To be published in: M. Behrmann & R. Kimchi (Eds.),
Perceptual Organization in Vision: Behavioral and Neural Perspectives

Perceptual grouping in space and in spacetime: An exercise in phenomenological psychophysics

Michael Kubovy and Sergei Gepshtein
University of Virginia

The first two sections of this chapter are devoted to empirical and theoretical studies of grouping in space and in spacetime. In the first section we summarize recent research on grouping by proximity. We show that grouping by proximity can be modeled with a simple model that has few of the characteristics that one might expect of a Gestalt phenomenon. In the second section we (a) review the literature on the relation between grouping by spatial proximity and grouping by spatiotemporal proximity, and (b) summarize our research that shows that these two processes are inextricably entangled. The third section is meta-methodological. We do *phenomenological psychophysics*. Because the observers' responses are based on phenomenal experiences, which are still in bad repute among psychologists, we conclude the chapter with an explication of the roots of such skeptical views, and show that they have limited validity.

Contents

A clarification of terms **2**

Grouping by spatial proximity **2**
Generalizing the Gestalt lattice . . . 3
The instability of lattices . . . 5
The pure distance law . . . 7

Grouping by proximity in spacetime **9**
Sequential and interactive models . . . 9
Evidence that seems to favor interactive models, but doesn't . . . 10
Evidence for interactive models . . . 14

In praise of phenomenological psychophysics **17**
Subjectivity and objectivity in perceptual research . . . 17
The role of traditional psychophysical tasks . . . 18

References **19**

Perceptual organization lies on the border between our experience of the world and unconscious perceptual processing. It is difficult to study because it involves both bottom-up and top-down processes and because it is—like respiration—a semi-voluntary process. For example, when we first glance at a Necker cube, we usually see a cube below eye level. Over this response we have no control; it is spontaneous and automatic. But as soon as we have seen the cube reverse we seem to have some control over our interpretation of the drawing.

In this chapter we summarize our work on grouping in static and dynamic stimuli. In this work we have developed new methodologies which have allowed us to explore perceptual organization more rigorously than had hitherto been possible. These methodologies rely on the spontaneity and the multistability of grouping while taking care to minimize the effects of whatever voluntary control observers might have over what they see.

The first two sections of this chapter deal with empirical and theoretical studies of grouping. The third is meta-methodological. This third section is needed because 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

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 at School of Optometry, 360 Minor Hall, University of California, Berkeley, CA 94720-2020 (e-mail: *sergeg@uclink.berkeley.edu*).

to supplant ours. So we conclude the chapter with an explication of the roots of such skeptical views, and show that they have limited validity.

A clarification of terms

Grouping by spatial proximity is the process by which temporally concurrent regions of a scene become perceptually linked across space. Small-scale regions are customarily called *elements*. Figure 1A shows two such elements—❶ and ❷—being connected into a larger-scale entity. The two elements in Figure 1A could persist at their locations; in that case they would had been depicted as horizontally elongated ovals in the space-time diagram.

If ❶ and ❷ are not concurrent (❶ appears at time 1 and ❷ appears at time 2), to perceive ❶ and ❷ as a single element in motion (Figure 1B), vision must link them across space *and* across time, a process we call *grouping by spatiotemporal proximity* (Gepshtein & Kubovy, 2001).

Now consider a more complicated display: two elements at time 1 (❶ and ❷) and two elements at time 2 (❸ and ❹). When motion is seen, one of two things happens: [i] ❶ may be linked with ❹, and ❷ is likely to be linked with ❸ so that the elements are matched independently (Figure 1C[i]); or [ii] the elements may form a grouping and be seen as a single moving entity (❶❷⟶❸❹, Figure 1C[ii]). In the latter case, the motion of the elements is the same as the motion of the group (Figure 1C[iii]). In this process the visual system establishes a *correspondence* between elements visible at successive instants; the successive visual entities that undergo grouping by spatiotemporal proximity are called *correspondence tokens* (Ullman, 1979). Grouping by spatiotemporal proximity is also known as *matching*; the entities are then called *matching units*.

The representation of entities by dots in Figure 1 should not be taken too literally. Research on perceptual grouping is not only about the perception of displays that consist of elements which are discrete in time and space. In the section *Grouping by spatiotemporal proximity* we discuss several examples of grouping between regions which at every instant appear connected, but which behave as separate matching units in grouping by spatiotemporal proximity. In general, we prefer to think of grouping as a process that causes us to see certain regions of the scene as being connected (whether they are connected or not), rather than a process that causes us to see the connections among discrete entities.

Grouping by spatial proximity

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 2(a) one spontaneously sees horizontal grouping. In Figure 2(b) one can also see horizontal grouping, but with some difficulty. The tendency to see horizontal grouping is *weaker* in Figure 2(b) than in Figure 2(a). Such observations are the seed of a quantitative theory.

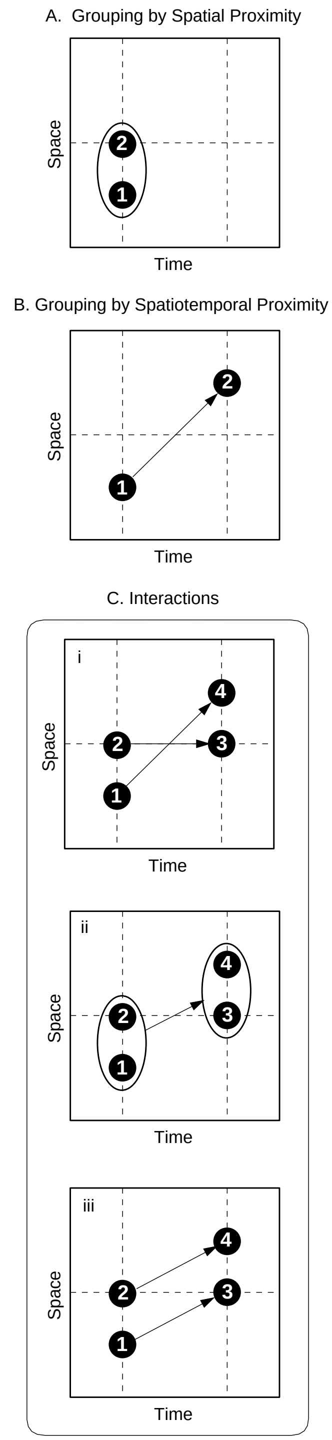
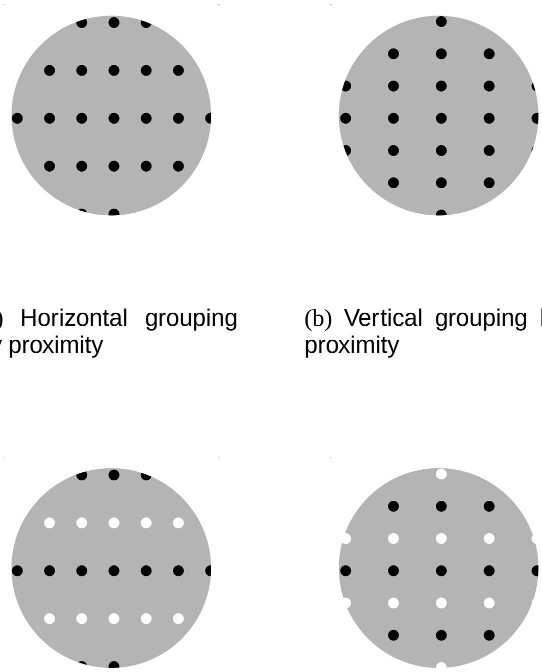


Figure 1. (A) Grouping by spatial proximity, (B) grouping by spatiotemporal proximity, and (C) their interactions. The configurations in C[i] and [ii] were introduced by Ternus (1936) who described two kinds of percepts: [i] *element-motion*: ❶⟶❹ and ❷⟶❸; and [ii] *group-motion*: ❶❷⟶❸❹. [iii] Ullman (1979) argued that what looks like group-motion may actually be element-motion ❶⟶❸ and ❷⟶❹ (see also Figure 16).



