha, \beta, \gamma, \delta\}$, then

$$\phi(\upsilon) = e^{-s\left(\frac{|\mathbf{v}|}{|\mathbf{a}|} - 1\right)}. \tag{1}$$

This means that grouping strength is a decaying exponential function of the distance between dots in the direction parallel to $\mathbf{v}$, $|\mathbf{v}|$, *relative to* the shortest distance between dots, $|\mathbf{a}|$. The computation of $\phi(\upsilon)$ is illustrated in Figure 9.

**Choice probability:** The four OTIDs are active concurrently, but the observer sees only one organization because the lattice is multistable. So we must distinguish overt responses from her internal states by using italic characters to refer to responses (i.e., $v$ represents the observer' indicating that the lattice appears organized into strips parallel to $\mathbf{v}$. Following Luce (1959) we assume that grouping strength is a ratio scale that determines the probability of choosing $v$, $p(v)$, in a simple way:

$$p(v) = \frac{\phi(\upsilon)}{\phi(\alpha) + \phi(\beta) + \phi(\gamma) + \phi(\delta)}. \tag{2}$$

The computation of $p(v)$ is illustrated in Figure 10.

*Entropy*.

Having proposed their model of grouping, Kubovy and Wagemans (1995) were in a position to predict the instability of the organization of any dot lattice. To test this prediction they used the model to calculate the expected entropy (also known as the average uncertainty) of the responses to each lattice (Garner, 1962). The reader will recall that the entropy of a discrete random variable $\mathbf{x}$, with sample space $X = \{x_1, \ldots, x_N\}$ and probability measure $P(x_n) = p_n$, is

$$H(\mathbf{x}) = -\sum_{n=1}^{N} p_n \log(p_n).$$

If the base of the logarithms is 2, the entropy is measured in *bits* (binary digits). Turning now to the predicted entropy of dot lattices, we have:

$$H = -\sum_{w \in W} p(w) \log_2 p(w),$$

where $W = \{a, b, c, d\}$. These predictions are shown in Figure 11. The results (Figure ), were encouraging, but not entirely satisfactory. The model underestimated the amount of entropy in the responses to the most unstable lattices (i.e.,, those with the highest predicted entropy). That is one reason why Kubovy, Holcombe, and Wagemans (1998a) revisited these data.

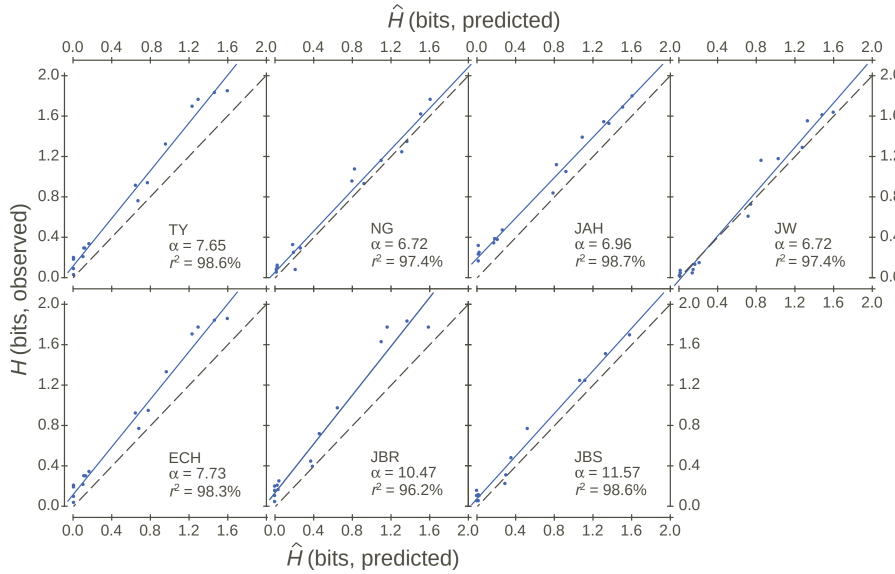



*Figure 12*. Observed entropy of responses as a function of entropy predicted by Kubovy& Wagemans's model (1995) for seven observers. (Figure after Kubovy & Wagemans (1995), Figure 11.)

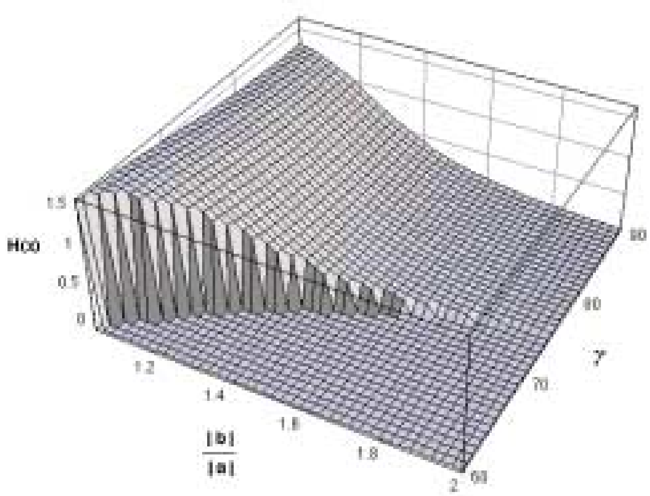


*Figure 11*. Entropy of responses as a function of $^{|\mathbf{b}|}/_{|\mathbf{a}|}$ and $\gamma$, as predicted by Kubovy & Wagemans (1995). The $x$ and $y$ axes are the same as in Figure 4. Note that where entropy is not defined (outside the curved boundary of the space of dot lattices) we let $H(x) = 0$. The value of $s$ $(= 5)$ is the same as in Figure 9.

## *The pure distance law*

Kubovy et al. (1998a) did not merely reanalyze the Kubovy and Wagemans (1995) data in order to improve the model, but to address a fundamental question. Did the data deviate from the model because the anti-Gestalt assumptions of the model were false? Did they contain a clue to an interesting Getsaltish interaction?

The data collected by Kubovy and Wagemans (1995) were ideally suited to answering these questions. The stimuli had been sampled (Figure ) so that each type of lattice was represented multiple times (except for the hexagonal and the square, of course, which are points in the space of lattices, Figure 4). As the reader will recall (see Figure 5), the different classes of lattices have different symmetries, and therefore have the potential to be organized differently. Kubovy et al. (1998a) reasoned that if they could show that the probability of choosing a vector $v$ depended on $\gamma$, or on the identity of the vector $(\mathbf{a}, \mathbf{b}, \mathbf{c}$, or $\mathbf{d})$, perhaps these dependencies would lead to a formulation of the Gestalt component of grouping in lattices.

They first noted that the data uses four probabilities—$p(a)$, $p(b)$, $p(c)$, and $p(d)$. Because there are only three degrees of freedom in these data they reduced them to three dependent variables: $^{p(b)}/_{p(a)}$, $^{p(c)}/_{p(a)}$, and $^{p(d)}/_{p(a)}$ (or $^{p(v)}/_{p(a)}$ for short). In addition, because in the data the range of these probability ratios was large, they used $\log[^{p(v)}/_{p(a)}]$ as their dependent variable(s).

The intricacies of the analyses conducted by Kubovy et al. (1998a) are beyond the scope of this article. To make a long story short, they faced two problems, both of which were most severe when $|\mathbf{b}|$ was large: low probabilities for responses $c$ and $d$, and larger probablities for response $d$ than for response $c$. There was little they could do about the first problem.[6] They were able to remedy the second problem, however, which was an unforseen consequence of the geometry of lattices. If one holds $|\mathbf{a}|$ and $\gamma$ constant and one increases the length of $\mathbf{b}$, the angle between $\mathbf{b}$ and $\mathbf{d}$ decreases. Thus the likelihood that an observer will respond $d$ when she

[6] We have since settled on a rule of thumb: not to use dot lattices in which $^{p(b)}/_{p(a)} \geq 1.5$

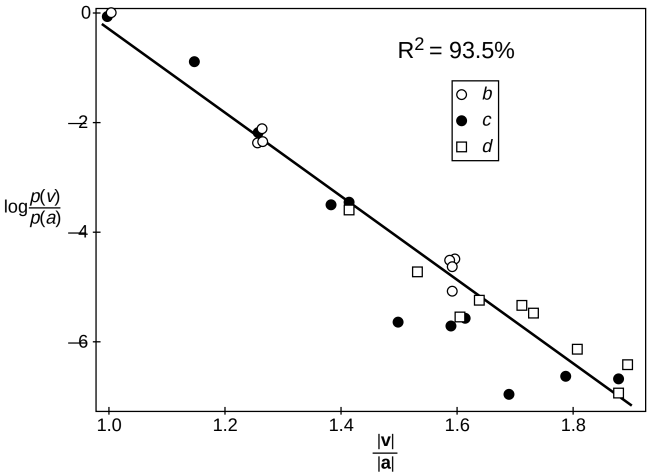



*Figure 13*. The pure distance law for dot lattices. Average data for the seven observers.

intended to choose $b$ increases with $|\mathbf{b}|$. (Of course the observer will also respond $b$ when she intended to choose $d$, but these are rare cases.) By carrying out an auxiliary experiment they were able to estimate the probability of this confusion and to develop a multinomial model that corrected for these errors.

