

**Record: 1**

**Article ID:** rev984569

**Title:** The How and Why of What Went Where in Apparent Motion: Modeling Solutions to the Motion Correspondence Problem

**Author(s):** Michael R. W. Dawson *University of Alberta, Edmonton, Alberta, Canada*

**Source:** Psychological Review. Vol. 98 (4) October 1991, pp. 569-603. American Psychological Association

**ISSN:** 0033295X

**Digital Object ID:** 10.1037//0033-295X.98.4.569

**Article Type:** Journal Article

**Abstract:** A model that is capable of maintaining the identities of individuated elements as they move is described. It solves a particular problem of underdetermination, the motion correspondence problem, by simultaneously applying 3 constraints: the nearest neighbor principle, the relative velocity principle, and the element integrity principle. The model generates the same correspondence solutions as does the human visual system for a variety of displays, and many of its properties are consistent with what is known about the physiological mechanisms underlying human motion perception. The model can also be viewed as a proposal of how the identities of attentional tags are maintained by visual cognition, and thus it can be differentiated from a system that serves merely to detect movement.

**PDF Image:**  PDF Full Text

**Database:** PsycARTICLES

## **The How and Why of What Went Where in Apparent Motion: Modeling Solutions to the Motion Correspondence Problem**

**By: Michael R. W. Dawson**

*University of Alberta, Edmonton, Alberta, Canada*

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Operating Grant A2038, by Equipment Grant 46584, by a grant from the Central Research Fund of the University of Alberta, and by a Province of Alberta Summer Temporary Employment Program (STEP) grant—all awarded to Michael R. W. Dawson—and by NSERC Operating Grant A2600 awarded to Zenon Pylyshyn. It has benefitted greatly from discussions with Zenon Pylyshyn and Richard Wright at the University of Western Ontario and from discussions with Vince Di Lollo, Walter Bischof, Charles Bourassa, Bill Rozeboom, Don Kuiken, Ngaire Nevin-Meadows, Don Schopflocher, and Nancy Digdon at the University of Alberta. Comments from Dennis Proffitt and an anonymous reviewer on an earlier version of the manuscript were also extremely useful. I would like to acknowledge Brian Harder, who died during the past year. He began work on this research with me and performed numerous simulation runs and prepared many of the figures.

Correspondence may be addressed to: Michael R. W. Dawson, Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9 Canada. Electronic mail may be sent to [mike@psych.ualberta.ca](mailto:mike@psych.ualberta.ca).

Many researchers have described the goal of visual perception as the construction of useful representations about the world (e. g. , [Horn, 1986](#); [Marr, 1976, 1982](#); [Ullman, 1979](#)). These representations are derived from the information projected from a three-dimensional visual world (the distal stimulus) onto an essentially two-dimensional surface of light receptors in the eyes. The interpretation of the distal stimulus must be determined

from the resulting pattern of retinal stimulation (the proximal stimulus).

However, the information represented in the proximal stimulus cannot, by itself, completely determine the nature of the distal stimulus. This is because the proximal stimulus does not preserve the full dimensionality of the physical world. The mapping from three-dimensional patterns to two-dimensional patterns is a many-to-one mapping and is not uniquely invertible (see [Gregory, 1970](#); [Horn, 1986](#); [Marr, 1982](#); [Richards, 1988](#)). Retinal stimulation geometrically underdetermines interpretations of the physical world. <sup>1</sup>

Underdetermination can also result because the information available from local measurements of the proximal stimulus is consistent with a large number of different global interpretations. Alone, the local measurements are not sufficient to determine which global interpretation is correct. One example of this is the aperture problem: local measurements of a contour's movement do not by themselves specify the contour's true velocity (e. g. , [Hildreth, 1983](#); [Marr & Ullman, 1981](#)). Another example is the stereo correspondence problem: local measurements do not by themselves specify which proximal stimulus element on the right and left retinas were produced by the same distal element (e. g. , [Grimson, 1981](#); [Marr & Poggio, 1976](#)).

A third example is called the *motion correspondence problem* (e. g. , [Attneave, 1974](#); [Ramachandran & Anstis, 1986b](#); [Ullman, 1979](#)), and in this article, a model for its solution is described as follows: First, the motion correspondence problem is defined, and three constraining principles to be used to generate solutions to this problem are discussed. Second, an model that applies these constraints is outlined. Third, the model's performance is used to focus discussion on several theoretical issues in the study of motion perception, including the relation between motion correspondence processing and the tag-assignment problem in visual cognition. Fourth, functional properties of the model are related to what is known about the physiological mechanisms that mediate motion perception by humans.

## Motion Correspondence Problem

The human visual system can produce the illusion of movement from a rapid succession of static images. <sup>2</sup> If elements depicted in these images are displaced a large amount (i. e. , a degree of visual angle or more), this illusory or apparent motion is usually presumed to be detected by the so-called long-range motion system (e. g. , [Anstis, 1978, 1980](#); [Braddick, 1974, 1980](#); [Petersik, 1989](#); [Ullman, 1981](#)). To generate apparent motion, the long-range system must identify an element in a position in one image (Frame 1) and another element in a different position in the next image (Frame 2) as constituting different glimpses of the same moving element. A *motion correspondence match* between a Frame 1 element and a Frame 2 element is such an identification.

In this article, an *element* is defined as an individuated component of the proximal stimulus. In other words, an element is some aspect of the proximal stimulus that can be referred to by a unique symbolic code (e. g. , a token, or a FINST [to be described] in the sense of [Pylyshyn, 1989](#)). [Ullman \(1979, Chap. 2\)](#) presented evidence suggesting that these tokens could represent components such as oriented parts of edges, corners, and terminators (i. e. , the type of information available in the primal sketch of [Marr, 1982](#)).

Measurements of Frame 1 and Frame 2 element positions underdetermine the motion correspondence matches that can be assigned to an apparent motion display. Several sets of motion correspondence matches are consistent with the same set of positions. This is illustrated in

Motion correspondence as a problem of underdetermination. ([a] An apparent motion display. Outline squares represent element positions in Frame 1; solid squares represent element positions in Frame 2. [b-g] Possible motion correspondence solutions for this display, in which solid lines represent assigned motion correspondence matches. Solution b is generated by the human visual system. )

Figure 1. In general, if there are  $N$  elements in Frames 1 and 2, and if one assumes a one-to-one mapping between frames of view, then there are  $N!$  sets of correspondence matches that are consistent with the proximal stimulus ([Ullman, 1979](#)). If a one-to-one mapping between frames of view is not assumed, then the number of possible solutions increases to  $2N$ . In order to solve the motion correspondence problem, one set of motion correspondence matches (a *global interpretation*) must be selected from the many that are consistent with the position measurements.

Because measurements of element locations are not sufficient to solve the motion correspondence problem, additional rules or principles must be exploited. These principles, in combination with the local measurements, must determine the set of motion correspondence matches assigned to a display. The discovery of the principles used by the human visual system has been an important goal of apparent-motion research (e. g. , [Attneave,](#)

1974; Petersik, 1989; Ramachandran & Anstis, 1986b; Ternus, 1938; Ullman, 1979).

The model to be described solves the motion correspondence problem by exploiting principles that are based on measurements of element positions: principles that minimize changes in element positions over time and that minimize changes in element positions relative to one another. The model does not exploit principles that are based on measurements of the figural appearance of elements. There were three major reasons for designing a model that was sensitive only to element positions.

First, human observers can easily experience apparent motion for displays in which all elements are of identical appearance (e. g. , all are dots or lines). For such displays, the assignment of motion correspondence matches can be based only on measurements of element positions; therefore, nonfigural principles must be important determinants of correspondence match assignments. Second, many psychophysical experiments have shown that the human visual system assigns motion correspondence matches primarily on the basis of element locations and not on the basis of element appearances. Whereas human observers are very sensitive to manipulations of element positions in apparent motion displays, they are much less sensitive to manipulations of figural properties such as shape, color, or spatial frequency (e. g. , Baro & Levinson, 1988; Burt & Sperling, 1981; Cavanagh, Arguin, & von Grunau, 1989; Dawson, 1990a; Kolers, 1972; Kolers & Green, 1984; Kolers & Pomerantz, 1971; Kolers & von Grunau, 1976; Krumhansl, 1984; Navon, 1976; Ullman, 1979, Chap. 2; Victor & Conte, 1990). Third, physiological evidence supports the existence of at least two predominantly independent anatomical pathways in the visual system, one of which is sensitive to movement but not to form (e. g. , Botez, 1975; Livingstone & Hubel, 1988; Maunsell & Newsome, 1987; Ungerleider & Mishkin, 1982). For example, many cells located late in this pathway are very sensitive to movement, regardless of stimulus shape, size, or contrast (e. g. , Albright, 1984; Albright, Desimone, & Gross, 1984; Dubner & Zeki, 1971; Maunsell & van Essen, 1983b; Rodman & Albright, 1987; Zeki, 1974). Furthermore, lesions in this area produce significant deficits in motion perception but do not appear to affect object perception (Hess, Baker, & Zihl, 1989; Newsome & Pare, 1988; Newsome, Wurtz, Dursteler, & Mikami, 1985; Zihl, von Cramon, & Mai, 1983). In sum, the psychophysical and physiological evidence indicates that element appearances play, at best, a minor role in the assignment of motion correspondence matches and therefore should not represent a major component of a motion correspondence model. <sup>3</sup>

## Part 1: Constraining Solutions to the Motion Correspondence Problem

Marr (1976, 1982) described a theory of a computation as an account of what is being computed and why. For a visual problem of underdetermination, such a theory describes a constraining principle to be used to choose the correct proximal stimulus interpretation from the set of possible interpretations. A constraining principle usually characterizes or exploits an attribute of a distal stimulus. The principle is applied as follows: if the constraining property is characteristic of some distal stimulus that could have caused the proximal stimulus, then this is the distal stimulus that is perceived. For instance, Ullman (1979) showed that the property "being rigid" can be used to determine three-dimensional wire-frame interpretations of dynamic (two-dimensional) proximal stimuli (for several other examples, see Marr, 1982, Chap. 3).