(a) Horizontal grouping by proximity

(b) Vertical grouping by proximity

(c) Proximity and similarity in concert

(d) Proximity and similarity in opposition

Figure 2. Examples of grouping by proximity and of the interaction of grouping by proximity and similarity

More than thirty years 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 thought that the 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 6×6 rectangular lattices of squares (Hochberg & Silverstein, 1956) or 4×4 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 2(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 luminance difference between the rows, the distance between rows for which observers reported an equilibrium between rows and columns increased (Figure 3). We call this is a *grouping indifference curve* because the observer is indifferent among the ⟨luminance-difference, row-distance⟩ pairs

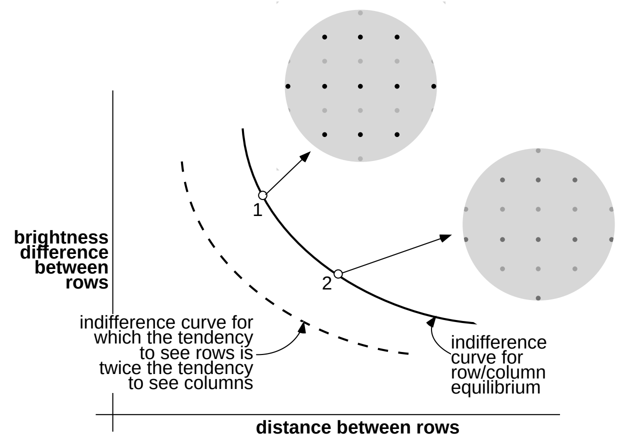


Figure 3. Two grouping indifference curves. Only the solid curve is achievable by methods such as Hochberg's (illustrated here for the trade-off between grouping by proximity and grouping by similarity) and Burt and Sperling (1981) (for the trade-off between grouping by spatial proximity and grouping by temporal proximity). Our method allows us to plot a *family* of indifference curves.

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 3). 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 2(a) and 2(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 4). 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 over-

[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, the ⟨meat, potato⟩ pairs ⟨1, 4⟩ and ⟨2, 1⟩ are said to lie on an indifference curve.

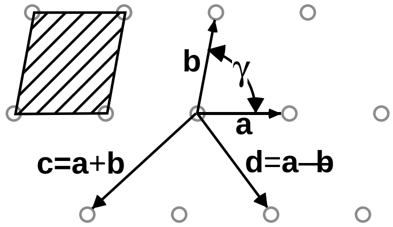


Figure 4. The features of a dot lattice (see text).

lay 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** leaves 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 4) 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 neighbors. Its distance from a neighbor is either **a**, **b**, **c**, or **d**. Another way to think about a dot lattice is to consider its building block: the *basic parallelogram*, *ABCD* in Figure 4. Its sides are **a** and **b**; its diagonals are **c** and **d**.

Any lattice can be defined by specifying three parameters: the lengths of the two sides of the basic parallelogram, $|\mathbf{a}|$ and $|\mathbf{b}|$, and the angle between them, γ. 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 γ.[3] Furthermore these parameters are somewhat constrained. The distances between dots are constrained by the inequalities $|\mathbf{a}| \leq |\mathbf{b}| \leq |\mathbf{c}| \leq |\mathbf{d}|$, and the angle γ is bounded, such that $60° \leq \gamma \leq 90°$ (Kubovy, 1994).

The two-parameter space of lattices is depicted in Figure 5. This space can be partitioned into six classes, whose names appear in Figure 5. The differences among these classes are portrayed in Figure 6, 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.

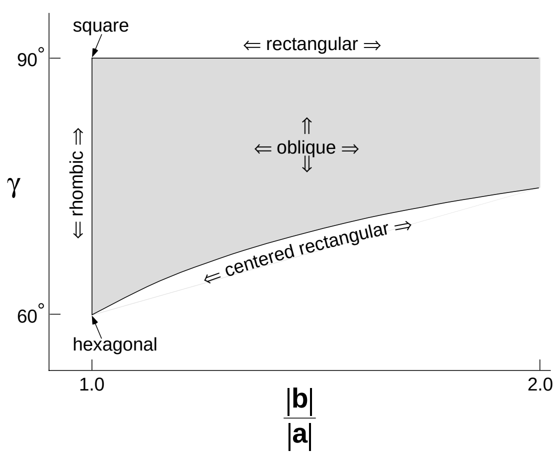


Figure 5. The space of dot lattices. See Figure 6.

The left to right order of the lattice classes in Figure 6 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.

---

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

[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).

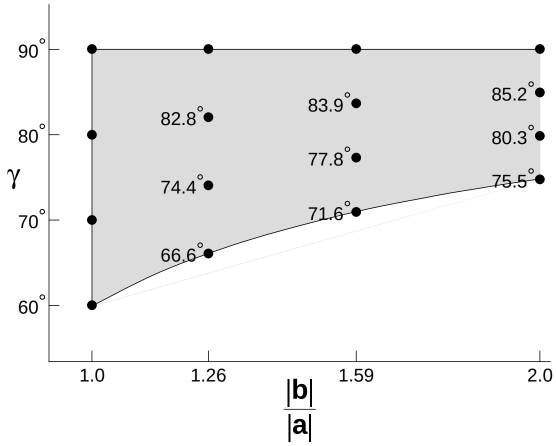
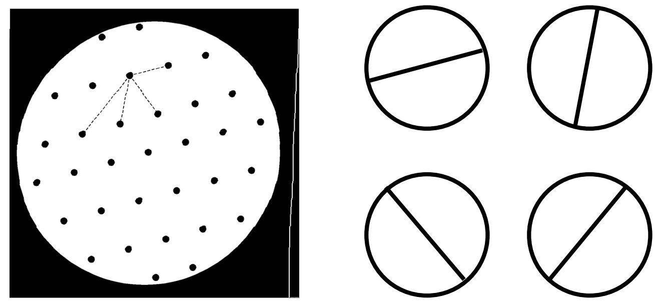


Figure 8. How the sixteen stimuli were sampled from the space of dot lattices. See Figure 5.



(a) A stimulus with four vectors

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

Figure 9. The Kubovy & Wagemans (1995) experiment.

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 7), sampled systematically from the space shown in Figure 5. The screen was divided into two regions, the aperture and the black mask around it (Figure 9(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 selecting one of four icons on the response screen (Figure 9(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 phenomenological psychophysics (see p. 17).

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—α, β, γ, and δ—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 υ 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 10.

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 internal states; we do so 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 11.