Figure  shows the results. This linear function, which we call the *attraction function*, whose slope is $s = 7.63$, accounts for 93.5% 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 $a$. The fact that all these observation fall on the same linear function supports our theory, and shows that the probability of choosing a vector $v$ does not depend on $\gamma$, or on the identity of the vector ($\mathbf{a}, \mathbf{b}, \mathbf{c}$, or $\mathbf{d}$). In other words, we have a *Pure Distance Law*. This is a quantitative law of grouping by proximity, which states that grouping follows a decaying exponential function of relative inter-dot distances. We refer to this empirical relationship as a law, because our evidence implies that it holds for all vectors in all possible dot lattices.

## *Where is the Gestalt?*

The pure distance law is as simple a law of grouping as we can imagine. The lattice activates four independent units whose output jointly determines the probability of a winner-take-all percept (we see only one organization at a time). Where is the Gestalt here? One might consider the fact that in the Kubovy-Wagemans model, all spatial distances are scaled relative to the shortest distance, $|\mathbf{a}|$. But this is no more a Gestalt phenomenon than any other perceptual scale-invariance or size-constancy.

The only Gestalt-like feature of grouping that remains is that it is a collective property. In other words, it is a property that requires a certain number of elements (perhaps at least a $4 \times 4$ lattice) to emerge.

# Grouping by proximity in spacetime

In the preceding section we saw that grouping by spatial proximity appears to be decomposable into separable mechanisms, and that nothing in our theory resembles the kind of complex dynamical system the Gestalt psychologists thought necessary. In this section we turn to the problem of grouping in space-time through the study of apparent motion, which is the prototypical Gestalt phenomenon (Wertheimer, 1912). Our starting point is similar in spirit to the approach taken by Kubovy and Wagemans. We ask, can we decompose apparent motion into two separable components: grouping by spatial proximity and grouping by spatiotemporal proximity? We first present some theory, and then turn to the work of Gepshtein and Kubovy (2000) for the answer.

## *A clarification of terms*

*Grouping by spatial proximity* is the process by which small, temporally concurrent, discrete elements of a scene (❶ and ❷ in Figure A) are perceptually linked across space to form larger and more complex perceived entities, culminating in the perception of objects and surfaces. This is the process we analyzed in the preceding section. Now suppose ❶ appears in frame 1 and ❷ appears in frame 2. To perceive ❶ and ❷ as a single element in motion (Figure B), vision must link them across space *and* across time, a process we call *grouping by spatiotemporal proximity*. Consider a slightly more complicated display, which consists of two elements in frame 1 (❶ and ❷) and two elements in frame 2 (❸ and ❹). When motion is seen, one of two things happens: [i] ❶ may be linked with ❹, and ❷ may be linked with ❸ so that elements move independently of each other (Figure C[i]); or [ii] the elements may form a grouping and be seen as a moving object (❶❷$\longrightarrow$❸❹, Figure C[ii]). In the latter case, the motion of the elements is the same as the motion of the object (Figure C[iii]). In this process the visual system establishes a *correspondence* between elements visible at successive instants, which is why grouping by spatiotemporal proximity is also known as *matching*. The successive visual entities that undergo grouping by spatiotemporal proximity are called *matching units* or *correspondence tokens* (Ullman, 1979).

*Sequential and interactive models*.

What is the relation between grouping by spatial proximity and grouping by spatiotemporal proximity? We consider two models:

**Sequential model (SM):** According to this model grouping by spatial proximity and grouping by spatiotemporal proximity are separable and serial, so that matching units are determined by their spatial proximity, independent of their spatiotemporal proximity. We will see that this model can account for many of the phenomena of motion perception and is often taken, tacitly, as the default model.

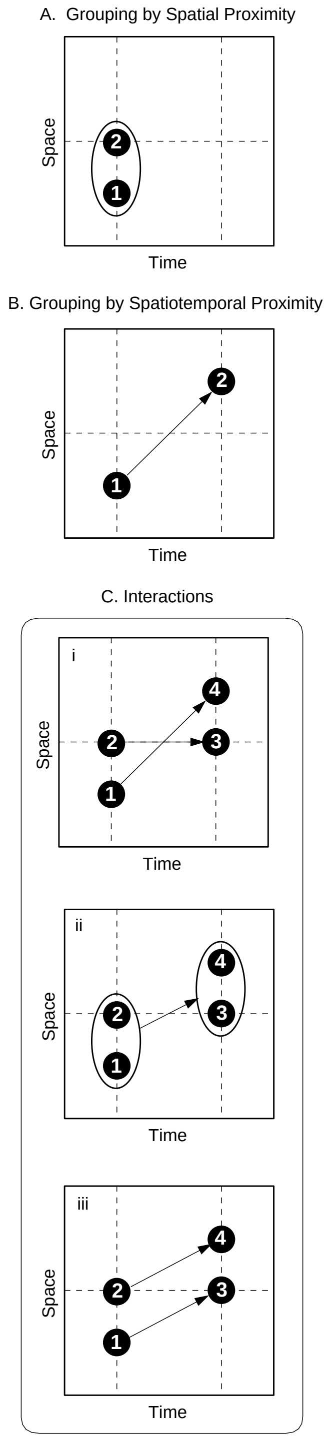



*Figure 14.* (A) Grouping by spatial proximity, (B) grouping by spatiotemporal proximity, and (C) their interactions. The configurations in C[i] and [ii] were introduced by Ternus (1936) who described two kinds of percepts: [i] *element-motion*: ❶⟶❹ and ❷⟶❸; and [ii] *group-motion*: ❶❷⟶❸❹. [iii] Ullman (1979) argued that what looks like group-motion may actually be element-motion ❶⟶❸ and ❷⟶❹ (see also Figure ).

**Interactive model (IM):** According to this model grouping by spatial proximity and grouping by spatiotemporal proximity are inseparable, so that grouping by spatiotemporal proximity can override the process of grouping by spatial proximity and matching units are derived by the combined operation of both processes.

The purpose of this section is to argue that the correct model of motion perception is the interactive model.

We follow McClelland (1979) in assuming that the visual system constructs a “cascade” of representations, a hierarchical set of spatial representations of a scene. In such a cascade, (a) more complex representations may contain entities from less complex ones, (b) and each representation may interact with both more complex and less complex representations. This view assumes that these alternative representations emerge concurrently, and can be accessed in parallel, as soon as they become available.

In the SM the mechanisms of temporal grouping can access alternative spatial representations in parallel, so that the most salient spatial organization becomes a matching unit. On this view, matching units can be thought of as “sliders” on the cascade of spatial perceptual organization because spatial entities of arbitrary complexity can serve as matching units. The most salient spatial organization determines *what* is seen to move.

The IM is different from the SM in that it implies 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 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. (We will present the IM in more detail later.)

Neither the SM nor the IM imply an ordering of the two kinds of grouping (by spatial proximity and by spatiotemporal proximity). According to one version of the SM (Neisser, 1967), motion perception is an integration of successive “snapshots” of the scene: grouping by spatial proximity alone determines which visual elements undergo motion matching. This cannot, however, be the whole story: motion matching may precede grouping by spatial proximity. Take random-dot cinematograms. In these dynamic displays each frame contains a different random and shapeless texture. If we correlate the frames, so that a compact set of elements, $f$, retains its texture across frames, and we change its location across the frames (while the remaining dots change their positions at random), each frame of the display will still look like a shapeles random texture. But when we show the sequence of frames we see $f$ segregated from the rest of the display. This could happen only if motion matching occurs before grouping by spatial proximity. The Gestalt psychologists referred to such organization as grouping by *common fate* (Wertheimer, 1923).