Three possible constraining principles are considered in the following sections for the motion correspondence problem. For each, three questions are briefly addressed: Does the principle entail the use of a general characteristic of distal stimuli? Does experimental evidence suggest that the principle is exploited by the human visual system? Can the principle be successfully applied to the correspondence problem? It is then argued that an adequate model of human motion correspondence processing requires (at least) that all three principles be applied simultaneously.

### Nearest Neighbor Principle

An experimental technique called the *motion competition paradigm* has been used to study how the human visual system solves the correspondence problem (e. g. , Ullman, 1979, Chap. 2). In the simplest motion competition display, two opposing paths of apparent motion (i. e. , two opposing motion correspondence matches) compete with one another for assignment. In Frame 1 of such a display, a single element is presented in the center. In Frame 2, the Frame 1 element has disappeared, and two lateral elements are now displayed, one to the right of center and the other to the left (see

The nearest neighbor principle governs perceptions of a standard motion competition display. (a and b) The central Frame 1 element is seen to move in the direction consistent with the shorter motion correspondence

match. (c) When both possible matches are of equal length, they are equiprobable. (In many cases subjects will report seeing the Frame 1 element split into two. )

Figure 2). Under appropriate temporal conditions, the Frame 1 element is seen to move either to the left or to the right. Of interest are the factors that determine the perceived direction of motion.

In a competition display, a very strong predictor of the perceived direction of motion is element displacement (i. e. , the distance between a Frame 1 element and a potentially corresponding element in Frame 2). The visual system prefers to assign correspondence matches that represent short element displacements (e. g. , [Burt & Sperling, 1981](#); [Ullman, 1979](#), Chap. 2). For example, if motion to the left in a competition display involves a shorter element displacement than motion to the right, then motion to the left will be preferred ([Figure 2a](#)). The visual system exploits a "nearest neighbor" principle, in which motion correspondence matches are created between Frame 1 elements and their nearest neighbors in Frame 2.

The nearest neighbor principle is consistent with the geometry of the typical viewing conditions for motion ([Ullman, 1979](#), pp. 114-118). When three-dimensional motion vectors are projected onto a two-dimensional surface (e. g. , the retina), their depth component is lost. As a result, slower two-dimensional movements are much more likely to occur than are faster movements. Because a preference for the nearest neighbor is equivalent to a preference for slowest two-dimensional velocities, it may be that the human visual system exploits this constraint because it has evolved in a visual environment in which low velocities are more frequent than high velocities.

The nearest neighbor principle is used in [Ullman's \(1979\) minimal mapping theory](#) of motion correspondence processing. According to this theory, a cost is associated with each possible motion correspondence match. This cost is proportional to element displacement, so that shorter motion correspondence matches have lower costs. The model selects the set of motion correspondence matches that minimizes the total cost. (The solution must also be consistent with what Ullman called the cover principle, which is discussed in detail later in this article. ) A computer implementation of minimal mapping theory demonstrated, for many displays, that the nearest neighbor principle can be used to emulate the correspondence solutions of the human visual system (an alternative implementation of minimal mapping theory was described by [Grzywacz & Yuille, 1988](#)). However, minimal mapping theory cannot generate correct solutions for displays in which element interdependencies can play a role (e. g. , [Dawson, 1987](#); [Ramachandran & Anstis, 1985](#)). An additional constraining principle is required for such displays.

## Relative Velocity Principle

A major assumption underlying minimal mapping theory is that the cost assigned to any particular motion correspondence match in a display is independent of the costs assigned to any other possible motion correspondence matches ([Ullman, 1979](#), pp. 84-86). This assumption is questionable in principle because the perceptible world consists primarily of coherent surfaces whose properties vary smoothly (i. e. , neighboring points on the surface have, in the vast majority of cases, nearly identical visual properties, as described by [Marr, 1982](#), pp. 44-51). To the extent that visual elements arise from physical features on such surfaces, the movement of neighboring elements should be similar. If there were a general way of characterizing the interdependent properties of the two-dimensional motion of points projected from a coherent surface, the result would be a property that could be used to constrain motion correspondence solutions.

[Yuille \(1983\)](#) provided an account of one such property for an apparent motion display constructed from different views of a moving, continuous contour. This contour is represented as the set of unit vectors tangent to the contour at all its points. A motion correspondence solution for this display describes the transformation that produces the Frame 2 view of the contour from the Frame 1 view. Yuille argued that the desired transformation minimizes the distortion of the contour as it moves. This transformation therefore matches each Frame 1 tangent vector to a Frame 2 tangent vector, minimizing the differences between matched vectors over the entire contour.

[Yuille \(1983\)](#) demonstrated that his measure of figural distortion is very strongly related to another measure called *motion smoothness* ([Hildreth, 1983](#); [Horn & Schunk, 1981](#)). The smoothness of motion is measured by integrating differences between local velocities over the entire moving contour: motion is smoothest when neighboring points on a contour have nearly identical velocities. [Hildreth \(1983\)](#) showed that coherent objects moving arbitrarily in three-dimensional space produce unique, smooth patterns of retinal movement. The relation between figural variation and smoothness indicates that motion correspondence solutions that minimize figural distortion represent unique, physically plausible solutions.

Yuille's (1983) analysis can be used to generate a hypothesis about a constraining property for displays that consist of discrete visual elements (rather than continuous contours). A motion correspondence match between discrete elements can be described as a motion vector because the match is an assertion that a Frame 1 element has moved in a particular direction, at a particular speed, to occupy a new Frame 2 position. It is hypothesized that the visual system selects the set of motion correspondence matches that minimizes the relative velocities between neighboring display elements in an attempt to minimize local changes in the configuration. *Relative velocity* is defined as the difference between two motion correspondence matches (interpreted as motion vectors) after they have been centered at a common origin (

Defining the relative velocity between matches. (a) Two of four possible motion correspondence matches for a simple apparent motion display. (b) The same two matches represented as motion vectors centered at a common origin. (The distance between the endpoints of the vectors, indicated by the brace, is the relative velocity between the two matches. )

Figure 3). This constraint is called the *relative velocity principle* (Dawson, 1986; Dawson & Pylyshyn, 1986, 1988).

Dawson (1987) provided empirical support for the relative velocity principle; this empirical support shows that the independence assumption in minimal mapping theory is incorrect. Human observers were presented competition displays embedded in the context of a moving configuration (

The effect of element interdependencies on motion correspondence matches. (a) A control competition display in which two correspondence matches are equiprobable. (b and c) When the control display is embedded in an unambiguous context, there is a strong preference to see the Frame 1 component move in the direction of the context (see

Figure 4). The presence of a moving context had a pronounced effect on the perceived direction of the Frame 1 element in comparison with control displays without contexts. There was a strong tendency to see the central element and the context move in the same direction. This demonstrates that element interdependencies are important determinants of motion correspondence matches. This result is also consistent with several others showing that the human visual system minimizes patterns of relative motion for various discrete element displays (e. g. , Cutting & Proffitt, 1982; Gogel, 1974; Johansson, 1950; Proffitt & Cutting, 1979, 1980; Ramachandran & Anstis, 1985, 1986b).

Computer simulations have shown that many motion correspondence problems can be solved through the use of *only* the relative velocity principle (Dawson, 1986; Dawson & Pylyshyn, 1986, 1988; see also Barnard & Thompson, 1980). Furthermore, when this constraint is applied to displays in which element interdependencies are important (e. g. , Figure 3), the solutions generated are more similar to humans' solutions than those produced by minimal mapping theory. However, the relative velocity principle alone is not sufficient to solve a number of elementary motion correspondence problems that are solved by minimal mapping theory. For example, displays in which there is only a single element in both frames have no relative velocity information at all. This indicates that motion correspondence solutions must be constrained by the use of relative velocity information in combination with other principles (e. g. , nearest neighbor).

## Element Integrity Principle

Experimental studies of motion perception suggest that the human visual system prefers one-to-one mappings between elements in different frames of view.

Preference for one-to-one matchings by the human visual system. (a and b) Correspondence solutions generated by the human visual system. (c) An incorrect solution for Part a generated by minimal mapping theory. (d) An incorrect solution for Part b generated by the relative velocity principle.

Figures 5a and 5b illustrate two examples of this. These displays can pose problems for motion correspondence models that are based on the two principles just described. Figure 5c depicts an incorrect solution for Figure 5a that is generated by minimal mapping theory (Ullman, 1979, p. 99). Figure 5d depicts an incorrect solution for Figure 5b generated by the relative velocity principle (Dawson, 1986, Figure 5,

Motion correspondence problems. (a) An example of apparent motion display. (b) The four possible motion correspondence matches for the display, labeled from 0 to 3. (c) An autoassociative network for this problem. The labeled circles depict processing units; each unit represents one of the correspondence matches from Part

b. Lines represent the weighted connections between processing units. (d) The representation of the network in linear algebra. (The column vector represents the activation values of each processing unit, which vary over time. The square matrix represents the connection weights for the network, which are defined on the basis of the three constraining principles. )

6,

Examples of the network's performance for some benchmark displays.

7,

An example of a problematic display. (a) An incorrect correspondence solution generated through the use of the "standard" settings. (b) The correct solution generated by change in the network's settings, as described in the text.

8 and

Solutions generated for the Ternus configuration. (a) The group motion solution, generated through the use of the standard settings. (b) The element motion solution, generated when the network's preference for short matches is increased. (c) The group motion solution, generated when the distance between elements is reduced from five units to one unit, even when preference for short matches is increased.

9).