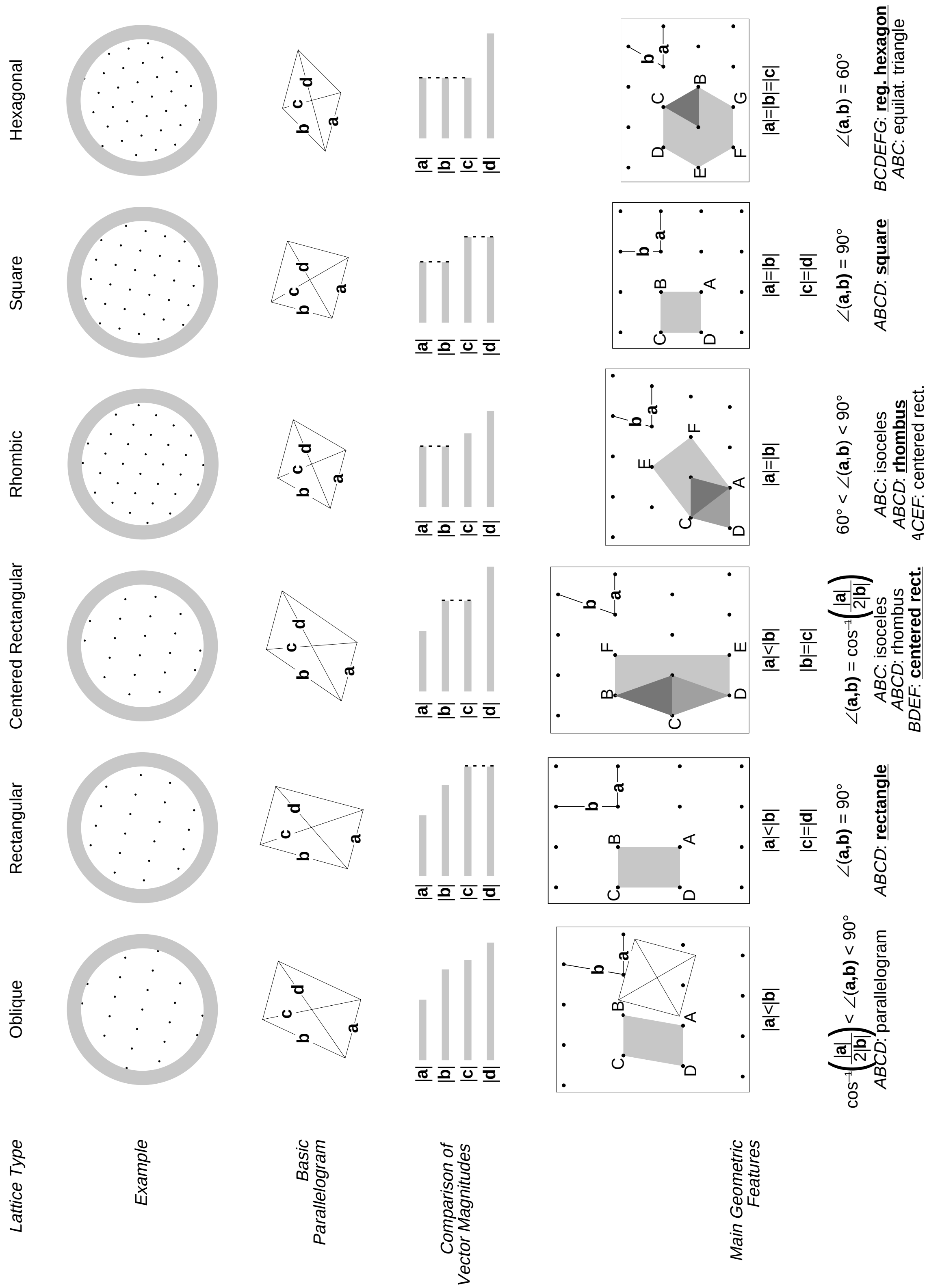


Figure 6. The six classes of dot lattice according to Kubovy (1994).

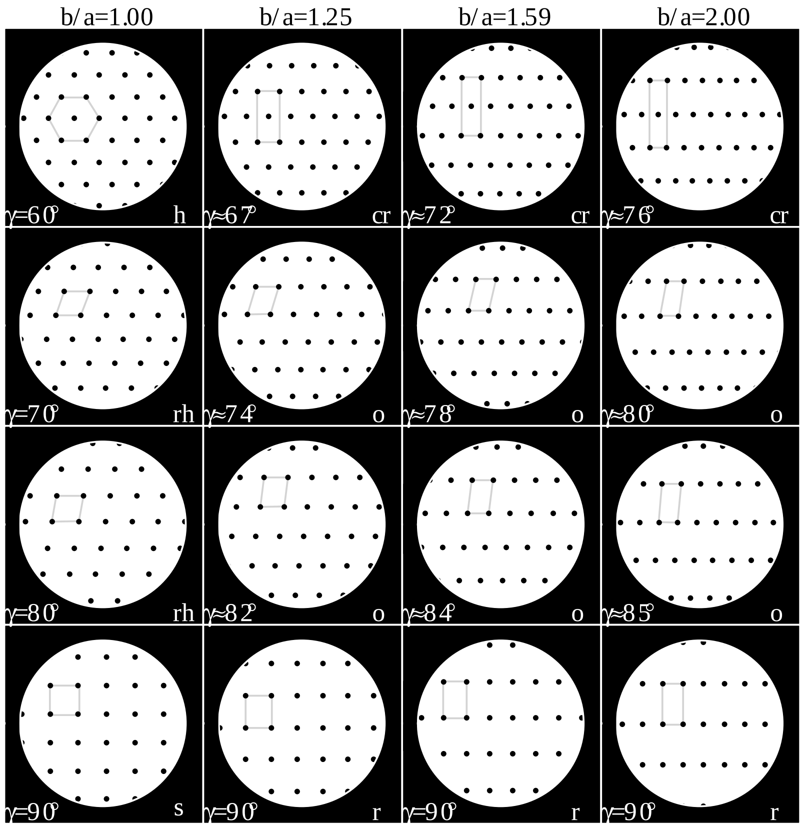


Figure 7. The sixteen dot lattices used by Kubovy & Wagemans (1995): *h*—hexagonal; *cr*—centered rectangular; *s*—square; *r*—rectangular; *o*—oblique. At the top of the figure, the $|\mathbf{b}|/|\mathbf{a}|$ ratio. In the lower left-hand corner of each panel, the value of γ.

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 logarithm 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 12.

The results, were encouraging (Figure 13), 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.

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 Gestaltish interaction?

The data collected by Kubovy and Wagemans (1995) were ideally suited to answering these questions. The stimuli had been sampled (Figure 8) 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 5). As the reader will recall (see Figure 6), 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 γ, or on the identity of the vector ($\mathbf{a}, \mathbf{b}, \mathbf{c}$, or $\mathbf{d}$), perhaps these dependencies would

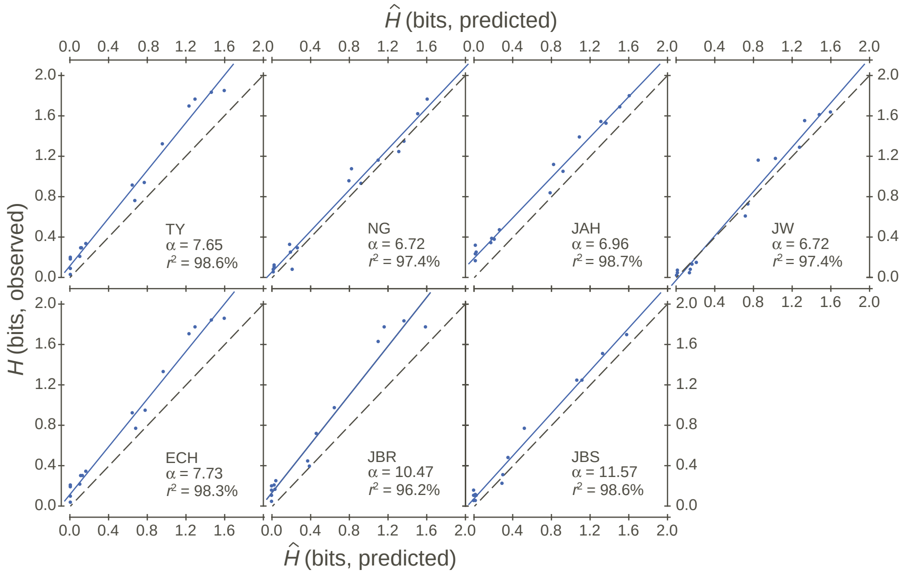


Figure 13. Observed entropy of responses as a function of entropy predicted by Kubovy& Wagemans's model (1995) for seven observers. (Figure after Kubovy & Wagemans (1995), Figure 11.)

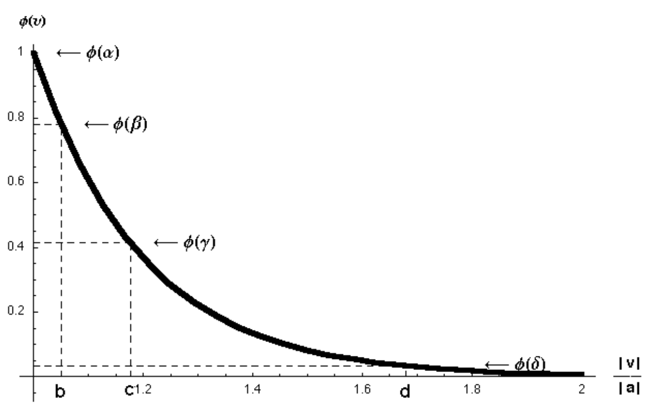


Figure 10. Grouping strength in the theory proposed by Kubovy & Wagemans (1995). Here we have set the slope of the grouping strength function to $s = 5$. We also illustrate how one calculates the four grouping strengths for an oblique lattice in which $^{|\mathbf{b}|}/_{|\mathbf{a}|} = 1.05$ and $\gamma = 70°$. 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 ϕ are $\phi(\beta) \approx 0.78$, $\phi(\gamma) \approx 0.41$, and $\phi(\delta) \approx 0.034$ ($\phi(\alpha)$ is always 1.0).