Before we embark on a discussion of which model is more adequate, we will consider the place of the SM and the IM in the theories of motion perception, and then turn to the evidence that supports the SM. We will review some parade examples of the relations between the two kinds of grouping and argue that the SM is able to explain a surprisingly wide spectrum of phenomena in motion perception. Only then, we will look for critical evidence in favor of one of the models.

## *Two kinds of 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 the perception of motion (Lu & Sperling, 1995), based on these alternative spatial representations:

1. 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 grouping by spatiotemporal proximity applied directly to the raw visual input.

2. 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.

3. 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 the two kinds of grouping seem to operate sequentially within these systems (see, e.g., Chubb & Sperling, 1989) the three-system conception of motion perception as a whole is mute on separability of the two kinds of grouping and therefore can be adapted to agree with either the SM or the IM. 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 (Figure A), or (b) the competition could occur after motion computation for each level in the cascade of spatial representations (Figure B). 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 grouping by spatiotemporal proximity will determine which level of spatial organization will dominate perception.

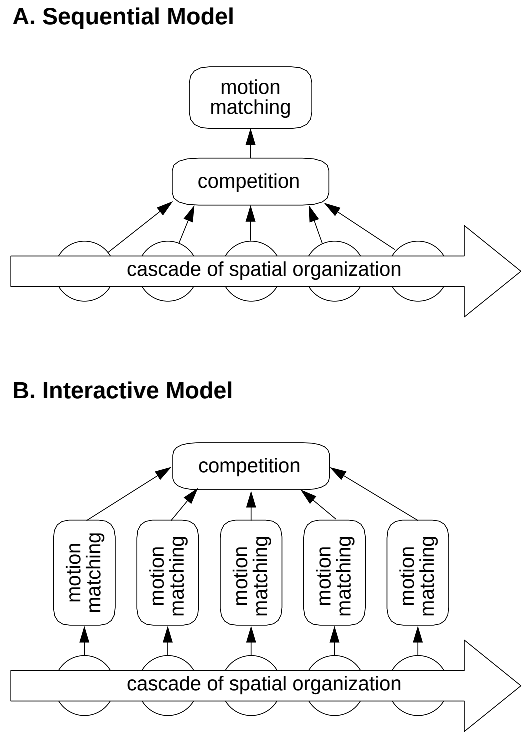



*Figure 15*. A. In the sequential model (SM), alternative spatial representations compete, so that the most salient one undergoes temporal grouping. Here, spatial grouping alone determines *what* is moving (see also Figure ). B. In the interactive model (IM), the competition occurs between the outputs of motion matching operations, so that both spatial and temporal grouping determine the spatial complexity of matching units. (The horizontal arrows in A and B correspond to the direction of increasing complexity of spatial organization in the cascade.)

*An explicit interactive model.*

In the literature on motion perception the IM is admittedly uncommon and, hence, less familiar to the researchers. To further clarify the distinction between the SM and the IM, consider an explicit example of the IM.

Wilson, Ferrera, and Yo (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 model is interactive because the competition between alternative spatial representations (or, we can say, between different spatial identities) takes place after motion matching within every representation (Figure B).

The model of Wilson et al. (1992) is truly interactive, but it was inspired by evidence which is also consistent with the SM:

• Physiological evidence that showed that the cortical areas responsible for motion perception receive both first- and second-order spatial information (i.e., Maunsell & New-

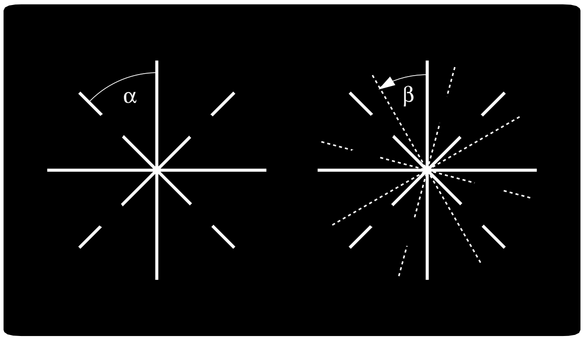



*Figure 16*. The "broken wagon wheel" demonstration of Ullman (1979). LEFT: Each other spoke of the "wagon wheel" is interrupted in the middle. The angle between the neighboring spokes is α. RIGHT: When the "wheel" is rotated anticlockwise by angle $\beta > \frac{\alpha}{2}$ one sees three rotating objects: 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.

some, 1987). The two inputs could compete before motion matching operation.

- Psychophysical evidence of motion integration that was launched by a seminal study of Adelson and Movshon (Adelson & 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. 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 & Albright, 1994 for a different view.)

## *The sequential model and matching units*

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 consider several examples that are compatible with the SM, for which the matching units are spatial organizations of increasing complexity. First, we will see how one can apply Ullman's model to displays that do not consist of small-scale elements. Then, we will see that the assumption of small-scale units does not work.

*The SM and Ullman's theory of matching units.*

Our first illustration concerns displays that contain one or more continuous lines, rather than a collection of small-scale discrete elements (Figure ). Whenever a line is moving behind an aperture that occludes its endpoints, we see motion orthogonal to the line. This observation raises two problems:

1. Any segment on the line at time $t_i$ may match any segment on the line at time $t_{i+1}$. This is the *correspondence problem*.

2. The size of these segments is unknown. This is the *matching unit problem*.

These two problems are parts of what has become known as the *aperture problem* (Wallach, 1935; Wallach & O'Connell, 1953; Hildreth, 1983). If we follow 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 (e.g., Figure ).

A similar analysis applies to other displays. For example, Wallach, Weisz, and Adams (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 (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 because they do not assume a spatial grouping process other than the processes that locate the spatial primitives which become matching units. Hence, on their view temporal grouping can have no influence on spatial grouping.

*Recursive grouping.*

Matching units can be be derived by the grouping by spatial proximity of entities which in turn are derived by grouping by spatiotemporal proximity. In such cases we are talking of hierarchical perceptual organization, where elements move within objects that are themselves moving.

Consider *grouping by common fate* (Wertheimer, 1923), which occurs when elements extracted by grouping by spatiotemporal proximity are segregated from the background and form a moving figure, when they move in a similar direction.[7] The elements extracted by this process 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 grouping by spatial proximity alone, which is followed by grouping by spatiotemporal proximity. This grouping by spatiotemporal proximity determines the directions and the

[7] Such organization by common fate occurs for both translation (Wertheimer, 1923) and rotation (Julesz & Hesse, 1970).

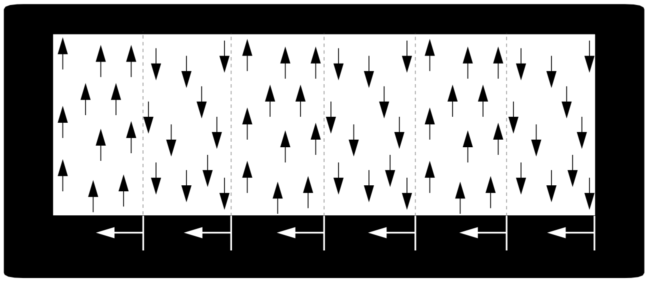


*Figure 17*. Second-order motion. In the stimulus of Cavanagh and Mather (1990), random texture elements moving upward alternate with strips of random texture elements moving downward. The elements moving in common direction group by common fate (see text), which allows the mechanism of spatial grouping 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.

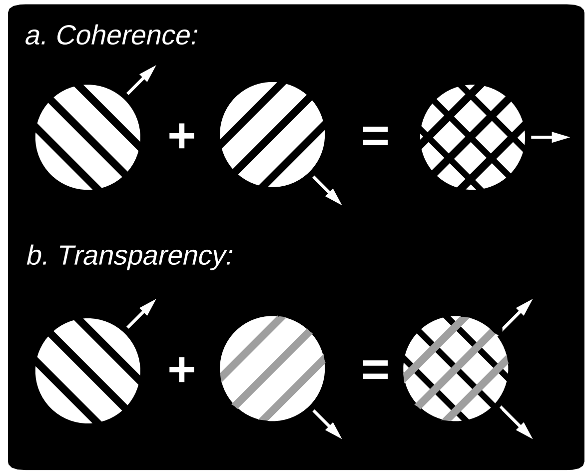



*Figure 18*. When identical drifting gratings are superimposed, they form a checkerboard pattern, a "plaid," which appears to move in the direction different from the directions of the component gratings (Adelson & Movshon, 1982). This effect can be explained by first applying spatial grouping to individual gratings, to form a plaid, which becomes a complex matching unit.

velocities of the elements which is used by the subsequent grouping by spatial proximity to derive the objects' shape.