The two incorrect solutions depicted in [Figure 5](#) show that for some displays, minimal mapping theory and a relative velocity model include motion correspondence matches that are discarded by the human visual system. To deal with this problem, [Ullman \(1979, pp. 97-101\)](#) modified the minimal mapping theory to include an element integrity principle. According to this principle, the splitting of one element into parts during movement, or the fusing together of different elements into one, should be penalized. Ullman incorporated these penalties into the nearest neighbor cost function and, as a result, extended the range of problems solved by minimal mapping theory.

The element integrity principle is consistent with general assumptions about the physical nature of moving surfaces (see also [Marr's \[1982, pp. 111-114\]](#) discussion of stereopsis). Specifically, proximal stimulus elements are assumed to correspond to parts of coherent, physical stimuli, such as edge segments, physical markings, and so on (see [Ullman, 1979, Chap. 2](#)). The physical coherence of surfaces (and therefore of surface parts) suggests that the splitting or the fusing of visual elements is unlikely. In addition, one-to-one mappings between elements over time will be correct everywhere except at surface discontinuities (e. g. , at an occluding edge where different elements may be suddenly appearing or disappearing). However, discontinuities make up a small proportion of scenes and images, and as a result the element integrity principle is likely to be true over most of an image.

Nevertheless, element integrity by itself is a very weak constraining principle. For instance, each of the possible motion correspondence solutions illustrated in [Figure 1](#) is consistent with the element integrity principle. Therefore, this principle alone does not generate *unique* solutions to the correspondence problem. The utility of this principle, illustrated in [Ullman's \(1979\)](#) modified minimal mapping theory, is that it may select one solution from a set that cannot be differentiated by other constraints, as in [Figure 5](#).

## The Need for Multiple Constraints

In the preceding sections, three potential constraining principles for the motion correspondence problem were described. The application of each is supported by empirical studies of motion perception, and each is consistent with general assumptions about the nature of moving surfaces. However, it is clear that none of these constraints are by themselves sufficient to emulate humans' solutions to the motion correspondence problem. The preceding examples show that if one of the constraints is omitted in a model, the model will generate incorrect solutions (i. e. , solutions not generated by the human visual system) for elementary displays. Given this situation, the working hypothesis for the model to be described next was that all three must be applied simultaneously. Once this working hypothesis is adopted, the researcher's task is to design an effective procedure for this simultaneous application.

## Part 2: A Motion Correspondence Model

In the following section, a model is presented for solving the motion correspondence problem by simultaneously applying the three constraints just described. The model is an *autoassociative network* that iteratively modifies the activation pattern of a set of simple interconnected processing units (representing local measurements) until a stable pattern is achieved (representing a *global interpretation*). The model's properties are described in three steps. First, it is shown how the three constraining principles are used to define the strengths of connections between processing units. Second, rules for iteratively updating the network are described and are analyzed to determine the nature of the stable states to which the network converges. Third, in order to examine the performance of the model, the solutions that it generates are compared with those generated by the human visual system.

## Autoassociation and Problems of Underdetermination

Within cognitive science, there is considerable interest in connectionist models of perception and cognition. Connectionist models are networks of simple processing units (e. g. , [Clark, 1989](#); [Rumelhart, Hinton, & McClelland, 1986](#); [Smolensky, 1988](#)). A single processing unit is characterized by a numeric activation value, which is changed as a function of the total signal processed by the unit. Connections between processing units in a network are communication channels that transmit numeric signals from one unit to another. A particular connection between two units is defined by a single number representing its strength, which is used to scale the numeric signal that it transmits.

One kind of connectionist model is an autoassociation network. It consists of a set of so-called massively parallel processing units; that is, each processing unit is connected to every other processing unit in the network. The initial pattern of network activation values produces changes in itself through a feedback loop enforced by the massively parallel connections. Given an incomplete initial pattern, an autoassociator can fill in missing details on the basis of pattern knowledge stored in its connections (e. g. , [Hinton & Sejnowski, 1986](#); [Hopfield, 1982](#)), and this knowledge can also be used to assign unique category labels to input patterns (e. g. , [Anderson & Mozer, 1981](#); [Anderson, Silverstein, Ritz, & Jones, 1977](#)).

An autoassociator can also solve problems of underdetermination by applying constraints (e. g. , [Grzywacz & Yuille, 1988](#)). Consider the general structure of such a model for the motion correspondence problem. Assume that there are  $N$  Frame 1 and  $M$  Frame 2 elements in an apparent motion display. This means that there are  $N \times M$  possible motion correspondence matches. The model must select a subset of these possible matches as being true of the display. Each possible correspondence match is represented by one of the processing units in the autoassociation network. A match is included in the solution to the correspondence problem if the processing unit that represents it has a high activation value at the end of processing. Otherwise, the match is excluded from the solution.

## Defining Connection Strengths

The connections among processing units in the network are defined by whatever constraints are being applied to the problem. For example, imagine that only one constraint is being used. Let  $U_i$  and  $U_j$  be two processing units in the autoassociation network, each representing a different correspondence match. If both these units represent matches that are consistent with the applied constraint, the connection between them should be assigned a large, positive weight. As a result, high activity in  $U_i$  would produce high activity in  $U_j$ , and vice versa. If these two units represent matches that are not consistent with the applied constraint, the connection between the two should be assigned a large, negative weight. As a result, high activity in  $U_i$  would produce low activity in  $U_j$ , and vice versa.

In the computer simulation, the weighted connections between processing units are represented in a square matrix  $\mathbf{C}$  with  $N \times M$  rows and columns. Each entry  $c_{ij}$  in  $\mathbf{C}$  is a numeric value that represents the connection strength between processing units  $U_i$  and  $U_j$ .  $\mathbf{C}$  is a symmetric matrix; that is,  $c_{ij} = c_{ji}$ . Processing units can be connected to themselves; when some constraints are applied, it is possible that  $c_{ii} \neq 0.00$ .

The connections among units, represented in the constraint matrix  $\mathbf{C}$ , are defined by a combination of the independently determined influences of the nearest neighbor principle, the relative velocity principle, and the element integrity principle. As a result, it is convenient to describe  $\mathbf{C}$  as the weighted sum of three connection matrices, each identical to  $\mathbf{C}$  in size and each defined by one of the three principles being exploited. Of importance is that these three matrices are described for expository purposes only: There is only one set of

connections among processing units in the network, and **C** is its only representation.

## Nearest neighbor weights.

Let **NN** be a square matrix that represents network connections defined by only the nearest neighbor principle. Recall that this principle encourages the selection of shorter motion correspondence matches. **NN** is defined by the creation of a single excitatory connection between each unit and itself; as a result, **NN** is a diagonal matrix. The strength of each connection in the diagonal of **NN** ranges from 0.00 to 1.00 and is a function of the length of the correspondence match represented by the unit. The shorter this match is, the stronger is the excitatory connection. Because of this, units representing shorter matches (all other factors being equal) increase their activation values at a faster rate than do units representing longer matches.

An exponential function is used to transform  $\|m_i\|$ , the length of motion correspondence match  $i$  (which in principle can range from zero to infinity), into a connection strength within the desired range of 0.00 to 1.00. An exponential function was chosen because of Ullman's (1979, pp. 114-118) argument that the probability distribution for patterns of movement projected onto the retinas is exponential in shape. Equation 1 defines the nearest neighbor principle ( $nn$ ) as operationalized in the computer simulation: The constant  $a$  is a positive parameter that defines the preference of the network for short element displacements (i. e., small values of  $\|m_i\|$ ). When  $a$  is large, the network has a very strong preference for short element displacements and generates very small connection strengths for long element displacements. When  $a$  is small, the preference for short element displacements still exists but is not as strong. In the simulations to be described,  $a = 0.25$ .

## Relative velocity weights.

Let **RV** be a matrix representing network connections defined by only the relative velocity principle. Recall that this principle is used to assign neighboring elements similar movements (i. e., correspondence matches of similar direction and length). In order to define weights through the use of this principle, two issues must be considered: (a) an operational definition of relative velocity and (b) an operational definition of *neighboring*.

Relative velocity is the distance between two vectors **mi** and **mj**, representing two motion correspondence matches, after they have been centered at a common origin (Figure 3). Because it is a distance measure, it can (in principle) range from zero to infinity. In the model, this distance is transformed into a connection strength limited to the range from -1.00 to 1.00. This means (all other factors being equal) that processing units representing similar motion correspondence matches increase each other's activation values. Processing units representing dissimilar motion correspondence matches decrease each other's activation values.

Relative velocity ( $r_{ij}$ ) is transformed to a connection strength in the desired range by the following exponential function: The constant  $\beta$  in this equation determines the preference of the network for small relative velocities, in the same manner that was described for the constant  $\beta$  in Equation 1. In the simulation to be described,  $\beta = 0.25$ . The value  $\epsilon_{ij}$  in Equation 2 is a parameter used to operationalize the notion of neighboring Frame 1 elements. The neighborhood parameter is such that the effect of two Frame 1 elements on one another (with respect to correspondence match assignment) decreases as an exponential function of their separation. This is consistent with Dawson's (1987) finding that the effect of a context decreased exponentially as it was moved farther from a motion competition display.

The neighborhood parameter  $\epsilon_{ij}$  is defined as follows: Let  $x$  be the coordinates of the Frame 1 element from which  $m_i$  originates, and let  $y$  be the coordinates of the Frame 1 element from which  $m_j$  originates. The distance between the two Frame 1 elements is therefore  $\|x - y\|$ . (The neighborhood parameter is not computed if this distance is zero—i. e., if the two matches originate from the same Frame 1 element.) The neighborhood parameter is defined by the exponential equation The parameter  $\epsilon_{ij}$  determines the extent to which the assignment of a correspondence match to one Frame 1 element is affected by match assignment to other, distant Frame 1 elements. When  $\epsilon_{ij}$  is large, these distant elements have little effect; their effect increases as  $\epsilon_{ij}$  is decreased. In the simulations to be described,  $\epsilon_{ij} = 0.15$ .