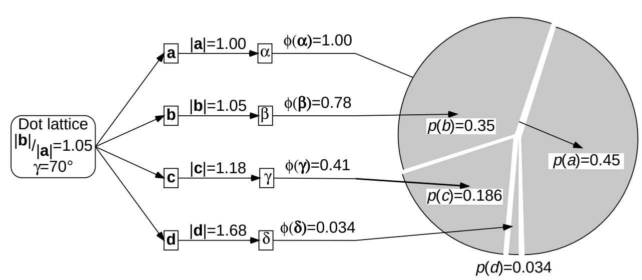


Figure 11. How $p(v)$ is computed, according to the model of Kubovy & Wagemans (1995). The parameters are the same as those in Figure 10.

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 probabilities 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 unforeseen consequence of the geometry of lattices. If one holds $|\mathbf{a}|$ and γ 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 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

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

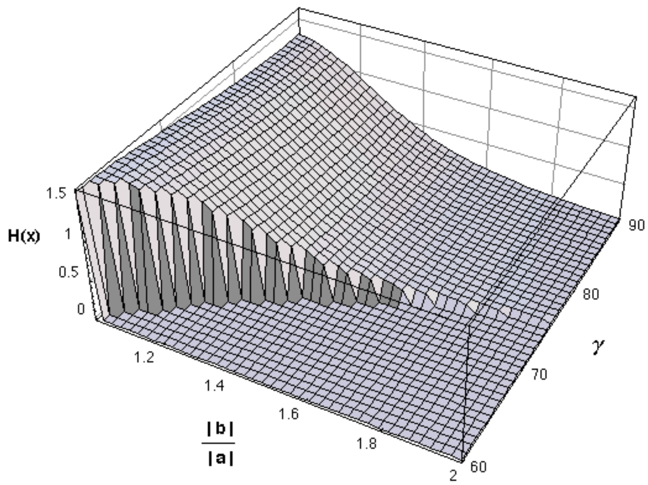


Figure 12. Entropy of responses as a function of $|\mathbf{b}|/|\mathbf{a}|$ and γ, as predicted by Kubovy & Wagemans (1995). The x and y axes are the same as in Figure 5. Note that where entropy is not defined (outside the curved boundary of the space of dot lattices) we let $H(x) = 0$. The value of s $(= 5)$ is the same as in Figure 10.

errors.

Figure 14 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 γ, 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's the Gestalt?. The Gestalt psychologists were interested in emergent properties, in phenomena where the whole is different from the sum of the parts. Grouping by proximity is indeed an emergent property: it is not a property that holds for sets of dots smaller than say a 4×4 lattice. Nevertheless, the pure distance law is as simple a law as we can imagine. We can model the law by assuming that the lattice activates four independent units whose outputs jointly determine the probability of a winner-take-all percept (we see only one organization at a time). This independence is not in the spirit of the complex interactive processes we have come to associate with Gestalt-like theorizing.

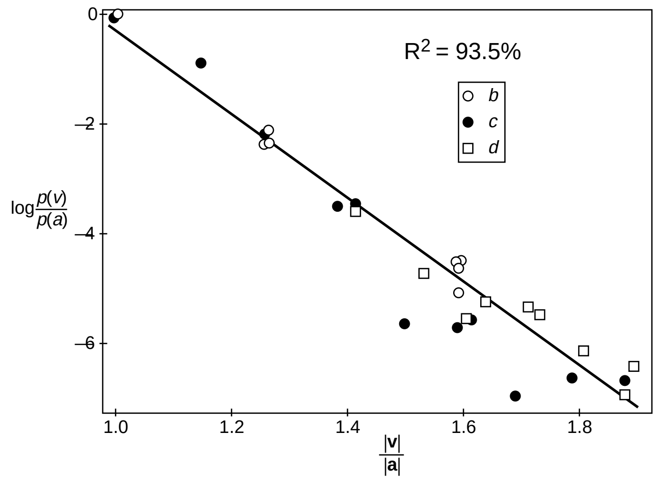


Figure 14. The pure distance law for dot lattices. Average data for the seven observers.

Grouping by proximity in spacetime

In the preceding section we saw that grouping by spatial proximity is decomposable into separable mechanisms, does not require the kind of holistic system the Gestalt psychologists proposed. Now we turn to the study of apparent motion (AM), a prototypical Gestalt concern (Wertheimer, 1912). Is AM decomposable into separable mechanisms: grouping by spatial proximity and grouping by spatiotemporal proximity?[7] The answer is that we cannot.

Sequential and interactive models

What is the relation between grouping by spatial proximity and grouping by spatiotemporal proximity? In this section of our chapter we consider two mutually exclusive classes of models:

Sequential models (SMs): Grouping by spatial proximity and grouping by spatiotemporal proximity are separable and serial, so that matching units are determined by their spatial proximity, independently of their spatiotemporal proximity. As we will see, this model can account for many phenomena of AM and is often tacitly taken as the default model.

Interactive models (IMs): Grouping by spatial proximity and grouping by spatiotemporal proximity are inseparable: the latter can override former; matching units are sometimes derived by the combined operation of both processes.

Although many phenomena of AM are consistent with SMs, we will review evidence (including an experiment of our

[7] We defined these terms and others that we will need, in the section *A clarification of terms*.

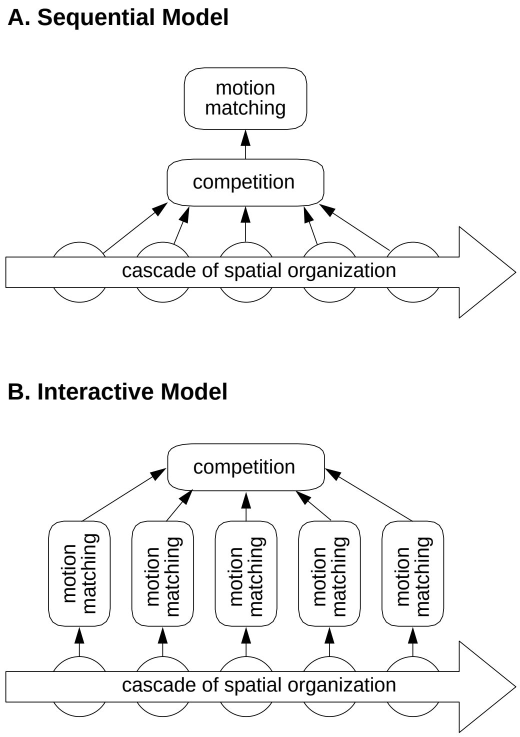


Figure 15. A. In the SM, alternative spatial representations compete, so that the most salient one undergoes temporal grouping. Spatial grouping alone determines *what* is moving (see also Figure **??**). B. In IMs, outputs of motion matching operations compete, so that both spatial and temporal grouping 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.)

own) which shows that AM is better described by IMs than by SMs.

We assume (following McClelland, 1979) that in processing a scene the visual system constructs a "cascade" (a hierarchical set) of spatial representations. In such a cascade, (a) more complex representations may contain entities from less complex ones, (b) and each representation may interact both with more and less complex representations. Alternative representations emerge concurrently and can be accessed in parallel, as soon as they become available.

In SMs, mechanisms of temporal grouping can access alternative spatial representations in parallel, so that the most salient spatial organization becomes a matching unit. Matching units can be thought of as "sliders" along the cascade of spatial perceptual organization because spatial entities of arbitrary complexity can serve as matching units (Figure 15A). The most salient spatial organization determines *what* is seen to move.

The IM implies that *both* spatial and temporal grouping determine the level of spatial organization at which matching units arise. Competition occurs between the outputs of parallel motion matching operations applied to different levels in the cascade of spatial organization (Figure 15B). Thus, the salience of both spatial and temporal grouping contributes into the formation of matching units.