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 (Figure 17). 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 grouping by spatial proximity, $S_1$, which cannot do much grouping because the elements in each frame are random. The elements in each frame are matched by grouping by spatiotemporal proximity ($T_1$) and identified as dots moving in up or down. Dots moving in the same direction undergo grouping by spatial proximity ($S_2$) 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 $T_2$. $T_2$ compares successive boundaries and detects their leftward motion (called *second-order motion* by Cavanagh & Mather, 1990). Note, however, that the output of $T_2$ does not depend on the fact that the boundaries between the strips are derived using grouping by spatiotemporal proximity. These boundaries could have been produced by grouping by spatial proximity based on a common property other than common motion, such as luminance or color. Thus, $T_2$ is independent from $T_1$, as the SM requires.

*Matching of groupings*.

We saw above that matching units can be spatial primitives or spatial aggregates of similar moving spatial primitives. Now consider the cases when grouping by spatial proximity organizes visual primitives into groupings that become matching units.

Adelson and Movshon (1982) showed observers two superimposed moving gratings through a circular aperture (Figure ). When either moving grating was presented alone, it was seen to move at right angles to the orientation of its bars, because of the matching unit problem and the correspondence problem (which is related to the aperture problem, which we will discuss later). When the superimposed gratings were identical (as in Figure *a*), 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 (as in Figure *b*), 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). The outcome of this experiment depends on the propensity of the gratings to fuse when overlaid as *static* gratings. 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 can infer the output of the spatial grouping by similarity that derives the matching units—the gratings or the plaids—independent of grouping by spatiotemporal proximity.[8]

To be sure, the components of a figure do not need to overlap to group in space and be seen as parts of a single moving entity (Shiffrar & Pavel, 1991). Consider, for example, the displays where a rectangle is moving vertically behind an opaque screen, seen through a variable number of circular apertures (Figure ; Ben-Av & Shiffrar, 1995). When only one edge of the rectangle is visible, it appears to move or-

[8] For yet another example, consider Stoner, Albright, and Ramachandran (1990) who showed that visual clues for static depth layout of the plaids (due to the apparent transparency, occlusion, or fusion of grid intersections) controls whether observers experience coherent or transparent motion. This and other relevant work is reviewed by Stoner and Albright (1994).

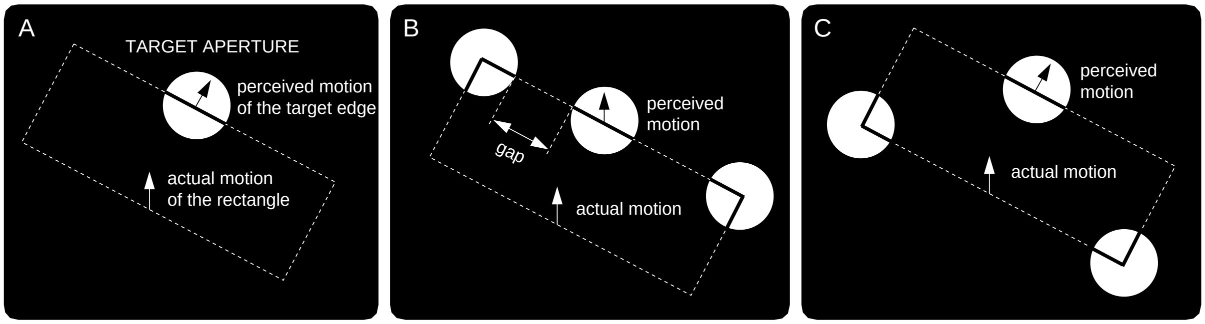



*Figure 19*. An outline of a rectangle (the white dotted outline in the diagrams) is translating *vertically* behind an opaque screen. A. When an edge (shown with the black solid line) of the rectangle is visible through a round aperture (labeled “target aperture”), the edge appears to move in a direction orthogonal to its orientation. B-C. Two corners of the rectangle are made visible in addition to the edge. Observers are more likely to see the target edge move vertically (i.e., veridically) when the visible corners are aligned with the target segment (B) than when they are not (C). The likelihood of seeing the target segment move vertically in B is greater when the “gap” is short (Ben-Av & Shiffrar, 1995).

thogonal to its orientation, which is the default solution of the aperture problem (which we will soon discuss). When a single corner is visible, it is seen to move vertically. Ben-Av & Shiffrar asked whether the motions of the corners can capture (or disambiguate) the motion of the edge, when both the corners and the edge are visible. They found that the motion of the corners did capture the motion of the edge when the visible corners were (a) collinear with the edge and (b) when the distance between the corners and the edge (the “gap” in Figure B) was short. When the corners were collinear but remote, or when they were not collinear, no matter how close, the motion of the edge was not affected by the motion of the corners; the edge appeared to move orthogonal to its orientation. In agreement with the SM, one can explain the findings of Ben-Av & Shiffrar by the grouping of corners and edges into matching units. When the visible components of the rectangle are collinear and close to each other, they group in space, so that grouping by spatiotemporal proximity occurs between the composite matching units. When the components are remote or not collinear they do not group by a spatial feature, and grouping by spatiotemporal proximity links the components.

*Matching of high-level units*.

Organizations more complex than aggregates of similar elements can become matching units. For example, Shepard and Judd (1976) showed that a rapid alternation of two images of a three-dimensional object (such as in Figure ) creates an impression of the object rotating through an appropriate angle in depth. To derive this motion grouping by spatiotemporal proximity must establish correspondence between the homologous parts of the perceived spatial object, rather than between small-scale spatial primitives (Rock, 1988, p. 57). In keeping with the SM, it was the grouping by spatial features of the frames that derived the complex matching units in the displays of Shepard and Judd.

If grouping by spatiotemporal proximity had matched small-scale entities between the frames in the display of Shepard and Judd the percept would have been different. Ramachandran, Armel, and Foster (1998) created a display that

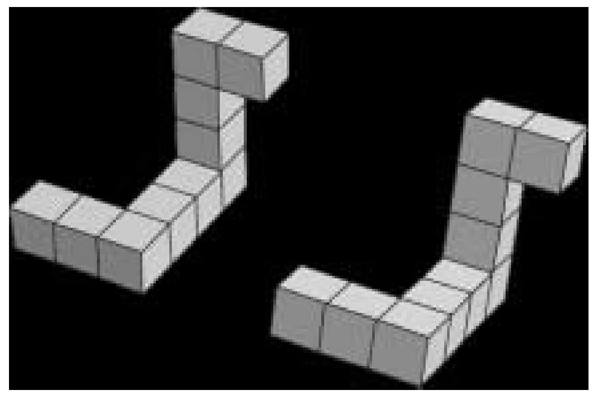


*Figure 20*. When the two images are shown in rapid alternation, observers see a three-dimensional object rotating through an appropriate angle (Shepard & Judd, 1976). To see such motion vision must establish correspondence between the parts of the three-dimensional objects, rather than between spatial primitives.

showed just that. These authors 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 (Figure ). 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. As in the study of Shepard and Judd, here it is the grouping by spatial features within the frames which derives complex matching units. Yet, the study of Ramachandran et al. goes beyond the finding of Shepard and Judd by showing that the familiarity of the nascent object can facilitate the grouping by spatial features of elements into complex matching units, and thus determine the level in the cascade of spatial organization which is accessed by grouping by spatiotemporal proximity.

Other evidence is available in support of the interaction between object familiarity and grouping by spatiotemporal proximity: e.g., Shiffrar and Freyd (1990), perceptual bias, forward-facing motion, direction of motion, cognition

(1992), Tse and Cavanagh (2000).

*Form and motion*.

Apparent motion has been commonly studied using displays of spatial shapes spatially well-segregated from the rest of the scene. In these displays the spatial distance between the successive shapes has usually been much greater than the distance between concurrent elements within the shapes. Under such conditions, the grouping by spatial proximity between the concurrent elements of a shape is much stronger than the grouping by spatiotemporal proximity between the elements of the shape in successive views. Hence, grouping by spatiotemporal proximity is given little chance to compete with the grouping by spatial proximity between the concurrent elements. It is not surprising that the studies using such displays always support the SM.