## Element integrity weights.

Let **EI** be a matrix representing network connections defined by only the element integrity principle. Recall that this constraining property is used to inhibit the splitting or fusing of display elements during movement. **EI** is a

symmetric matrix constructed in the following manner: All pairs of processing units that represent motion correspondence matches emanating from the same Frame 1 element are given mutually inhibitory connections. Specifically, if units  $U_i$  and  $U_j$  represent two matches originating from the same Frame 1 element (i. e. , a split),  $e_{ij}$  and  $e_{ji}$  are both set to  $-1.00$ . Similarly, all pairs of processing units that represent motion correspondence matches that terminate at the same Frame 2 element (i. e. , a fusion) are given mutually inhibitory connections. All other connection strengths represented in  $EI$  are assigned values of  $0.00$ .

## Defining the overall connection matrix.

The connection matrix  $\mathbf{C}$  for the network is defined as the weighted sum of the three individual constraints matrices just described. Specifically, where  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are constants specifying the relative importance of each constraining principle and  $d$  is a fraction that determines the rate of convergence of the algorithm. In the simulations to be reported,  $d = 0.10$ , and  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  all =  $1.00$ .

Because  $\mathbf{NN}$ ,  $\mathbf{RV}$ , and  $\mathbf{EI}$  are all symmetric matrices,  $\mathbf{C}$  must also be symmetric. This ensures that the network converges to a stable state, as described later. Also,  $\mathbf{C}$  is not a learned matrix, as is typically the case in autoassociative networks (e. g. , [Anderson et al. , 1977](#); [Hopfield, 1982](#)). The values of  $\mathbf{C}$  are determined a priori from the properties of the input display.

## Updating the Network

The preceding section demonstrated how the connections among the processing units in the network are defined by the three constraining principles. This section of the article describes how these connections are used to iteratively change the activation values of the processing units, so that eventually these units represent a solution to the motion correspondence problem.

Linear algebra can be used to describe how to iteratively update an autoassociative network capable to identifying the motion correspondence solution that is most consistent with the constraints used to define connection weights. (For an introduction to linear algebra in the context of connectionism, see [Jordan, 1986](#). ) Let  $\mathbf{a}$  be a column vector of  $N \times M$  entries, which represent the activation values of every processing unit in the network. The entry  $a_i$  in this vector represents the activation value of  $U_i$ . The state of all the units at time  $k$  is represented as  $\mathbf{a}^k$ . [Figure 6](#) illustrates the relation among the apparent motion display, the autoassociative network, and the representation in linear algebra.

Two equations are used to describe how the network is updated. The first describe how  $\mathbf{a}$  is changed over time: In [Equation 5](#), the matrix  $\mathbf{W}$  is the connection matrix  $\mathbf{C}$  plus the identity matrix  $\mathbf{I}$  and is used as a compact notation to describe network updating. In [Appendix A](#) this compact notation is used to prove that the network converges to a stable state.

[Equation 5](#) is the updating rule applied in the "brainstate-in-a-box" model of pattern categorization (e. g. , [Anderson & Mozer, 1981](#); [Anderson et al. , 1977](#)). At each processing step  $k$ , every unit  $U_i$  in the network adds, to its current activation value, the total signal transmitted to it by the units to which it is connected. This total signal is equal to  $\sum_j c_{ij} \cdot a_j$ . Because the connection strengths dictate the consistency among the motion correspondence matches, the activation values of the units representing to-be-included matches are strengthened. The activation values of units representing to-be-excluded matches are weakened.

[Equation 5](#) represents a feedback loop that, operating alone, would usually lead processing units to unboundedly increase or decrease their activation value (e. g. , [Anderson et al. , 1977](#), p. 427). A second equation, to be applied after [Equation 5](#) is computed, is required to restrict the growth of the activation values: [Equation 6](#) constrains the growth of  $\mathbf{a}$  by ensuring that after the network is updated,  $\mathbf{a}$  has a length of  $1.00$ . As shown later, the application of this equation also ensures that the network converges to a stable state that represents the set of activation values that are most consistent with the constraints represented in  $\mathbf{C}$ .

The autoassociative network governed by [Equations 5](#) and [6](#) is very similar to the model proposed by [Anderson et al. \(1977\)](#). Qualitatively speaking, the network just described is a "brain-state-in-a-sphere. " The vector  $\mathbf{a}$  is the unit radius of a hyperdimensional sphere. Repeated application of the matrix  $\mathbf{C}$  rotates  $\mathbf{a}$ , so that it points to different directions from the origin of the hypersphere. As processing proceeds,  $\mathbf{a}$  is rotated to point in a stable direction that represents a solution to the problem being solved.

## Convergence Properties of the Network

Vector normalization is not the only means by which the growth of activation values in  $\mathbf{a}$  could be checked. In the brain-state-in-a-box model, activation values are restricted to the range (-1.00 to 1.00). Although this approach could be adopted to solve the correspondence problem (e. g. , [Dawson, 1988](#)), there are two reasons why it is not ideal. First, although it can be shown that the processing units in the brainstate-in-a-box converge to a stable state (i. e. , a corner of a hyperdimensional box in which every activation value is equal to  $\pm 1.00$ ), this requires that the nonzero eigenvalues of the constraint matrix all be positive (see [Anderson et al. , 1977](#), p. 428). Although this property follows from the learning procedure used by Anderson et al. , it is not necessarily true of  $\mathbf{C}$ , as described earlier. Second, the brainstate-in-a-box can have a large number of stable states. In order to solve a problem of underdetermination, a network with fewer stable states is desirable.

Through the use of vector normalization to restrict activation value growth, it can be shown that the autoassociation network converges to a stable state. In [Equation 5](#), both  $\mathbf{I}$  and  $\mathbf{C}$  are symmetric. Therefore, the matrix  $\mathbf{W}$  is also symmetric. This ensures that there exists some vector  $\mathbf{e}$  such that  $\mathbf{W} \cdot \mathbf{e} = \lambda \cdot \mathbf{e}$ , where  $\lambda$  is a real scalar value called an eigenvalue and  $\mathbf{e}$  is called an eigenvector of  $\mathbf{W}$ . When  $\mathbf{W}$  is used to premultiply one of its eigenvectors, the result is that the length (but not the direction) of the eigenvector is changed. Let  $\mathbf{a}_0$  be the initial activation values of the network, and assume that  $\mathbf{a}_0$  is normalized to be of unit length. Let  $\mathbf{e}_1$  be the most dominant eigenvector of  $\mathbf{W}$  (i. e. , the eigenvector associated with the largest eigenvalue). [Appendix A](#) presents a proof that except in special cases, a network updated by the procedures described in [Equations 5](#) and [6](#) rotates the activation vector  $\mathbf{a}$  until it converges to either  $\mathbf{e}_1$  or  $-\mathbf{e}_1$ , depending on the relationship of  $\mathbf{a}_0$  to  $\mathbf{e}_1$ .

This proof helps establish another important property of the model. The goal for overcoming a problem of underdetermination is to converge to a solution that is *unique*: A system should produce the same answer to the problem each time that it is presented. The current model converges to a unique solution because the vector  $\mathbf{e}_1$  is uniquely defined for matrix  $\mathbf{W}$ . The model fails to generate unique solutions only in rare, special cases in which  $\mathbf{W}$  has more than one dominant eigenvector (e. g. , when two eigenvectors have the same eigenvalue and this value is greater than all other eigenvalues for  $\mathbf{W}$ ). In this special case, the model always converges to a solution that is a linear composite of the two dominant eigenvectors, and as a result, identical solutions are usually not achieved when a problem is presented repeatedly (see [Hall, 1963](#), pp. 63-66). However, this special situation is rarely encountered. It is highly unlikely that a matrix that represents naturally occurring properties would have more than one dominant eigenvector (W. W. Rozeboom, personal communication, October 15, 1990). Indeed, this situation has not been encountered for any of the displays used to test the performance of the model.

## Stable States Are Optimal Problem Solutions

The results just summarized, and detailed in [Appendix A](#), indicate that the most dominant eigenvector (multiplied by 1 or -1) of the constraint matrix  $\mathbf{W}$  represents the stable state to which the network converges in almost all cases. It is also important to show that this stable state is meaningful with respect to the constraints defined in  $\mathbf{W}$ . In other words, it must also be shown that this stable state represents the best solution to the problem of underdetermination, given the constraints used to create  $\mathbf{C}$ .

[Hopfield \(1982\)](#) developed a measure of the cost, or energy, of patterns of activation in a Hopfield net, which is a particular example of an autoassociative network. The iterative processing in a Hopfield net serves to decrease this cost measure until a minimal-energy network state is reached. This minimal-energy state is the optimal response of the network, given the constraints specified in its connections: Changes in any of the final activation values result in a higher-energy state (i. e. , a state that is less consistent with the constraints defined by network connections).

[Appendix B](#) develops a minor generalization of [Hopfield's \(1982\)](#) cost measure to be applied to the network characterized by [Equations 5](#) and [6](#). It is shown that  $\mathbf{e}_1$ , the most dominant eigenvector of the connection matrix, represents the least-energy state of the network, as defined by this cost measure. Thus when the network converges, it has reached a state representing the solution that is most consistent with the constraints defined in  $\mathbf{C}$ .

Of course, this least-energy state of the network is optimal in another sense. The simulation results to be described indicate that when this state is achieved, the solution represented by the network is, for a wide variety of displays, the same as the solution generated by the human visual system.