It is obvious that IMs preclude giving primacy to grouping by spatial proximity or grouping by spatiotemporal proximity. It may not be as obvious that SMs does not require giving priority to grouping by spatial proximity. For example, according to Neisser's (1967) early account of AM, motion perception integrates successive "snapshots" of the scene: grouping by spatial proximity has priority—it alone determines which visual elements undergo motion matching. But this cannot be the whole story: motion matching may precede grouping by spatial proximity. This is the case in random-dot cinematograms (RDCs), where each frame contains a different random texture. If we introduce a correlation between successive frames, so that a compact region of elements, f, moves from one frame to the next while retaining its texture (and the remaining dots are uncorrelated between frames), we see f segregated from the rest of the display, even though none of the individual frames is distinguishable from random texture. This is possible only if motion matching can occur *before* grouping by spatial proximity. The Gestalt psychologists referred to such phenomena as grouping by *common fate* (Wertheimer, 1923, who did not, however, have such unambiguous examples as RDCs).

Evidence that seems to favor interactive models, but doesn't

As the exploration of AM progressed, increasingly complex motion displays have been studied. We now review some of these, and show that the complexity of an AM display does not count as evidence against SMs.

SMs *and Ullman's theory of matching units*. Our first illustration is Ullman's (1979) "broken wagon wheel" demonstration Figure 16. Every other spoke is interrupted in the middle. The angle between the neighboring spokes is α (Figure 16, left). If between frames we rotate the spokes counterclockwise by an angle $\beta > \frac{\alpha}{2}$ one sees three rotating objects. The outer segments of the spokes are seen moving clockwise. The same is true of the inner segments of the spokes. In addition one sees a counterclockwise motion of the gaps. This result could be taken as evidence against SMs, because *entities that are not present in the static image are created in the* AM *display*. However, we should not commit the *isomorphism error*, the error of thinking that what we experience is necessarily isomorphic with the underlying process. 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 we can explain this effect in the spirit of SMs. (As we will see, Ullman's model is probably too restrictive since complex organizations can serve as matching units.)

The second illustration is the *aperture problem* (Wallach, 1935; Wallach & O'Connell, 1953; Hildreth, 1983). Whenever a line moves behind an aperture that occludes its endpoints, we see motion orthogonal to the line (illustrated on

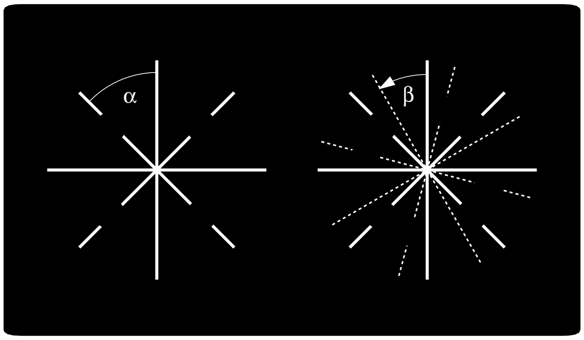


Figure 16. The "broken wagon wheel" demonstration of Ullman (1979).

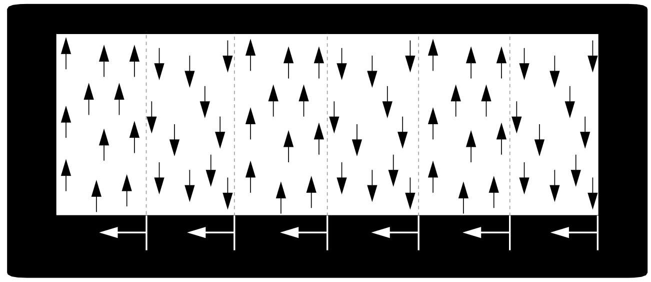


Figure 17. Second-order motion as an example of recursive grouping (Cavanagh and Mather, 1990).

the left side of Figure 17). 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*.

Ullman's (1979) model, which is a SM, can predict the visual system's solution to the aperture problem (e.g., Figure 16).

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 the contour of the ellipse. She then showed that the effect of aspect ratio is predicted by a system that finds the smoothest velocity field that maps successive contour fragments onto each other.

According to Ullman and Hildreth the first stage of the matching process locates spatial primitives which then become matching units: this is a SM; temporal grouping can have no influence on this process. Although the Ullman-Hildreth approach is parsimonious, the data for which it account are not inconsistent with IMs.

Recursive grouping. Matching units can be derived by the grouping by spatial proximity of entities which in turn are derived by grouping by spatiotemporal proximity. Such matching units are part of a hierarchical perceptual organization, in which elements move within moving objects. Such cases are easily described by SMs.

On such example is*grouping by common fate* (Wertheimer, 1923): elements extracted by grouping by spatiotemporal proximity are segregated from the background and form a moving figure. It occurs for both translation (Wertheimer, 1923) and rotation (Julesz & Hesse, 1970). The resulting elements may be subject to further spatial organization, which might produce, for example, a three-dimensional object (*shape-from-motion* Ullman, 1979). This phenomenon is consistent with SMs because the matching units are derived by grouping by spatial proximity alone (S_1), which is followed by grouping by spatiotemporal proximity (T_1). T_1 determines the directions and the velocities of the elements which is used by a subsequent grouping by spatial proximity (S_2) to derive the objects' shape. The recursive operation is $S_1 \longrightarrow T_1 \longrightarrow S_2$.

Cavanagh and Mather (1990) produced another instance of recursive grouping. They created a stimulus composed of a set of adjacent vertical bands, in each of which randomly positioned short-lived elements move vertically. Adjacent bands contain elements moving in opposite directions (Figure 17). The boundaries between these band are easily visible. When they are made to drift to the left, observers readily see the motion. A recursive SM would describe the phenomenon as follows. S_1: the short-lived random elements are output by grouping by spatial proximity, which cannot do much grouping because the elements in each frame are random. T_1: the elements in each frame are matched by grouping by spatiotemporal proximity and identified as dots moving up or down. S_2: dots moving in the same direction undergo grouping by spatial proximity (grouping 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). The output of T_2 does not depend on the fact that the boundaries between the strips are derived by T_1 which is a grouping by spatiotemporal proximity. These boundaries could have been produced by grouping in space by luminance or color. The recursive SM is: $S_1 \longrightarrow T_1 \longrightarrow S_2 \longrightarrow T_2$.

Matching of groupings. We have seen that matching units can be spatial primitives or spatial aggregates of similar moving spatial primitives. Can SMs account for 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 (Fig-

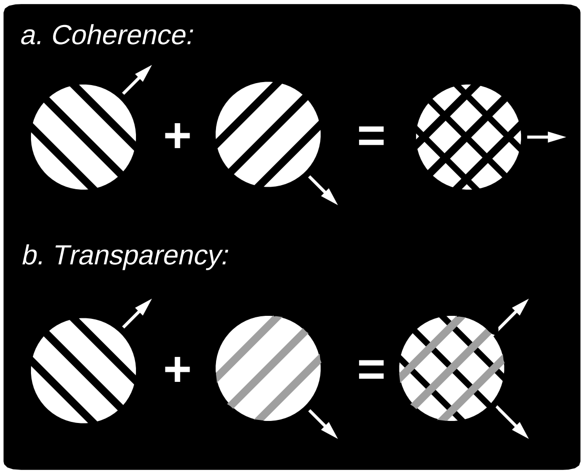


Figure 18. The Adelson & Movshon (1982) plaids.

ure 18). When either moving grating is presented alone, it is seen moving at right angles to the orientation of its bars, which is the visual system's solution of the aperture problem. When the superimposed gratings are identical (as in Figure 18*a*), the gratings are fused and seen as a single plaid moving in an orientation different from the motion of the individual gratings. However, when the superimposed gratings are sufficiently different (as in Figure 18*b*), they are not fused; they are seen as two overlaid gratings, each moving at a right angle to the orientation of its bars. 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. This is consistent with a SM.