The assumption of the SM has been carried over into the (more general) studies of the interaction of form and motion. In this literature, vision has been thought to derive the form of an object before the grouping by spatiotemporal proximity between the objects takes place. Because of the bias in favor of the SM, the question of form-motion interaction has been generally posed in a way that excludes the IM: Do the form properties of moving objects affect grouping by spatiotemporal proximity? An answer to this question has been sought in two directions:

- One branch of this literature has been dealing with the *similarity* of object's form across the successive views of an object (e.g., Orlansky, 1940; Kolers, 1972; Burt & Sperling, 1981; Oyama, Simizu, & Tozawa, 1999).
- The other branch has explored more complex—*transformational*—relations between the successive forms (e.g., Warren, 1977; Eagle, Hogervorst, & Blake, 1999).

Within both branches of the literature, researchers arrived at disparate conclusions about form-motion interactions, i.e., no consensus has been reached on whether grouping by spatiotemporal proximity is sensitive to form differences of the grouped entities or not.

The distinction between the SM and IM, which we explore in this review, bears immediate implications on the research of form-motion interactions. If the IM is correct, the question of whether grouping by spatiotemporal proximity and object form affect each other should be explored under conditions where the strength of grouping by spatiotemporal proximity of objects is comparable with the strength of grouping by spatial proximity between the concurrent elements. Only in this case we will be able to discover the conditions under which the form of nascent objects affects the interactions between the concurrent and successive elements. (We know that such interactions do occur from the studies mentioned in section *Form and motion* on page 14.)

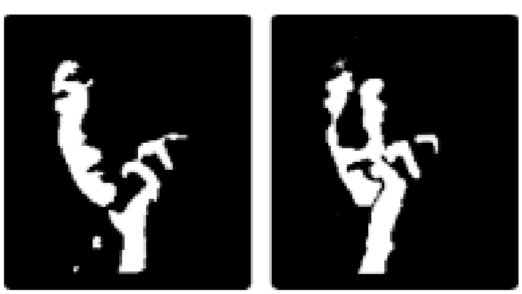


*Figure 21*. The two images of "Mooney faces" are shown to observers in rapid alternation. When the observers see a face, they perceive it rotating in depth. When they do not see a face, they perceive an incoherent motion in the picture plane (Ramachandran *et al.*, 1998). In keeping with the SM, spatial grouping forms faces which become matching units.

## *Against the sequential model*

*Ternus display and ISI effect*.

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

Consider the Ternus display (Ternus, 1936), in which dots can occupy three equally-spaced collinear positions (Figure C–E). These displays consist of two rapidly alternating frames, represented with two vertical dotted lines in Figure . The dots in one frames are 1 and 2; the dots in the other frame are 3 and 4. This display can give rise to two percepts: (a) *Element motion* (*e*-motion), which occurs when a single dot appears to move between the positions 1 and 4, and dot 2 appears immobile when replaced by dot 3 (Figure C); (b) *Group motion* (*g*-motion), which occurs when two dots appear to move back-and-forth as a group, from 1–2 to 3–4 (Figure D).

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

1. 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 grouping by spatial proximity rather than to grouping by spatiotemporal proximity, and thus the ISI effect is consistent with the SM.

2. 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 (Figure C[iii]). 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 & Sperling, 1981). Ullman (1979) also explained the percept of g-motion in the Ternus display in terms of the grouping by spatiotemporal proximity of individual elements.

*Dynamic superposition*.

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 grouping by spatial proximity alone cannot derive matching units when objects and surfaces are revealed gradually.

Take, for example, the perception of kinetic occlusion (Michotte, Thinès, & Grabbé, 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 & Rock, 1974; Kellman & Cohen, 1984; Tse, Cavanagh, & Nakayama, 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 & Kellman, 1993; Cicerone, Hoffman, Gowdy, & Kim, 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 this case, as in the examples above, 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, 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).

## Inseparability of grouping by spatial proximity and grouping by spatiotemporal proximity

*Motion lattices*.

To refute the SM, we must demonstrate that grouping by spatial proximity and grouping by spatiotemporal proximity interact even when a simple correspondence between the successive frames is possible and no specialized, or inferential, mechanism is required. We will now review a study by Gepshtein and Kubovy (2000) who tested the SM using spatiotemporal dot lattices, called *motion lattices*. These stimuli allowed Gepshtein and Kubovy to vary the strength of grouping by spatial proximity and grouping by spatiotemporal proximity independently of each other, by manipulating spatial proximity between concurrent and successive dots (Figure ).

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) grouping by spatial proximity may consolidate as the ISI grows, and (b) longer ISIs may favor matching over a different spatial range. Therefore in motion lattices Gepshtein and Kubovy held ISI constant, and varied the strength of grouping by spatiotemporal proximity by manipulating the *spatial proximity between successive dots*.

In motion lattices displays, as in the Ternus displays, 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 the former 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 (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. Their stimulus did not allow Burt and Sperling to explore the effect of relative proximity between concurrent and successive dots, which we will examine presently. In motion lattices each frame contains a two-dimensional pattern of dots, which allowed us to set up a competition between alternative *spatial* organizations within a frame and ask whether grouping by spatiotemporal proximity affects grouping by spatial proximity.

*A critical test of the SM*.

According to the SM, the propensity of elements to form virtual objects within frames, and thus yield *g*-motion, is independent of the determinants of grouping by spatiotemporal proximity, i.e., grouping between *successive* dots. As Kubovy, Holcombe, and Wagemans (1998b) showed, spatial grouping within static dot lattices is determined by relative proximity between the *concurrent* dots. That is to say, the angles between alternative organizations of the lattice and its symmetry properties do not affect its organization. Gepshtein and Kubovy (2000) used this property of static dot lattices to test the SM by asking whether the frequency of *g*-motion changes when we hold constant the relative proxim-

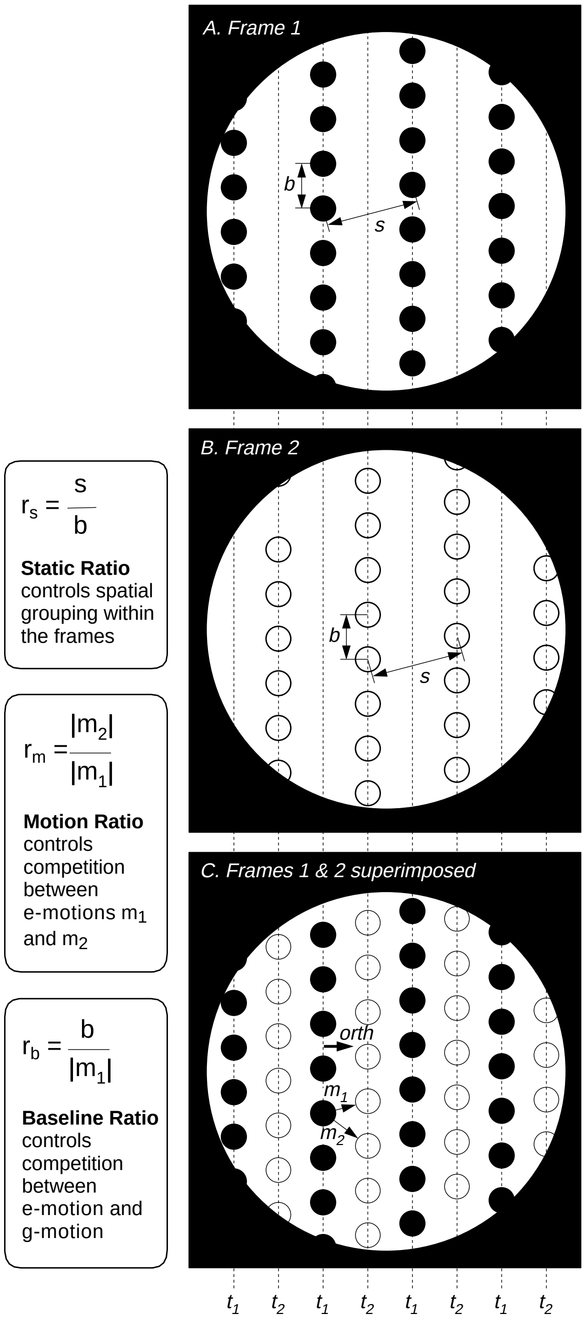



*Figure 22*. Design of motion lattices. Two frames of a motion lattice are shown schematically in *A* and *B*; the frames are superimposed in *C*. Distances *b* and *s* correspond to the shortest inter-dot distances within the frames (shown in *A and B*). Vectors $m_1$ and $m_2$ (shown in *C*) are the most likely e-motions, i.e., motions derived by matching of individual elements. When vision derives motion by matching dot groupings (called *virtual objects*), rather than dots themselves, motion orthogonal to the virtual objects is seen (g-motion). In *C*, g-motion is horizontal (notated *orth* ), because the virtual objects are vertical.