## Interpreting Converged States

In order for a network to make decisions, there must be a nonlinear component to its processing (e. g. , [Blake & Zisserman, 1987](#), Section 1. 2. 3). The computations summarized in [Equations 5](#) and [6](#) define a system that is linear.<sup>4</sup> As a result, an additional component is required if the stabilized activity values are to be interpreted as decisions about the inclusion or the exclusion of correspondence matches. Specifically, continuous activity values must be translated into discrete assertions about inclusion or exclusion.

Nonlinear components are characteristic of autoassociation networks. For example, in a Hopfield net (e. g. , [Hopfield, 1982](#)), changes in the state of a processing unit require that the total signal to the unit exceed a threshold. Similarly, restricting the range of activation values in the brainstate-in-a-box model introduces a nonlinear processing component ([Anderson et al. , 1977](#)).

The nonlinearity that is introduced in the current model is a threshold-testing operation that only occurs after network convergence is achieved. An arbitrary threshold is selected for the network. If  $a_i$  is greater than or equal to the threshold, then the match represented by processing unit  $U_i$  is included in the correspondence solution. Otherwise, the match is not included in the solution. In the simulation just described, the threshold was equal to 0. 13.

## Specifying the Starting State

In order for network processing to commence, the initial activation values of the processing units (i. e. , vector  $\mathbf{a}_0$ ) must first be specified. In the simulations to be reported, all motion correspondence matches were assumed to be equally likely at time 0. Thus each processing unit was initially assigned the same positive value (1. 0), which was then scaled by normalizing  $\mathbf{a}_0$  to unit length. In adopting this method, the model is following a procedure similar to that of relaxation labeling (e. g. , [Zucker, 1976](#)): Initially, all possible labels (motion correspondence matches) are asserted to be equally true of the display, and as iterative processing proceeds, labels inconsistent with the constraints in  $\mathbf{C}$  are discarded.

## Performance of the Model: Preliminary Remarks

The performance of the motion correspondence model was examined in a series of computer simulations. The only information provided to the model is the numbers of elements in Frame 1 and in Frame 2, as well as the  $x$  and  $y$  coordinate of each element. Unless otherwise stated, the distance between nearest Frame 1 neighbors in the following figures was five arbitrary units, and the diagrams are drawn to scale. The "standard" settings noted earlier for the equation parameters were used to compute the network's connection weights for each simulation. The network was always initialized so that the processors representing different motion correspondence matches had equal, positive activation values. These values, however, varied from display to display because the vector that represented them was normalized. As a result, initial activation was a function of the number of processors in the network. Iterative processing continued until the network converged upon a solution. The convergence index that was used was the sum of the squared differences in processor activation values from iteration  $k$  to iteration  $k + 1$ . Convergence occurred when the value of this index reached zero.

Before the network's performance is considered, it is important to place claims about the model in the proper perspective. The model is not proposed as a specific performance theory of human motion correspondence processing. It is highly unlikely that human motion correspondence is performed in exactly the manner dictated by the network. For example, the model has too many degrees of freedom: Several equation parameters can be freely varied, and numerous alternative equations could be derived for each of the constraints.<sup>5</sup>

Although the network is not being proposed as a performance theory, it is being proposed as a general framework for human competence in motion correspondence processing. Specifically, the model is presumed to apply the same kind of constraints to the motion correspondence problem as does the human visual system. In general, then, the network's performance reflects the adequacy and the utility of the constraining principles, even if the model uses specific procedures that may differ from those of the human visual system.

When viewed as a working competence theory, the model provides a powerful qualitative tool for generating insights, and for raising questions, about human motion perception. What kinds of solutions can be solved by the

simultaneous application of the three constraints? Are there any emergent properties of the network, so that it solves problems that were not considered during its creation? If such qualitative questions are dealt with first, the foundation is laid for later quantitative attempts to model human perceptions of apparent-motion displays (cf. Köhler, 1947/1975, chap. 2; see also Dawson, 1990b).

With this perspective noted, the performance of the model is described in two parts. In the following section, the model is shown to generate the same qualitative solutions as the human visual system to a set of benchmark displays. In Part 3 the simulation's performance on some additional displays is used as a focal point for considering several theoretical issues that have arisen in the study of motion perception.

## Performance of the Model: Benchmark Displays

When a model of motion correspondence processing is developed, it is difficult to specify an optimal procedure to test its capabilities. The number of potential displays that the model could process is infinite. The problem is to choose an interesting and informative subset of these potential tests. The strategy adopted to test the current model was to choose a set of so-called benchmark displays to investigate the capabilities of the model. For the most part, these benchmark displays were quite simple, primarily because the qualitative nature of human motion correspondence solutions is known only for relatively simple displays (see, for instance, the many examples given by Kolars, 1972). The particular benchmarks tested were selected for a variety of reasons. Many of the displays posed severe problems for ancestors of the current model (e. g. , Dawson & Pylyshyn, 1986); correct solutions for these displays therefore indicated definite progress in modeling. Some of the displays tested the utility of a particular constraint or posed potential challenges because they violated a constraint exploited by the model. Still others were selected to differentiate the current model from those developed by other researchers. Figure 7 depicts the model's performance on some benchmark displays, which are described in detail in the following sections. In each case, the model assigned the same set of correspondence matches that are assigned by the human visual system.

### Single element translation.

Figure 7a illustrates the simplest possible apparent motion display that could be presented to a human observer: a single element presented in different positions in Frames 1 and 2. The model solves this problem easily, converging to the correct solution (i. e. , the solution generated by the human visual system) after only one iteration. This performance is notable only in comparison with that of the model's ancestors. For instance, the network proposed by Dawson and Pylyshyn (1986) exploited only the relative velocity principle and, as a result, could not solve the correspondence problem for this elementary display; when only one element is in the display, no relative velocity information exists.

### Multiple element translation.

The model also generates correct solutions when  $N$  elements are translating from Frame 1 to Frame 2. This solution can be generated when the elements are moving in parallel (65 iterations for the Figure 7b solution) or when the elements are moving in different directions (41 iterations for the Figure 7c solution). This latter display was presumed to pose a greater challenge for the model because the relative velocity principle cannot be exploited as readily as for the Figure 7b display.

### Multiple element rotation.

The model can generate the correct solution when a configuration of elements is rotated about the origin. Figure 7d illustrates a display in which elements are located at the vertices of a square. This square configuration was rotated 10° clockwise about the origin (indicated by the small circle) from Frame 1 to Frame 2. The model required 48 iterations to generate the Figure 7d solution. It was viewed as a challenge for the current model because elements opposite one other across the origin move in opposite directions, which is contrary to the relative velocity principle.

### Nearest neighbor sensitivity.

Figure 7e illustrates the model's performance for three motion competition displays. In the first two displays, one

Frame 2 element is twice as close as the other to the central Frame 1 element. In both cases, the model assigns the shortest correspondence match after 49 iterations. In the third display, both Frame 2 elements are the same distance from the central Frame 1 element. In this case, the model generates a splitting solution after only one iteration.

The model's solutions to these competition displays are important in two respects. First, they show that the network is implementing a nearest neighbor solution to these problems, as does the human visual system. Second, they show how the model performs when there is an unequal number of elements in the two frames of view. In the first two cases, no motion correspondence match was assigned to one of the Frame 2 elements, which can be interpreted as an assertion that this element suddenly appeared. Human observers often interpret such a display in this fashion. However, this kind of interpretation is not possible for minimal mapping theory because it must exploit the so-called cover principle, which forces a correspondence match to be assigned to every display element. The fact that the current model can function without requiring the cover principle is an important advance and is discussed in detail in Part 3.

## Context sensitivity.

The major motivation for incorporating the relative velocity principle into the current model was evidence that such information is an important determinant of motion correspondence matches for human observers (Dawson, 1987). Figure 7f illustrates a solution, generated in 58 iterations, that depends on this principle. The Figure 7f display can be viewed as being identical to the third display of Figure 7e with an additional contextual element that provides disambiguating relative velocity information. The solution generated by the model illustrates that it can generate simple field effects that are not unrelated to those reported by Ramachandran and Anstis (1985). Field effects are discussed in more detail in Part 4.

## Motion shear.

Figure 7g depicts a solution for a problem that is difficult for any model that uses the relative velocity principle. In this display, nearest neighbors move in exactly opposite directions. The correct solution was generated after 293 iterations. The fact that it could be generated at all indicates that the nearest neighbor and the element integrity principles combined are capable of overcoming alternative (and incorrect) solutions that are more consistent with the relative velocity principle.

## Stationary elements.

Figure 7h illustrates a solution for a degenerate apparent-motion display; in this display, no motion is perceived because the presented elements do not change position over time. The solution that is generated (after 50 iterations) is consistent with human perceptions of stationary elements. This display is a benchmark because it poses tremendous difficulties for certain operationalizations of the constraining principles. For example, instead of defining the relative velocity principles as in Equation 2, one could define relative velocity as the cosine of the angle between neighboring matches (after centering at a common origin). Such a definition has the attractive advantage of providing a natural scaling of relative velocity into the range (-1 to 1) but cannot be applied to stimuli in which one or more vectors have zero length, as is the case in this display.

## Necessity of all three constraints.

Although not illustrated in any of the figures, it can be easily shown that the model's performance depends on the application of all three constraints (Dawson & Harder, 1989). If any one of the constraints is removed from **C**, there will be some displays for which the model will generate solutions that differ from those generated by humans. This is consistent with the arguments made earlier in which experimental evidence was provided for each of the constraints.

## Problems with "standard" settings.

Some displays provide problems for the network when the so-called standard settings are used in the equations that define connection strengths. One example of an incorrect solution generated by the standard network is

illustrated in [Figure 8a](#). Typically, cases like this can be remedied with some minor changes of the network's settings. For instance, the correct motion correspondence solution of [Figure 8b](#) is generated by increasing the system threshold from 0.13 to 0.23.