SMs are applicable even when the components of a figure do not overlap but still group in space. Consider, for example, displays (Shiffrar & Pavel, 1991) in which a rectangle moves vertically behind an opaque screen, seen through circular apertures (Figure 19; Ben-Av & Shiffrar, 1995). When only one edge of the rectangle is visible (the black solid line in Figure 19A) through the circular aperture (labeled "target aperture"), its motion is orthogonal to its orientation, the default solution of the aperture problem. When a single corner is visible, it is seen to move vertically. Ben-Av and Shiffrar asked whether the motions of the corners can capture (or disambiguate) the motion of the edge, when two corners and an edge are visible. They found that the motion of the corners did capture the motion of the edge (and produce veridical vertical motion) 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 19B) was short. When the corners were collinear but remote, or when they were not collinear, no matter how close (Figure 19C), the motion of the edge was not affected by the motion of the corners; the edge appeared to move orthogonal to its orientation. The findings of Ben-Av and Shiffrar are consistent with SMs: corners and edges group into matching units. When the visible components of the rectangle are collinear and close to each

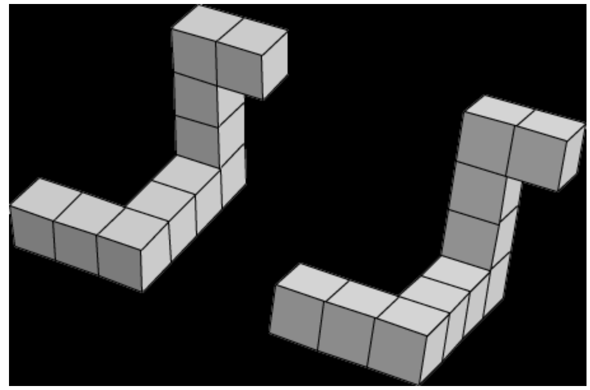


Figure 20. When the two images are shown in rapid alternation, observers see a rotating three-dimensional object (Shepard & Judd, 1976).

other, they group in space, so that grouping by spatiotemporal proximity occurs between the composite matching units.

Matching of high-level units. Organizations more complex than aggregates of similar elements can become matching units. These phenomena too are consistent with SMs. For example, Shepard and Judd (1976) rapid alternation of two images of a three-dimensional object (such as in Figure 20) looks like the object is rotating in depth. To derive this motion grouping by spatiotemporal proximity must match homologous parts of the object, rather than small spatial primitives (Rock, 1988, p. 57). According to SMs, it was the grouping by spatial features of the frames that derived the complex matching units in the displays of Shepard and Judd.

The Shepard and Judd displays suggest that if grouping by spatiotemporal proximity had matched small-scale entities, the percept would have been different. Ramachandran, Armel, and Foster (1998) created a display which showed just that. They created pairs of fragmented patterns, called "Mooney faces," that are sometimes seen as a face, and sometimes as a random pattern (Figure 21). Observers who saw the pattern as a face, experienced motion in a different direction from the one specified by matching the individual fragments. As in Shepard and Judd's display, the grouping by spatial features within the frames derives complex matching units, and hence this phenomenon is consistent with a SM. However, Ramachandran et al. go further: they show 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's have show interactions between object familiarity and grouping by spatiotemporal proximity: Shiffrar & Freyd, 1990; McBeath, Morikawa, & Kaiser, 1992; Tse & Cavanagh, 2000.)

Form and AM. AM 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

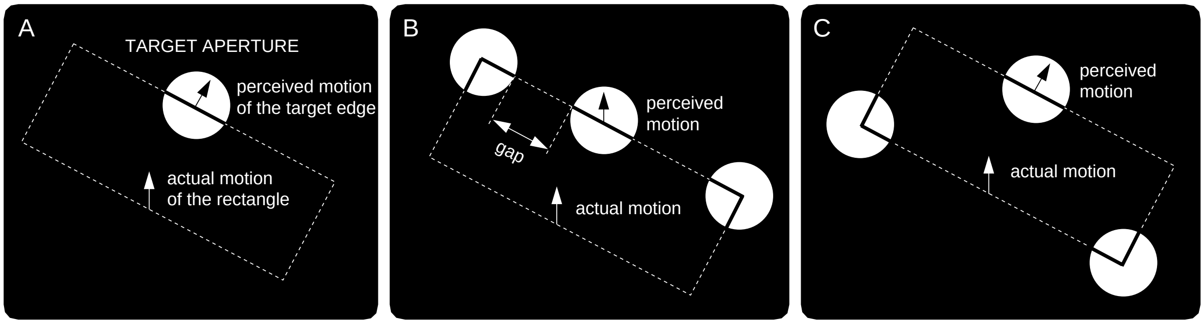


Figure 19. An outline of a rectangle (the white dotted outline in the diagrams) moves *vertically* behind an opaque screen (Ben-Av & Shiffrar, 1995). Panel B: the only configuration that produces veridical motion perception.

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. Such displays always support SMs.

SMs have also been assumed in studies of the interaction of form and AM. In this literature, researchers assume that vision derives the form of an object before the grouping by spatiotemporal proximity between the objects takes place. Because of this bias in favor of SMs, the question of form-motion interaction has been generally posed in a way that excludes IMs: Do form properties of moving objects affect grouping by spatiotemporal proximity? An answer to this question has been sought in two directions: (a) The *similarity* of an object's form across successive views of it (e.g., Orlansky, 1940; Kolers, 1972; Burt & Sperling, 1981; Oyama, Simizu, & Tozawa, 1999). (b) The *transformational* relations between successive forms (e.g., Warren, 1977; Eagle, Hogervorst, & Blake, 1999). In neither of these directions has a consensus been regarding the sensitivity of grouping by spatiotemporal proximity is sensitive to form differences of the grouped entities.

The distinction between SMs and IMs has consequences for research on form and AM. If IMs are 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 then we find the conditions under which the form of nascent objects affects the interactions between concurrent and successive elements.

Figure 21. The two images of "Mooney faces" are shown in rapid alternation. When observers see a face, they perceive it rotating in depth. When they do not, they perceive incoherent motion in the picture plane (Ramachandran *et al.*, 1998).

In a study we will presently describe, we (Gepshtein & Kubovy, 2000) show that spatial form affects motion matching when the spatial distances between concurrent elements are large enough to compete with the spatial distances between successive elements. One could also demonstrate the influence of spatial form on AM by reducing the spatial distances between successive elements, to the point that successive elements overlap. This approach has been adopted in the ingenious "transformational AM" displays by Tse and colleagues (Tse, Cavanagh, & Nakayama, 1998; Tse & Logothetis, 2001, in press).

Seeming evidence against SMs. The Ternus display has been offered as evidence against the SM. In this section we show that it is not. In this display (Ternus, 1936) dots occupy three equally-spaced collinear positions (Figure 1C[i]–[iii]). These displays consist of two rapidly alternating frames, represented by two vertical dotted lines in Figure 1. The dots in one frames are ❶ and ❷; the dots in the other frame are ❸ and ❹. 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 ❶ and ❹, and dot ❷ appears immobile when replaced by dot ❸ (Figure 1C[i]); (b) *Group motion* (*g*-motion), which occurs when two dots appear to move back-and-forth as a group, from ❶❷ to ❸❹ (Figure 1C[ii]).

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 grouping by spatiotemporal proximity 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 SMs for two reasons:

1. Longer ISIs could have two effects: (i) they could weaken grouping by spatiotemporal proximity, as Kramer and Yantis assumed, or (ii) they could allow more time for grouping by spatial proximity to consolidate the organization of concurrent dots. If the latter is true, we could attribute the ISI effect to grouping by spatial proximity rather than to grouping by spatiotemporal proximity, and conclude that the ISI effect is consistent with SMs.

2. If an observer sees *g*-motion, one cannot tell whether the matching units were dots or dot groupings, because in either case matching yields motion in the same direction (Figure 1C[iii]). Therefore, the group motion percept may actually be the result of matching of individual dots, just as in *e*-motion; 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.)

Evidence for interactive models

The only types of motion perception in the current literature that truly undermine the generality of SMs involves overlapping objects and surfaces whose relation is changing dynamically, which we call *dynamic superposition*. The perception of dynamic superposition poses a challenge to SMs 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 et al., 1998). In such cases a simple correspondence between 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. 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 SMs, it is too specific to carry the burden of refuting SMs in favor of IMs. 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).

To refute the class of SMs, 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. That is we conducted a study (Gepshtein & Kubovy, 2000) in which we tested SMs using spatiotemporal dot lattices, called *motion lattices*.

Motion lattices. Motion lattices allowed us to independently vary the strength of grouping by spatial proximity and grouping by spatiotemporal proximity by manipulating spatial proximity between concurrent and successive dots (Figure 22).

As we observed earlier (with regard to Ternus displays), the duration of the ISI does not necessarily determine the strength of grouping by spatiotemporal proximity, 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 our motion lattices we held ISI constant, and varied the strength of grouping by spatiotemporal proximity by manipulating the *spatial proximity between successive dots*.