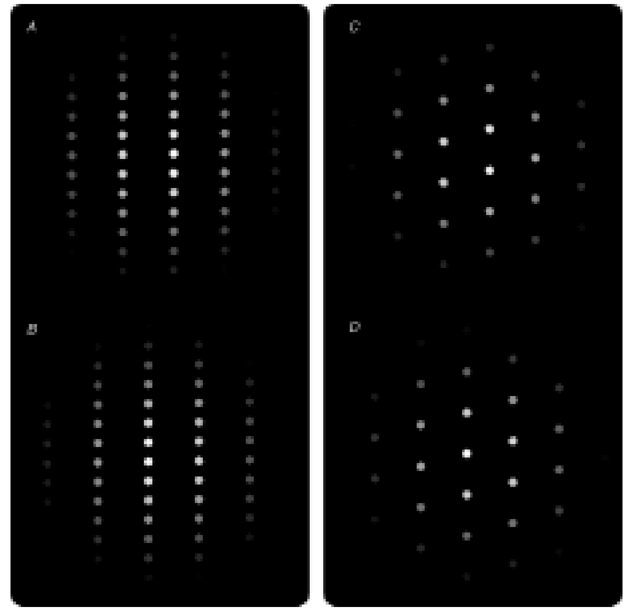


*Figure 23*. Screen snapshots of a motion lattices (not to scale). *A-B*. A lattice where g-motion is likely (high $r_s$ and low $r_b$). *C-D*. A lattice where e-motion is likely (low $r_s$ and high $r_b$).

ity between concurrent dots and vary the proximity between successive dots.

Figure  explains the geometry of a specific class of motion lattices used by Gepshtein and Kubovy (2000). These motion lattices are obtained by splitting static lattices into two frames, so that every frame contains every second column (or row) of the original lattice. (These particular lattices are called *two-stroke motion lattices* $M^2$; they belong to a general class of motion lattices $M^i$, where $i$ is the number of component frames.) When the component frames of a motion lattice are shown in rapid alternation, observers see a flow of apparent motion under appropriate spatial and temporal conditions. We say that observers experience **e-motion** when they report seeing dots flow in a direction of matching between individual dots; we say that observers experience **g-motion** when they report seeing dots flow in the direction orthogonal to the virtual objects formed within the frames.

In the schematic shown in Figure , the dots are arranged such that they are likely to group within frames into vertical virtual objects. If grouping by spatiotemporal proximity across frames occurs between the virtual objects, rather than between the dots, one sees motion orthogonal to the virtual objects (i.e., horizontal motion in Figure ). Frame snapshots of an actual $M^2$, where horizontal g-motion is likely, are shown in Fig.  (*A–B*). If dots are arranged such that virtual objects within the frames are less salient, then g-motion is less likely. For example, the two frames shown on panels *C–D* of Figure  belong to an $M^2$ where g-motion is less likely than in the $M^2$ whose frames are shown on panels *A–B*.

According to the SM, the rate of g-motion depends on the propensity of concurrent dots to form virtual objects within the frames, and does not depend on grouping by spatiotem-

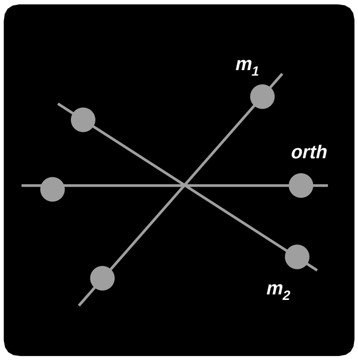



*Figure 24*. A response screen corresponding to the lattice shown in Figure C-D (not to scale). In every trial observers click on a circle attached to the radial line parallel to the perceived direction of motion. Response labels $m_1$, $m_2$, and *orth* do not appear on the actual response screen.

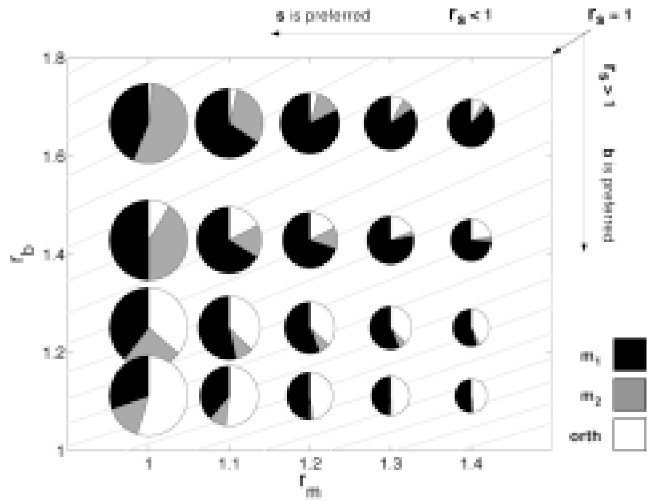



*Figure 26*. Results of the experiment of Gepshtein & Kubovy (2000). The pie charts on the top panel show the distributions of three responses ($m_1$, $m_2$, and *orth*) for twenty configurations of motion lattices. The gray lines in the background are iso-$r_s$ lines; within these lines the salience of spatial virtual objects is invariant (see text).

poral proximity. In other words, the SM holds that matching units are derived by grouping by spatial proximity alone, so that only spatial proximities between concurrent dots determine the likelihood of whether e-motion or g-motion is seen. In contrast, the IM holds that the grouping by spatial proximity of dots within frames is affected by grouping by spatiotemporal proximity between successive dots.

To pit the two models against each other, Gepshtein and Kubovy (2000) measured the relative rate of e-motion and g-motion under conditions of equivalent spatial grouping within the frames. Within the frames, the salience of virtual objects does not change as long as the ratio between relevant spatial distances is invariant. Thus, the propensity of dots to form vertical virtual objects in Figure  does not change, as long as $r_s = \frac{s}{b}$ does not change. The SM predicts that the rate of g-motion relative to the rate of e-motion will not change when we vary the strength of grouping by spatiotemporal proximity, as long as $r_s$ is constant. The IM predicts that the rate of g-motion will drop relative to the rate of e-motion, as we increase the strength of temporal grouping, while keeping $r_s$ constant.

The experiments of Gepshtein and Kubovy (2000) unequivocally support the prediction of the IM. The pie charts of Figure 26 show the distributions of three responses—$m_1$, $m_2$, and *orth*—for different configurations of motion latices. Three trends in these data are noteworthy:

1. The frequency of $m_1$ motion grows as a function of $r_m$.
2. The frequency of *orth* motion drops as $r_b$ grows.
3. The frequency of *orth* motion varies within the sets of iso-$r_s$ conditions, marked with oblique gray lines.

Gepshtein and Kubovy explicated the last observation in two steps:

1. They constructing a statistical model of the data shown in Figure 26. (The model accounted for 98% of variance in the data.)

2. They interpolated motion frequencies within the iso-$r_s$ sets of parameters. A result of this computation is shown in Figure 2, where each curve plots the relative frequency of g-motion and e-motion within a corresponding $r_s$ set.

Figure 2 shows that, when $r_s$ is ($s \gg b$, as in the top iso-$r_s$ curves), grouping by spatial proximity within the frames tends to derive vertical virtual objects (vertical in the coordinate system used in Figure ), and horizontal g-motion is likely. When $r_s$ decreases ($s$ approaches $b$, as in the bottom iso-$r_s$ curves), the salience of vertical virtual objects drops,

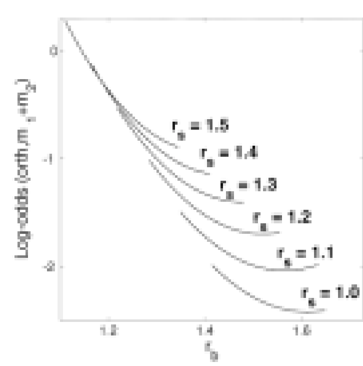



*Figure 27*. Results of the experiment of Gepshtein & Kubovy (2000). The relative likelihood of g-motion (*orth* responses) and e-motion ($m_i$ responses) changes within the iso-$r_s$ sets of conditions, in contrast to the prediction of the sequential model (see text).