This flexibility is indicative of the many parameters that can be freely varied in the model and would be an undesirable property if the model were to be proposed as an explanation of human motion correspondence. To increase the explanatory power of the model, it must eventually be translated from a general (qualitative) framework for motion correspondence competence to a specific (quantitative) model of motion correspondence performance. The collection of the experimental data required for such a translation is an important component of an ongoing research program (e. g. , [Dawson, 1987, 1990a](#); [Dawson & Wright, 1989](#)). However, even in its current qualitative state, the model sheds some interesting light on some specific motion perception issues, as is shown in the following discussion.

### Part 3: Motion Correspondence and Motion Perception

The previous section indicated that the motion correspondence model was capable of generating the same qualitative solutions generated by human observers to a variety of benchmark apparent-motion displays. In the section that follows, additional examples of the model's performance are used to focus discussion on a number of theoretical issues that have arisen in the study of motion perception. These issues include the two-process distinction in motion perception, particular motion correspondence modeling assumptions, the role of figural properties in apparent-motion perception, and the relation between motion correspondence and visual attention.

A major theme underlying this part of the article concerns the utility of designing effective procedures for studying psychological phenomena. One tradition in the study of motion correspondence processing, illustrated earlier, is to argue that a small number of constraining principles are exploited by the visual system and then to design a working computer model that implements these constraints (e. g. , [Grzywacz & Yuille, 1988](#); [Ullman, 1979](#)). When such a model is constructed, one can determine the extent to which the constraining principles account for phenomena not originally considered when the model was designed. One measure of the model's strength is its ability to account for a wider range of phenomena than was originally intended. Examples of emergent explanations generated by the current model include its ability to generate both versions of the Ternus configuration ([Ternus, 1938](#)), its ability to follow the cover principle under certain element displacement conditions, and its ability to generate least-change transformations (see the following section).

In contrast, a second tradition in the study of motion correspondence processing is to use psychophysical experiments to compose a catalog of independent variables that affect which matches are assigned (e. g. , [Ramachandran & Anstis, 1986b](#); [Sekuler et al. , 1990](#)). This tradition fulfils the important role of refining descriptions of correspondence processing. However, it does little to *offer explanations* of how this processing actually occurs. For example, [Ramachandran and Anstis \(1986b\)](#) proposed that the visual system applies a set of strategies or heuristics "from what is in effect a bag of tricks" (p. 102). However, they offered few concrete proposals for how these strategies could be realized as an effective procedure, a step that many authors believe is necessary to provide explanations of phenomena (e. g. , [Johnson-Laird, 1983](#), pp. 4-6). Furthermore, experimental results may provide misleading information about what such a bag of tricks may contain. For instance, after reviewing the experimental evidence, [Attneave \(1974\)](#) proposed that one rule governing correspondence match assignment is a preference for symmetric matches. However, such a rule need not be explicitly implemented. [Ullman's \(1979\)](#) minimal mapping theory and the current model (see [Figure 8](#)) can generate symmetric patterns of matches without explicitly assuming a symmetry rule. This kind of discovery is possible only when one explores processing with explicit proposals about effective procedures.

### The Two-Process Distinction in Motion Perception

In recent years, much of the research on human motion perception has been guided by the putative distinction between two motion perception systems (for reviews, see [Anstis, 1980, 1986](#); [Braddick, 1980](#); [Petersik, 1989](#)). The first, called the *short-range motion system*, is thought to detect movements involving short element displacements (e. g. , 15-30 minutes of visual angle) and brief temporal intervals (e. g. , interstimulus intervals of 40 ms or less). [Anstis \(1980\)](#) proposed that the short-range system detects motion before the extraction of figural properties. As a result, it is typically modeled as some form of spatiotemporal correlation between image intensities that vary continuously over time (e. g. , the correlator class of detector proposed by [Reichardt, 1961](#); for related models, see [Adelson & Bergen, 1985](#); [Burr, Ross, & Morrone, 1986](#); [Dawson & Di Lollo, 1990](#); [Farrell & Kesler, 1988](#); [Marr & Ullman, 1981](#); [Morgan & Watt, 1983](#); [van Santen & Sperling, 1984, 1985](#); [Watson &](#)

Ahumada, 1985).

The second process, called the *long-range motion system*, is thought to detect movement involving much longer element displacements (e. g. , several degrees of visual angle) and temporal intervals (e. g. , interstimulus intervals of more than 100 ms). The long-range system is usually proposed as the system that mediates the perception of classical apparent motion (i. e. , the motion considered by [Kolers, 1972](#)), and must solve the correspondence problem, as was assumed earlier in this article. [Anstis \(1980\)](#) proposed that the long-range system detects motion after some figural properties have been extracted from the stimulus. Accordingly, it is typically modeled as the matching of discrete tokens over time, whereby each token represents the properties of an individuated element (e. g. , [Ullman, 1979, 1981](#)).

Recently, researchers have questioned the validity of the two-process distinction. Many of the classical differences between the two systems have not stood up to experimental inquiry (for a review, see [Cavanagh & Mather, 1989](#); the contrasting view was presented by [Petersik, 1989](#)). Cavanagh and Mather generalized the short- and long-range distinction to one between first and second-order motion but then proceeded to argue that first- and second-order motion detectors do not differ qualitatively.

Researchers who have argued against a qualitative distinction between the short- and long-range motion systems have also proposed that all motion perception can be modeled with some variation of a Reichardt detector (e. g. , [Cavanagh & Mather, 1989](#)). Token-based schemes are viewed skeptically. [Adelson and Bergen \(1985, p. 284\)](#) criticized feature-based schemes for not making precise claims about which figural properties are explicitly represented. [Ramachandran and Cavanagh \(1987\)](#) asked, "How does the visual system know which spot goes with which? Our answer is that the visual system doesn't care"; instead, the short-range motion signal derived from the low spatial frequencies of the stimulus is "spontaneously attributed to the spots themselves" (p. 105).

One problem with this position is that it mistakenly equates the detection of element motion with the maintenance of element identity. Typical correlator models (e. g. , [Dawson & Di Lollo, 1990](#); [Reichardt, 1961](#); [van Santen & Sperling, 1984](#)) generate a numeric value that can be interpreted as asserting that "motion to the left was detected in the display" or "no motion was detected in the display. " However, these assertions are quite different from the assertion that "Frame 1 element *x* and Frame 2 element *y* are the same entity," which in some cases may be completely unrelated to movement. For instance, the states of the motion correspondence network represent assertions about element identities when no movement has occurred (e. g. , [Figure 7a](#)). Similarly, motion signals can be generated for displays in which elements may not have been individuated (e. g. , [Daugman, 1988](#); [Mather, 1984](#)).

A token-based model—the motion correspondence network—can generate both long- and short-range identity matches for a bistable display. The Ternus configuration ([Ternus, 1938](#)) has long been studied by apparent-motion researchers. The display consists of a group of three elements that are translated in one direction from Frame 1 to Frame 2. The amount of translation is such that two stimulus locations in both Frame 1 and Frame 2 always represent the position of an element (see [Figure 9](#)). The Ternus configuration can support two motion percepts. One is the group motion percept, in which all three elements are perceived to translate in one direction as a whole group ([Figure 9a](#)). The other is the element motion percept, in which two elements remain stationary while the third moves from one end of the group to the other ([Figure 9b](#)). In the context of the two-process distinction, researchers have argued that the long-range system generates the group motion percept and that the short-range system generates the element motion percept (e. g. , [Braddick & Adlard, 1978](#); [Pantle & Picciano, 1976](#)).

When the standard settings for the motion correspondence model are used, the model generates correspondence matches that are consistent with the group motion percept for the Ternus configuration ([Figure 9a](#)). This is not surprising, given the contention that the network simulates one component of the long-range system. However, a minor adjustment of the model's settings, to increase its preference for short correspondence matches, results in the network's generating a solution that has been ascribed to the short-range system. The matches consistent with element motion were produced after the value of *a* in [Equation 1](#) was increased from 0. 25 to 0. 50 ([Figure 9b](#)).

The motion correspondence model also provides an additional account of another Ternus display regularity. [Breitmeyer and Ritter \(1986\)](#) argued that group motion results when there is a reduction in the visible persistence of activation produced by Frame 1 of the configuration. The results of several experiments have supported this argument. For instance, the visible persistence of visual elements is known to decrease substantially when display elements are very near one another (e. g. , [Di Lollo & Hogben, 1987](#)). When the distance between Ternus configuration elements is decreased, the group motion percept predominates

(Breitmeyer & Ritter, 1986; Petersik, 1986). This finding supports the visible persistence hypothesis. However, it is also consistent with the assumptions underlying the correspondence network: When the distance between elements is decreased from five units to one unit, group motion correspondence matches are assigned, even when  $a = 0.50$  (Figure 9c). This is because decreasing the distance between elements in Frame 1 increases the effect of the relative velocity principle through the neighborhood parameter in Equation 2.

The fact that relatively minor changes in network settings can produce both correspondence solutions to the Ternus configuration is at first glance consistent with Cavanagh and Mather's (1989) contention that separate motion detection systems may differ quantitatively but not qualitatively. However, the network's performance does not entail a rejection of a two-process distinction. This is because although the network maintains the identity of moving elements, it does not represent their movement. Later in this article it is argued that the two-process distinction may not be between different motion perception systems but instead may be between a low-level motion detector and an attentional tracking system that is part of visual cognition (see also Marr, 1982, pp. 202-204; Petersik, 1989).

## Cover Principle

Ullman's (1979) minimal mapping theory exploits the nearest neighbor principle in conjunction with a second constraint, called the *cover principle*. The cover principle requires that a valid correspondence solution account for ("cover") every Frame 1 and Frame 2 element with at least one motion correspondence match. In other words, this constraint stipulates that Frame 1 elements cannot suddenly disappear and that Frame 2 elements cannot suddenly appear.