Why not use other AM displays? First, consider Ternus displays, in which—as in motion lattices—either element motion (*e*-motion) or group motion (*g*-motion) can be seen. In Ternus displays, however, the directions of *e*-motion and *g*-motion do not differ. In motion lattices the direction of *e*-motion is determined by matching individual dots in successive frames of the display, whereas the direction of *g*-motion is determined by the matching of dot groupings (strips of dots, or *virtual objects*) in successive frames. In motion lattices the direction of *g*-motion is orthogonal to the orientation of the objects, which is different from the direction of *e*-motion.

Second, consider 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, they displaced the row both horizontally and downward, so that under appropriate conditions observers saw the it moving downward and to the right (or to the left). Burt and Sperling studied the trade-off between space and time in motion matching, and the effect of element similarity on matching. But their stimulus did not allow them to explore the effect of

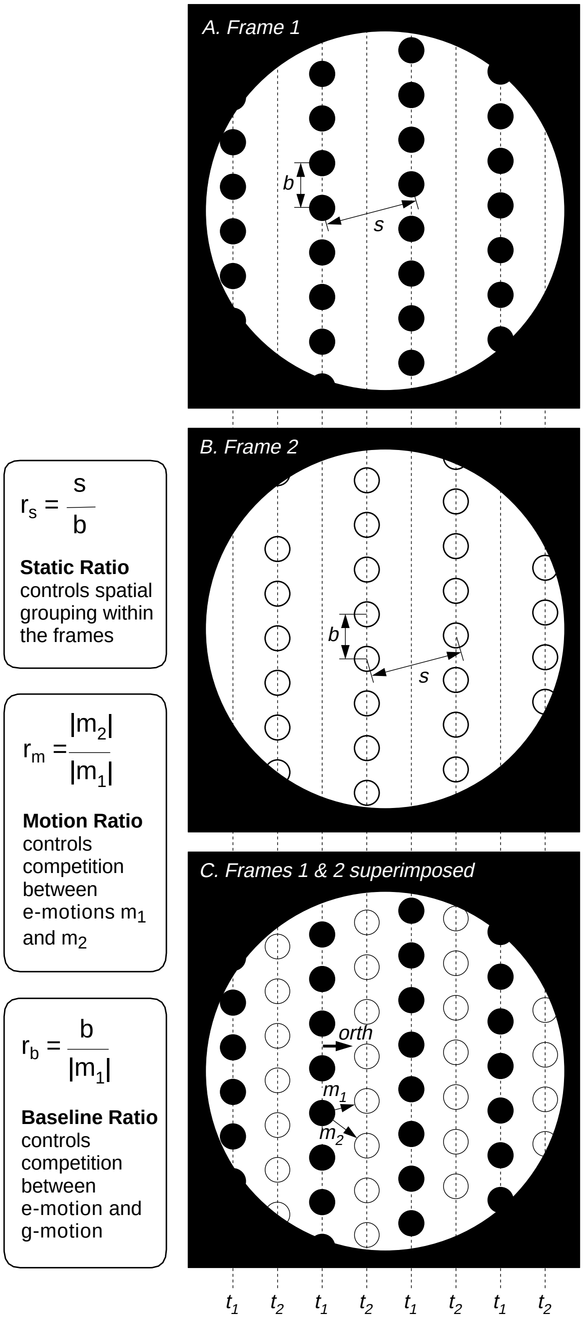


Figure 22. The design of motion lattices. Two frames of a motion lattice are shown schematically in *A* and *B*; the frames are superimposed in *C*. Distances *b* and *s* correspond to the shortest inter-dot distances within the frames (shown in *A and B*). Vectors m_1 and m_2 (shown in *C*) are the 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.

relative proximity between concurrent and successive dots. Motion lattices allowed us to set up a competition between alternative *spatial* organizations within a frame, because in these stimuli each frame contains a two-dimensional pattern of dots. We thus could ask whether grouping by spatiotemporal proximity affects grouping by spatial proximity.

A critical test of the SM *(Gepshtein & Kubovy, 2000)*. 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, grouping by spatial proximity within static dot lattices is only 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. We used this property of static dot lattices to test the SM by asking whether the likelihood of *g*-motion is affected by variations in the proximity between successive dots when the proximity between concurrent dots is held constant.

Figure 22 describes our motion lattices. We obtained them by splitting static lattices into two frames, so that every frame contains every other column (or row) of the original lattice (we call them *two-stroke motion lattices* M^2). When the frames of a motion lattice are shown in rapid alternation with the appropriate spatial and temporal parameters, observers see a flow of AM. When they report dots flowing in a direction of matching between individual dots, we say that they are seeing *e-motion*. When they report dots flowing in a direction orthogonal to virtual objects formed within the frames, we say that they are seeing *g-motion* .

In Figure 22, the dots are likely to group into vertical virtual objects within frames. If grouping by spatiotemporal proximity across frames occurs between virtual objects, rather than between dots, observers see motion orthogonal to the virtual objects (i.e., horizontal motion in Figure 22). Fig. 23 (*A*–*B*) shows frames from an M^2 motion lattice in which horizontal *g*-motion is likely. If we arrange dots within the frames so that virtual objects are less salient, *g*-motion is less likely. For example, the two frames shown on panels *C*–*D* of Figure 23 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 SMs, the likelihood of seeing *g*-motion rather than *e*-motion depends on the propensity of concurrent dots (within the frames) to form virtual objects, and does not depend on grouping by spatiotemporal proximity. Because SMs hold that matching units are derived by grouping by spatial proximity alone, only spatial proximities between concurrent dots determine the likelihood of whether *e*-motion or *g*-motion is seen. In contrast, IMs hold that the grouping by spatial proximity of dots within frames is affected by grouping by spatiotemporal proximity between successive dots.

To pit the models against each other, we measured the relative frequency of reports of *e*-motion and *g*-motion under conditions of equivalent spatial grouping within frames. Within frames, the salience of virtual objects cannot change as long as the ratio between relevant spatial distances is in-

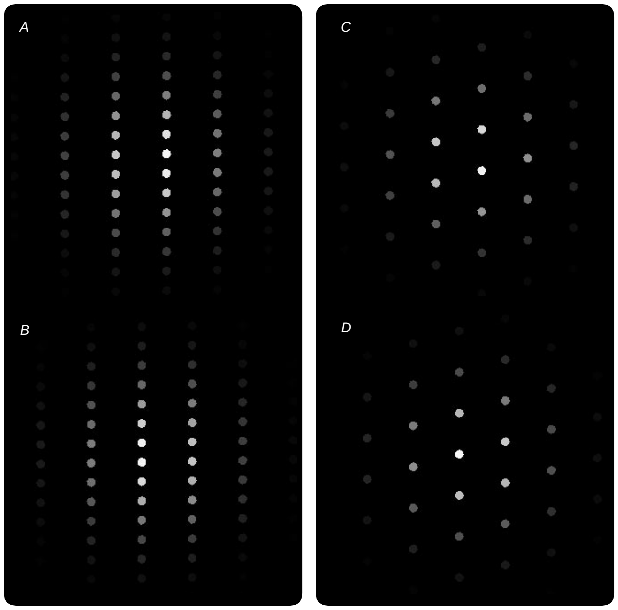


Figure 23. Frames of a motion lattices (not to scale). *A-B*. Two frames of a lattice in which g-motion is likely (high r_s and low r_b). *C-D*. Two frames of a lattice in which e-motion is likely (low r_s and high r_b).

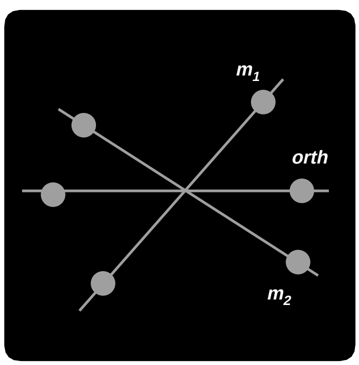


Figure 24. A response screen corresponding to the lattice shown in Figure 23 *C-D* (not to scale). Observers click on a circle attached to the radial line parallel to the perceived direction of motion. (Response labels m_1, m_2, and *orth* did not appear on the response screen.)

variant. Thus, as long as $r_s = \frac{s}{b}$ (Figure 22) does not change, the propensity of dots to form vertical virtual objects in does not change. In our experiment we holding r_s constant while varying the strength of grouping by spatiotemporal proximity. Under these conditions, according to SMs, the likelihood of seeing g-motion rather than e-motion should not change, but according to IMs the likelihood of g-motion should drop.