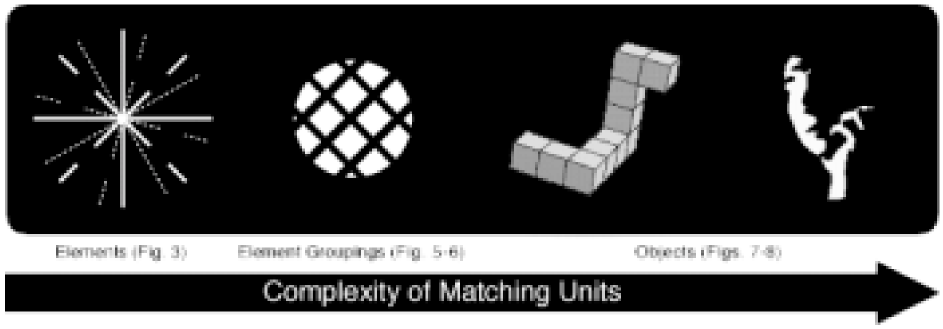



*Figure 25*. Spatial entities of various complexity compete to become matching units. Two instances of objects being matching units are shown because objects can become matching units independent of whether they belong to a category familiar to the observer (faces) or not (block-figures). According to the sequential model, grouping by spatial proximity solely determines the identity of matching units. The interactive model holds, in contrast, that the identity of matching units emerges in the interaction of grouping by spatial proximity and grouping by spatiotemporal proximity. (This figure complements Figure .)

and the likelihood of g-motion decreases. Critically, the fact that the frequency of g-motion changes within the iso-$r_s$ sets indicates that it is not only the spatial proximities within the frames that determine *what* is seen to move in motion lattices. This evidence unequivocally demonstrates the validity of the IM.

### *Where is the Gestalt?*

We believe that we have found what we were looking for: a complex dynamical system that does not allow us to decompose motion perception into two successive grouping operation: grouping by spatial proximity and grouping by spatiotemporal proximity. Rather we have found that they are intertwined in ways that will be a challenge to understand.

## In praise of experimental phenomenology

### *Subjectivity and objectivity in perceptual research*

Palmer (this volume) opens his discussion of Methodological Approaches to the study of perceptual organization with his Figure 2 (p. xxx), in which he shows several demonstrations of grouping. His discussion of such demonstrations concludes that the "phenomenological demonstration is a useful, but relatively blunt instrument for studying perceptual organization." (p. xxx). We wholeheartedly concur. Two reasons Palmer gives for worrying about phenomenological demonstrations are: (1) they do not produce quantifiable results, and (2) they have a subjective basis. Palmer believes that the quantification problem can be overcome, by using what he calls "quantified behavioral reports of phenomenology," an approach we prefer call *experimental phenomenology*.[9] All the experiments we describe in this chapter belong to this category.

Although experimental phenomenology may solve the quantification problem, does it solve the problem of subjectivity? Twenty years ago, when Pomerantz and Kubovy (1981) wrote the overview chapter of *Perceptual Organization*, they did not think so:

> . . . the pragmatic streak in American psychology drives us to ask what role . . . experiences, however compelling their demonstration, play in the causal chain that ends in action. Thus we ask whether such phenomenology might not be a mere epiphenomenon, unrelated to behavior. [p. 426]

Palmer's scepticism is very much in line with this position. His solution to the problem—to use "objective behavioral tasks"—is also in agreement with Pomerantz and Kubovy:

> . . . if we can set up situations in which we ask subjects questions about the stimulus that have a correct answer, and if organizational processes affect their judgments (and so their answers), then the experimentalists' skepticism about the importance of organizational phenomena should be dispelled. This book presents a wealth of organizational phenomena that can be demonstrated by both the phenomenological method and by objective experimental techniques. [p. 426]

We have come to disagree with Pomerantz and Kubovy's views on this matter and therefore disagree with Palmer's. First of all there is the matter of the contrast between *phenomenological* and *objective*. It is tendentious to use the terms subjective or objective in this context, for two reasons.

[9] We prefer our term, because we think that the data produced by such a method should be called quantified only if they have been described by a metric mathematical model. In experimental phenomenology, responses of different kinds can be counted, and therefore statistics may be applicable. They may or may not lend themselves to mathematical modeling.

First, because subjectivity is widely thought to be inconsistent with the scientific method, whereas objectivity is its hallmark. Second, because objectivity bespeaks unbiasedness; in current English it has an honorific connotation.

When we study perceptual organization we are studying perceptual experiences that are *phenomenal but not idiosyncratic*. The Merriam-Webster Dictionary gives several definitions for the adjective “subjective,” two of which are relevant here.

**Subjective = phenomenal:** “A characteristic of or belonging to reality as perceived rather than as independent of mind.”

**Subjective = idiosyncratic:** “Peculiar to a particular individual ... arising from conditions within the brain or sense organs and not directly caused by external stimuli.”

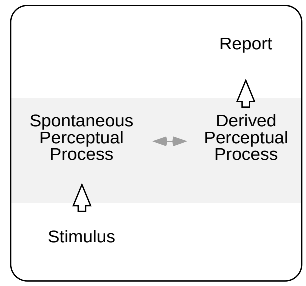
One source of the concern with the “subjectivity” of the phenomena of perceptual organization is the conflation of these two senses. Judging that something is *red* is accompanied by a subjective experience which is phenomenal but is not idiosyncratic. One can easily find an object and viewing conditions under which an overwhelming majority of people would agree that the object is red. Judging that an object is *beautiful* is also accompanied by a subjective experience, but this experience is both (1) phenomenal *and* (2) idiosyncratic. It is not so easy find an object and viewing conditions under which an overwhelming majority of people would agree that the object is beautiful.

That is why, when we study perceptual experiences that are phenomenal but not idiosyncratic, we say that we are doing experimental phenomenology rather than studying subjective experience. We recommend that the discipline eschew the use of the terms *objective* or *subjective* to characterize perceptual research methods. They can only lead to confusion.

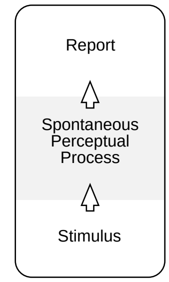
## *The role of traditional psychophysical tasks*

What about the concern voiced then by Pomerantz and Kubovy, and now by Palmer: Can one determine whether an experience is epiphenomenal (perhaps Palmer would call it “purely subjective”)? We take this concern seriously. After all, “measuring perceived grouping is fundamentally different from measuring perceived size or perceived distance, which have well-defined objective measures against which people’s behavioral reports can be compared for accuracy” (Palmer, p. xxx). Perceived size can be studied by the traditional methods of psychophysics. Can perceptual organization be studied by embedding it in an experimental task for which responses can be judged to be correct or incorrect, i.e., a traditional psychophysical task? This involves, to quote Palmer, “changing what is actually being studied from subjective grouping to something else” (p. xxx).

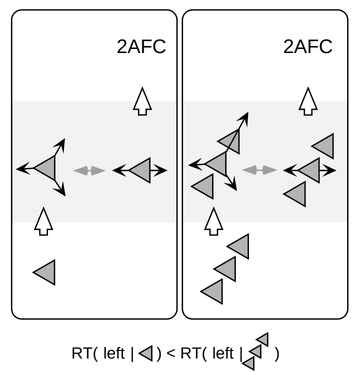
We will now show just what this transformation of one task into another entails and what it achieves. Then, we will show what role experimental phenomenology can play in the study of perceptual organization. As opposed to experimental phenomenology, traditional psychophysical tasks are indirect. This idea is illustrated in Figures 28(a) and 28(b). In natural viewing conditions, as well as in the tasks used in experimental phenomenology, certain aspects of the visual scene (“stimulus” in the figure) lead to a corresponding percept by means of a private perceptual process. The latter is labeled as a “spontaneous perceptual process” in the figure to emphasize that the process occurs naturally, just as it does when the observer views the stimulus outside of the



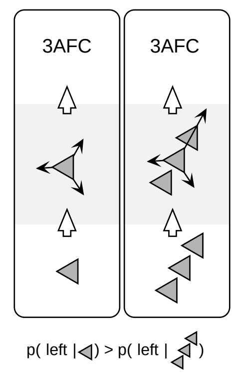

(a) Traditional psychophysical procedure




(b) Experimental phenomenology procedure




(c) The Palmer & Bucher (1981) experiment as a traditional psychophysical procedure (2AFC).