In minimal mapping theory, the cover principle serves a special function that differentiates it from other constraints. It forces a system governed by the nearest neighbor principle to include at least some matches in a correspondence solution. Without this principle, Ullman's (1979) model would produce the zero-cost solution that does not include any motion correspondence matches at all.

The special function of the cover principle is reflected in how it is applied in minimal mapping theory. Other constraints are included in the cost function that is minimized. Because of this, these constraints are what connectionists would call "weak"; they can be violated to a certain extent if the result is a better overall solution. The cover principle is not part of the cost function but instead defines a set of necessary conditions on cost minimization. Thus the cover principle is a "strong" constraint: it cannot be violated.

Adoption of the cover principle could be defended on grounds similar to those used to defend the element integrity principle (e. g. , Dawson & Pylyshyn, 1988). However, evidence shows that the cover principle can be violated by the human visual system. For instance, Figures 2a and 2b represent displays in which one element is seen to move in one direction while a second element suddenly appears. This sudden appearance is a violation of the cover principle.

Element displacement appears to determine whether the cover principle is followed.

The cover principle and element displacement. (a) When element displacement is small, all six Frame 2 elements are covered by motion correspondence matches. (b) When element displacement is doubled, three of the Frame 2 elements are not covered by matches; they are seen by human subjects to suddenly appear.

Figure 10 illustrates two displays in which three elements are presented in the same locations in both Frames 1 and 2 and another three elements are presented only in Frame 2. If this second set of elements is close to the first set (short element displacement), then all Frame 2 elements are covered by motion correspondence matches: Human subjects see three elements move from behind the three stationary elements (Figure 10a). If this second set of elements is displaced farther from the first set, then not all Frame 2 elements are covered: Human subjects see three stationary elements, followed by the sudden appearance of another three elements (Figure 10b).

Violations of the cover principle clearly indicate that it should not be a strong constraint on motion correspondence processing. Ideally, weak constraints should be sufficient to identify a unique and correct solution. The motion correspondence network does not apply the cover principle because it does not specify that a particular subset of processing units must have high activation values. Yet it generates appropriate "covers" in the solutions to which it converges—it generates both of the Figure 10 solutions—and in so doing is sensitive to element displacement.

## Least-Change Transformations

The projected movement of rigid objects in three-dimensional space should produce two-dimensional patterns of movement that are consistent with the nearest neighbor, relative velocity, and element integrity principles. As a result, when these three principles are applied to a proximal stimulus produced by the movement of a rigid distal stimulus, the resulting correspondence solution should be physically plausible. In particular, the assigned correspondence matches should not contradict the rigid structure of the distal stimulus (see [Dawson & Pylyshyn, 1988](#)).

In some cases, however, there is more than one way to perceive motion and preserve physical plausibility. Consider the three displays presented in

Three examples of least-transformation solutions. (The outline circle represents an origin about which the group of three elements could be described as being rotated by [a] 45°, [b] 135°, or [c] 180°. See text for details. )

Figure 11. Each of these could be geometrically described, and plausibly perceived, as a rigid clockwise rotation of three points about the marked origin. Whereas motion correspondence matches consistent with this interpretation are generated by the model for [Figure 11a](#), alternative (physically plausible) solutions are produced for the other two displays.

The [Figure 11b](#) solution is consistent with the perception of the three elements translating downward, accompanied by a 45° clockwise rotation about the middle point. The [Figure 11c](#) solution is consistent with the perception of the three elements translating downward. (The model does not generate element trajectories per se; it generates only identity matches that can be viewed as being consistent with trajectory interpretations. ) The alternative matches are assigned for the [Figure 11b](#) and [11c](#) displays because they produce lower total element displacement and relative velocity than do the matches that are consistent with a rigid clockwise rotation. Thus when the model is faced with choosing between physically plausible interpretations, it selects the interpretation that represents a "least-change" transformation: the interpretation that produces the least cost with respect to the three applied constraints.

The human visual system also prefers least-change transformations of this type (e. g. , [Dawson & Pylyshyn, 1988](#); [Farrell & Shepard, 1981](#); [Shepard & Judd, 1976](#)). This has led some researchers to propose models of apparent-motion perception that apply explicit principles of least change (e. g. , [Caelli & Dodwell, 1980](#); [Foster, 1978](#); [Mori, 1982](#); [Restle, 1979](#); [Shepard, 1984](#)).

For example, [Shepard \(1978, 1981, 1982, 1984\)](#) proposed a geometric theory of mental representation that can be applied to apparent-motion perception. The basic assumption of the theory is that object motion is represented as a path of activation on a representational surface called a *manifold*. In general, each location on the manifold represents an object in a particular position and orientation in three-dimensional space. Shepard proposed that the visual system processes apparent-motion displays as follows: Frames 1 and 2 of an apparent-motion display produce two points of activation on a specific representational manifold. One point of activation represents the figural appearance of Frame 1; the other represents the figural appearance of Frame 2. The illusion of motion is then produced by a spread of activation from the first manifold point to the second: The path of activation from the Frame 1 point to the Frame 2 point represents the appearance of the object as it moves.

However, there are infinitely many paths between the activation that represent the two frames of view. Therefore, Shepard (e. g. , [1982, 1984](#)) assumed a "minimum principle": The visual system selects the shortest path between the two activation points. In addition, however, all points of the path must lie on the manifold because the manifold is defined by the set of physically plausible transformations that could be applied to represented objects. If the path was not on the manifold, a physically implausible transformation would be represented.

In comparison with the current model, [Shepard's \(1978, 1981, 1982, 1984\)](#) manifold approach has the advantage of providing an explicit account of how objects change appearance during apparent movement. However, the manifold model still requires the correspondence problem to be solved. For example, an apparent-motion display composed of several objects of the same type would result in several Frame 1 and Frame 2 activations on a single manifold. The correct correspondence matches would therefore be required to direct the spread of activation between appropriate manifold locations. This suggests that solving the correspondence problem and generating the figural appearance of moving objects are functionally different components of apparent-motion perception.

Evidence does support a functional dissociation between the processes that assign correspondence matches and those that generate object appearances in apparent motion. Although correspondence matches are assigned in a two-dimensional coordinate system (e. g. , [Green & Odom, 1986](#); [Mutch, Smith, & Yonas, 1983](#); [Tarr & Pinker, 1985](#); [Ullman, 1978](#)), the appearance or the quality of apparent motion is sensitive to manipulations of apparent three-dimensional depth (e. g. , [Attneave & Block, 1973](#); [Corbin, 1942](#)). Apparent-motion quality can be affected by manipulations of the meaningfulness, the consistency, or the familiarity of a display (e. g. , [Jones & Bruner, 1954](#)), but this is not the case for correspondence match assignment ([Dawson & Wright, 1989](#)). Similarly, the processes that generate the appearance of apparent-motion trajectories are sensitive to higher order stimulus variables, such as the topological structure of elements; motion correspondence processes are not ([Dawson, 1990a](#)).

These results indicate that although both the current model and the manifold approaches propose least-change transformations, they do so for different aspects of motion perception. Under the assumption that motion correspondence processing is a component of [Braddick's \(1980\)](#) long-range motion perception system, this indication leads to the speculation that the long-range system has at least two differentiable components: one for assigning motion correspondence matches and another for generating the appearances of objects during movement. [Ullman \(1979\)](#) took a similar position in proposing that the three-dimensional structure of moving objects is determined after motion correspondence matches have been assigned.

## Image-Matching Procedures

The motion correspondence network is not sensitive to the figural characteristics of elements, apart from their individuation from the background. This was justified earlier on the grounds that the human visual system is quite insensitive to manipulations of element appearance in apparent-motion displays (e. g. , [Baro & Levinson, 1988](#); [Burt & Sperling, 1981](#); [Dawson, 1989](#); [Kolers, 1972](#); [Kolers & Green, 1984](#); [Kolers & Pomerantz, 1971](#); [Kolers & von Grunau, 1976](#); [Krumhansl, 1984](#); [Navon, 1976](#); [Ullman, 1979](#), Chap. 2).

Some recent experiments have shown that this insensitivity is not complete. The motion correspondence matches assigned by the human visual system can be affected by certain figural properties, including orientation, spatial frequency, some aspects of geometric shape, and possibly some topological features (e. g. , [Chen, 1985](#); [Green, 1986](#); [Mack, Klein, Hill, & Palumbo, 1989](#); [Prazdny, 1986](#); [Ramachandran, Ginsburg, & Anstis, 1983](#); [Shechter, Hochstein, & Hillman, 1988](#); [Ullman, 1980b](#); [Watson, 1986](#)). Although these results are far from conclusive (see [Footnote 3](#) [Anstis's \(1980\)](#) proposal that the long-range motion perception system (which is presumed to include motion correspondence processing) detects motion after some figural analysis of a scene has occurred (see also [Ramachandran, Rao, & Vidyasagar, 1973](#)).

In principle, figural characteristics could be used to assign correspondence matches, although in practice the human visual system does not appear to rely on such a strategy. In an image-matching procedure, motion correspondence matches are made between elements of similar figural appearance. A dynamic scene can be sampled so quickly that drastic changes in appearances between frames of view are unlikely. Under such conditions, image matching usually produces the correct motion correspondence matches, as shown by many successful models in computer science (e. g. , [Aggarwal & Duda, 1975](#); [Ferrie, Levine, & Zucker, 1982](#); [Jain, Martin, & Aggarwal, 1979a, 1979b](#); [Jain, Militzer, & Nagel, 1977](#); [Price, 1985](#); [Price & Reddy, 1977](#); [Tsuji, Osada, & Yachida, 1979, 1980](#); [Yalamanchili, Martin, & Aggarwal, 1982](#)).