Our experiments supported IMs. The pie charts of Figure 25 show the distributions of three responses—m_1, m_2, and *orth*—for different configurations of motion lattices. 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.

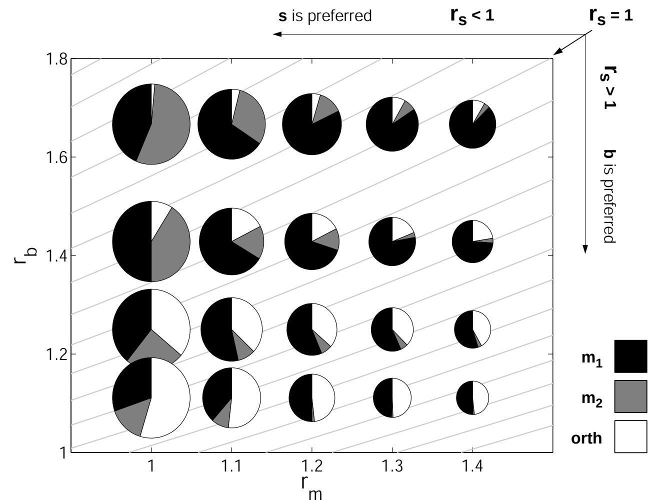


Figure 25. Results of our experiment. The pie charts on the top panel show the distributions of three responses (m_1, m_2, and *orth*) for twenty configurations of motion lattices. The gray lines in the background are iso-r_s lines; within these lines the salience of spatial virtual objects is invariant (see text).

3. The frequency of *orth* motion varies within the sets of iso-r_s conditions, marked with oblique gray lines.

We can explain the third observation in two steps:

1. We constructed a statistical model of the data shown in Figure 25. (The model accounted for 98% of variance in the data.)

2. We interpolated motion frequencies within the iso-r_s sets of parameters. A result of this computation is shown in Figure 26, where each curve plots the relative frequency of g-motion and e-motion within a corresponding r_s set.

Figure 26 shows that when r_s is big (i.e., when $s \gg b$, as in

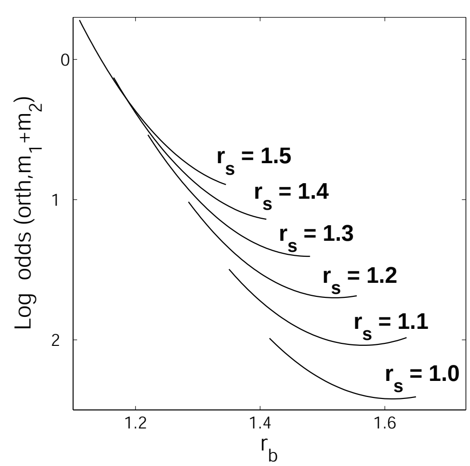


Figure 26. Results of our experiment. The relative likelihood of g-motion (*orth* responses) and e-motion (m_i responses) changes within the iso-r_s sets of conditions, in contrast to the prediction of the sequential model (see text).

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 Figures 22 and 23), 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, 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 demonstrates the validity of IMS.

Where's the Gestalt?. AM is an emergent property, just as is grouping by proximity. We have found that it is impossible to decompose motion perception into two successive grouping operations: grouping by spatial proximity and grouping by spatiotemporal proximity. This complexity is more in the spirit of Gestalt theories than the Kubovy & Wagemans model presented in the first part of this chapter.

In praise of phenomenological psychophysics

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 *phenomenological psychophysics*.[8] All the experiments we describe in this chapter belong to this category.

Although phenomenological psychophysics 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 skepticism 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. 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."

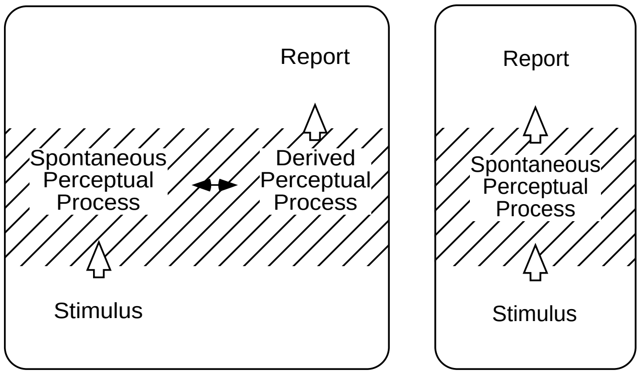
Subjective = idiosyncratic: "Peculiar to a particular individual . . . arising from conditions within the brain or sense organs and not directly caused by external stimuli."

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.

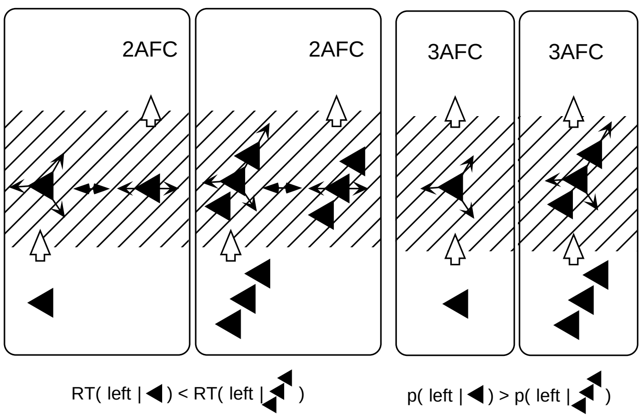
---

[8] 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 phenomenological psychophysics, responses of different kinds can be counted, and therefore statistics may be applicable. They may or may not lend themselves to mathematical modeling.



(a) Traditional psychophysical procedure

(b) Phenomenological psychophysics procedure



(c) The Palmer & Bucher (1981) experiment as a traditional psychophysical procedure (2AFC). "RT" stands for reaction time.

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

Figure 27. 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 hatched 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 phenomenological psychophysics. In that sense the latter are more direct than the former.

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 phenomenological psychophysics can play in the study of perceptual organization.

As opposed to phenomenological psychophysics, traditional psychophysical tasks are indirect. This idea is illustrated in Figures 27(a) and 27(b). In natural viewing conditions, as well as in the tasks used in phenomenological psychophysics, 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 laboratory. The hatched 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."[9]

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 27(c)). An equilateral triangle appears to point about equally often at $60^\circ, 180^\circ$, or 300° (Figure 27(c), left). If you align three such equilateral triangles along a common axis of mirror symmetry tilted 60° (Figure 27(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 27(c), left). Obviously, these triangles cannot point

[9] We recommend that the term for an indirect report be descriptive, such as "correct/incorrect task," rather than evaluative, such as "objective task."

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 27(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 27(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 27(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 27(d), right). For example, one could use an phenomenological psychophysics 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 27(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 make no sense without assuming the existence of a covert spontaneous organization, which under the appropriate eliciting circumstances would have led to the phenomenological report. Psychophysical studies of perceptual organization are no more than an *indirect assessment* of the effects of grouping. Hence, when available, we prefer experimental phenomenology.

This is not to say that tasks which have correct and incorrect responses can only serve to examine the epiphenomenality of a Gestalt phenomenon. When an organizational phenomenon is subtle or complex, such tasks may give us valuable information about underlying processes. Yet we hope that we have persuaded the reader that *indirect psychophysical methods do not have an intrinsic advantage over phenomenological methods*. Indeed, one of our goals in this chapter was 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.

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.

Gepshtein, S., & Kubovy, M. (2001). The weights of space and time in the perception of visual motion [abstract]. *Journal Of Vision*, *1*, 243a. (http://journalofvision.org/1/3/243, DOI 10.1167/1.3.243)

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.

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

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

McBeath, M. K., Morikawa, K., & Kaiser, M. K. (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.

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.

Tse, P. U., & Logothetis, N. K. (2001, in press). The duration of 3D form analysis in transformational apparent motion. *Perception & Psychophysics*.

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.