(d) Hypothetical study of the pointing using phenomenological report (3AFC

*Figure 28*. Comparison of the processes that take place in an observer engaged in different types of experimental procedures. The hypothetical events that are not public are marked by a grey band. The procedures of traditional psychophysics force the observer to do “perceptual work” (horizontal arrows), i.e., to transform their experience in order to meet the requirements of the procedure. Thus they engage additional perceptual processes compared to the procedures of experimental phenomenology. In that sense the latter are more direct than the former. See text for details.

laboratory. The gray regions in Figure are private in the sense that only the observer enjoys an immediate access to the outcomes of this process; this experience is made public, i.e., accessible to others, by means of a "report." The experimental phenomenologist strives to devise experimental conditions such as to make the report as close as possible to how observers would describe their experiences outside of the laboratory, but in a highly controlled environment. We will refer to such reports as "phenomenological."

In traditional psychophysics the natural perceptual experience is transformed. It is transformed by asking observers to judge certain aspects of the stimulus, which engages mechanisms normally not involved in the perception of natural scenes. Or, the perception of the stimulus is hindered, either by adding external noise to the stimulus or by presenting the stimulus at the threshold of visibility. We question whether such transformations of perceptual experience are indispensable in the studies of perceptual organization.

As an illustration of traditional psychophysics applied to the research of perceptual organization, consider the experiments in which Palmer and Bucher (1981) studied the pointing of equilateral triangles (Figure 28(c)). An equilateral triangle appears to point about equally often at 60°, 180°, or 300° (Figure 28(c), left). If you align three such equilateral triangles along a common axis of mirror symmetry tilted 60° (Figure 28(c), right), they appear to point most often at 60°. Palmer and Bucher used a 2-alternative forced-choice (2AFC) procedure; they asked observers to decide whether the triangle(s) can be seen pointing right or left (0° or 180°; Figure 28(c), left). Obviously, these triangles cannot point to the right (0°). We have seen that the isolated triangle appears to point spontaneously in all directions equally but when axis-aligned it tends to point at 60°. As a consequence, in the configuration shown in Figure 28(c) (right panel), observers were slower to decide whether the axis-aligned triangles point to the right or to the left than to decide whether the isolated triangle does. ("RT" in Figure 28(c) stands for reaction time.) We will say that the pointing induced by the common axis is forcing the observers in the experiments of Palmer and Bucher to do *perceptual work*: the observers must overcome the automatic effect of alignment on pointing, in order to focus on the properties of each triangle, and give a correct answer. Perceptual work is a transformation of spontaneous experience; it is represented in Figure 28(c) by horizontal arrows. It is this perceptual work that persuades us that the effect of common axis is not epiphenomenal (or purely subjective).

After one has established that the effect of common axis on pointing is not epiphenomenal, one could explore the effect directly, without forcing observers to do perceptual work (Figure 28(d), right). For example, one could use an experimental phenomenology procedure with a 3-alternative forced-choice (3AFC) in which the observer's task is to report (by pressing one of three keys) in which direction the middle (or single) triangle is pointing (Figure 28(d): "p(X)" stands for the probability of percept X.) This is a phenomenological report because the three report categories offered to the observers agree with the three likely spontaneous organizations of the stimulus.

The inferences involved in the interpretation of psychophysical studies of perceptual organization would not make sense without assuming the existence of a spontaneous organization, which would have led to the phenomenological report. They are an *indirect assessment* of the effects of grouping. We are not implying tasks that have correct and incorrect responses are useless *after* one has established that an organizational phenomenon is not epiphenomenal. These tasks can undoubtedly give us important information about the underlying process. But it would be unfortunate if we implied that these indirect psychophysical methods have an intrinsic advantage over phenomenological methods. Indeed, it was one of the goals of this chapter to demonstrate the power of experimental phenomenology.

## References


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

Ben-Av, M. B., & Shiffrar, M. (1995). Disambiguating velocity estimates across image. *Vision Research*, *35*(20), 2889–2895.

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.* N.p.: Crystallographic Society of America. (Original work published 1866)

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

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

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

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

Eagle, R. A., Hogervorst, M. A., & Blake, A. (1999). Does the visual system exploit projective geometry to help solve the motion correspondence problem? *Vision Research*, *39*(2), 373–385.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: Wiley.

Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences*, *97*, 8186–8191.

Grünbaum, B., & Shepard, G. C. (1987). *Tilings and patterns.* New York: W. H. Freeman.

Helmholtz, H. v. (1962). *Treatise on physiological optics* (Vol. III (Originally published in 1867)). NY: Dover Publications.

Hildreth, E. C. (1983). *The measurement of visual motion.* Cambridge, MA: The MIT Press.

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

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

Julesz, B., & 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.* NY: Praeger.

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

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

Kolers, P. A. (1972). *Aspects of motion perception.* Oxford, UK: Pergamon.

Korte, A. (1915). Kinematoskopische Untersuchungen [Kinematoscopic investigations]. *Zeitschrift für Psychologie*, *72*, 194-296.

Kramer, P., & 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., & Tversky, A. (1971). *Foundations of measurement* (Vol. I: Additive and polynomial representations.). New York: Academic Press.

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

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

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

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

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

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Martin, G. E. (1982). *Transformation geometry: An introduction to symmetry.* New York: Springer-Verlag.

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

McBeath, M., Morikawa, K., & Kaiser, M. (1992). Perceptual bias for forward-facing motion. *Psychological Science*, *3*(6), 362-367.

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., & Grabbé, G. (1964). Les compléments amodaux des structures perceptives. In *Studia psychologica*. Publications Universitaires de Louvain, Louvain, Belgium.

Neisser, U. (1967). *Cognitive psychology.* NY: Appleton Century Crofts.

Orlansky, J. (1940). The effect of similarity and difference in form on apparent visual movement. *Archives of Psychology (Columbia University)*, *246*, 1–85.

Oyama, T., Simizu, M., & Tozawa, J. (1999). Effects of similarity on apparent motion and perceptual grouping. *Perception*, *28*(6), 739-748.

Palmer, S. E., & Bucher, N. M. (1981). Configural effects in perceived pointing of ambiguous triangles. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 88–14.

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

Pomerantz, J. R., & Kubovy, M. (1981). Perceptual organization: An overview. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 423–456). Hillsdale, NJ: Lawrence Erlbaum.

Ramachandran, V. S., Armel, C., & 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.* Cambridge, MA: The MIT Press.

Rock, I. (1988). The description and analysis of object and event perception. In K. R. Boff, L. Kauffman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 33-1–36-71). New York: Wiley.

Shepard, R. N., & Judd, S. A. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science*, *191*, 952–954.

Shiffrar, M., & Freyd, J. F. (1990). Apparent motion of the human body. *Psychological Science*, *1*(4), 257–64.

Shiffrar, M., & Pavel, M. (1991). Percepts of rigid motion within and across apertures. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 749–761.

Shipley, T. F., & 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., & Rock, I. (1974). Stroboscopic movement based on perceptual intelligence. *Perception*, *3*, 9–28.

Smith, A. T. (1994). The detection of second–order motion. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 145–76). Academic Press.

Stoner, G. R., & Albright, T. D. (1994). Visual motion integration: A neurophysiological and psychophysical perspective. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 253–90). Academic Press.

Stoner, G. R., Albright, T. D., & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature*, *344*, 153–1551.

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

Tse, P., Cavanagh, P., & Nakayama, K. (1998). The role of parsing in high-level motion processing. In T. Watanabe (Ed.), *High-level motion processing* (pp. 249–66). Cambridge, MA: The MIT Press.

Tse, P. U., & Cavanagh, P. (2000). Chinese and americans see opposite apparent motions in a chinese character. *Cognition*, *74*(3), B27–B32.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: The MIT Press.

Wallach, H. (1935). Uber visuell wahrgenommene Bewegungsrichtung [On the visually perceived direction of motion]. *Psychologische Forschung*, *20*, 325–80.

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

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

Warren, W. H. (1977). Visual information for object identity in apparent movement. *Perception & Psychophysics*, *21*(3), 264-268.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung [Experimental studies on seeing motion]. *Zeitschrift für Psychologie*, *61*, 161–265.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II [Investigations of the principles of Gestalt, II]. *Psychologische Forschung*, *4*, 301–350.

Wilson, H. R. (1994). Models of two-dimensional motion perception. In A. T. Smith & R. J. Snowden (Eds.), *Visual detection of motion* (pp. 219–51). Academic Press.

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