Image matching could be added to the current model in a rather straightforward manner. An additional constraint matrix, in which connection weights are based on figural characteristics, could be defined. One example would be a diagonal matrix of units connected to themselves, similar to matrix **NN**. This matrix would define connection strengths in terms of the number of figural changes from Frame 1 to Frame 2 that are entailed by each correspondence match. A unit representing few (if any) changes would receive a high, positive connection strength. A unit representing many changes would receive a lower (possibly negative) connection strength. In such a "winner-take-all" structure, motion correspondence matches between elements similar in appearance would be favored over matches between elements dissimilar in appearance. This image-matching matrix would be added to the other constraint matrices when **C** was created. Because this new matrix is symmetric, the convergence properties of the network would not be affected by its inclusion.

However, from some points of view, it may not be desirable or necessary to add image-matching capabilities to the network. In the next section, it is proposed that motion correspondence processing is used to track attentional tags assigned in parallel by visual cognition. In one proposed model ([Pylyshyn, 1988, 1989](#)), these attentional tags do not represent figural properties but only encode the locations of features or feature clusters

that have been individuated.

## Cognitive Penetrability and Motion Correspondence

Some researchers have argued that (long-range) apparent motion is mediated by processes that are "perceptually intelligent": that is, processes no different from those that mediate thinking and problem solving (e. g. , [Goodman, 1978](#); [Rock, 1983](#)). This amounts to the proposal that apparent-motion perception is affected by the contents of intentional states, such as beliefs, goals, or expectations, and is therefore cognitively penetrable (e. g. , [Pylyshyn, 1980, 1984](#)).

This general proposal about apparent motion could entail the specific contention that correspondence processing is also cognitively penetrable. If this were true, then correspondence processes would be sensitive to semantic interpretations of display characteristics (e. g. , [Fodor & Pylyshyn, 1981](#); [Pylyshyn, 1984](#)). As a result, correspondence processing could not be described as the automatic application of a small number of constraints because the assignment of correspondence matches could, in principle, depend on any representational content available to general inferential processes (e. g. , [Fodor, 1983](#)).

It is argued in the next section that correspondence processing is itself cognitively impenetrable: It is a primitive component of perceptual processing and is insensitive to representational contents. However, it is also argued that cognitive penetrability may characterize the processes that individuate elements to which correspondence matches are assigned.

### Cognitive impenetrability of correspondence matches.

Experimental evidence indicates that motion correspondence processing is not cognitively penetrable. [Dawson and Wright \(1989\)](#) showed that beliefs about motion consistency do not affect perceptions of motion competition displays. [Ullman \(1979, Chap. 1\)](#) discussed some demonstrations in which perceived direction of motion is inconsistent with expectations, such as a "block train" in which the windows are seen to move in one direction as the train itself moves in another. [Kolers and Green \(1984\)](#) argued that if correspondence processing was perceptually intelligent, the visual system would not ignore simple interpretations (such as perceived rotation) to resolve figural differences between frames of view (see also [Petersik, 1987, 1989](#)).

Furthermore, the data-driven correspondence network just described is capable of generating solutions to which perceptually intelligent processing has been ascribed. For example, [Attneave \(1974\)](#) suggested that the perception of motion symmetry may require perceptual intelligence. However, the network can generate correct symmetric motion solutions (see [Figure 8b](#)), as can minimal mapping theory ([Ullman, 1979](#)). Similarly, correspondence match assignments can be affected by a previous apparent-motion sequence (e. g. , [Ramachandran & Anstis, 1986b](#)). The sequence establishes expectations for correspondence matches that differ from those that would be assigned in the absence of the sequence. However, these expectations need not be semantic in nature. Expectations in a connectionist system can be modeled as the processing unit activations that persist after the network has completed some previous task. The current motion correspondence model is state dependent in this fashion. The solution to which the network converges depends on the activation values in  $\mathbf{a}_0$ , and these values can be preset to mimic expectancy effects.

These results do not imply that all aspects of apparent motion perception are cognitively impenetrable. The processes that generate the appearance of elements in motion do appear to be sensitive to intentional states. [Kolers \(1972\)](#) proposed that "when meaningfulness plays a role, its chief effect is on the clarity and vividness of the motion perception" (p. 129). [Jones and Bruner \(1954\)](#) showed that meaningful apparent motion is easier to perceive and is better maintained in suboptimal conditions than is apparent motion that is less interpretable or meaningful. Nevertheless, the higher order factors that influence the quality or vividness of motion do not affect the correspondence matches assigned in motion competition displays (e. g. , [Dawson, 1990a](#); [Dawson & Wright, 1989](#); [Ullman, 1978](#)).

### Cognitive penetration and element individuation.

However, it may be that the elements that enter into motion correspondence processing are individuated by top-down, cognitively penetrable processes. An enormous variety of proximal stimulus conditions can be used to define elements to which motion correspondence matches are later assigned. For example, apparent motion (and motion correspondence matches) can be determined for such visual entities as simple points of light (e. g. ,

Kolers, 1972, Chap. 1), oriented line segments (e. g. , Ullman, 1980b), spatial frequencies (e. g. , Green, 1986; Ramachandran et al. , 1983), geometric shapes (e. g. , Kolers & Pomerantz, 1971), rigid configurations of discrete stimulus parts (e. g. , Rock, 1983, Figure 7-6; Ullman, 1979, Figure 2-20), dynamic "clouds" (e. g. , Marr, 1982, pp. 191-192; Pantle & Picciano, 1976), topologically defined entities (e. g. , Chen, 1985; Prazdny, 1986), segregated textures (e. g. , Ramachandran & Anstis, 1986a; Ramachandran et al. , 1973), and subjective figures (e. g. , Mather, 1988; Ramachandran, 1985). Furthermore, entities defined haptically or auditorally can also be used to produce apparent motion that follows spatiotemporal rules similar to those that govern visual apparent motion (e. g. , Perrott, 1982; Sherrick & Rogers, 1966) and may possibly be mediated by similar (or, as suggested later, identical) mechanisms.

Given the enormous variety of stimulus conditions just listed one is hard pressed to define a set of necessary and sufficient conditions for the automatic, preattentive individuation of apparent motion elements. "Shape, contour, or form is itself rather plastic and unstable; nor is it a monolithic property of objects" (Kolers, 1983, p. 27). Instead, element individuation may be largely determined by a variety of processes that work from the top down, reflecting the computational needs, the goals, and the relevant knowledge of the observer (cf. Ullman, 1984).

For example, Ternus (1938, pp. 155-156) found if subjects attended to the middle elements of the Figure 9 configuration, the element motion percept was produced. In addition, if some elements can be interpreted to be occluding surfaces, then the assignment of correspondence matches can change (e. g. , Ramachandran & Anstis, 1986b; Ramachandran, Inada, & Kiama, 1986; Rock, 1983, pp. 167-172; Sigman & Rock, 1974).

## Visual attention and tag assignment.

The two preceding subsections can be summarized as follows: (a) The motion correspondence processes that maintain the identity of individuated elements work automatically and preattentively and are cognitively impenetrable. (b) The processes that actually individuate elements are cognitively penetrable; as a result, there is no finite set of necessary and sufficient properties to automatically individuate elements before correspondence processing. These conclusions suggest that the motion correspondence network can model an important aspect of visual cognition, as is discussed later.

Pylyshyn (1988, 1989) proposed a model for the simultaneous spatial indexing of proximal stimulus elements. Feature locations are indexed with an attentional tag called a FINST (for instantiation finger); FINSTs do not encode any properties of a feature cluster (e. g. , shape, color) but instead permit the cluster to be located for further examination. It is presumed that FINSTs are assigned in parallel across the visual field but that only a finite number can be assigned. FINST assignment may also be influenced by top-down processes.

The FINST mechanism is thought to play several important roles in visual cognition. First, spatial relations among individuated elements can be determined once FINSTs have been assigned. Second, FINSTs can be used to construct a geocentric (i. e. , a nonretinotopic) visual representation. "If the retinal feature cluster identified in this way maintains a reliable correlation (over time) with some particular feature of the distal scene, then the FINST will succeed in pointing to that distal feature, independent of its location on the retina" (Pylyshyn, 1989, p. 69; see also pp. 73-77). Third, if a FINST refers to a geocentric location, then it can be used to coordinate perceptual-motor functions.

In order to provide these three functions, the FINST mechanism must solve what Strong and Whitehead (1989) called the tag-assignment problem: A FINST must keep pointing to the same feature cluster as the cluster changes retinal position, even though the cluster may undergo changes in appearance because of its movement. Pylyshyn (1988, 1989) assumed that tag assignment is a primitive property of his model, but he did not provide details of how it is realized in the FINST mechanism.

## Motion correspondence and tag assignment.

The motion correspondence network is capable of solving the tag-assignment problem (see also Dawson, 1989) and therefore can be considered as a model of how FINST identities are maintained. There are several reasons for this proposal.

First, the network is a model of preattentive (i. e. , cognitively impenetrable) identity tracking, and Pylyshyn (1988, 1989) assumed that tag assignment is performed preattentively. Second, the network can track the identities of several elements in parallel, which is required by the FINST mechanism. Third, the network requires

only element positions as input, and feature location is the only information represented by a FINST. Fourth, the network is not defined by units with fixed and limited receptive fields and, as a result, can track elements as they move to arbitrary locations in the visual field (see also [Cavanagh & Mather, 1989](#)). Fifth, the assignment of FINSTs (i. e. , the individuation of elements) is affected by top-down processes, which is also likely true for the elements to which motion correspondence matches can be assigned.

One implication of this proposal is that motion correspondence processing is of limited capacity. This is because experimental evidence indicates that only a small number of FINSTs can be assigned to a display at one time ([Pylyshyn & Storm, 1988](#)). This raises the possibility that motion correspondence processing, and perhaps apparent-motion perception in general, can be greatly affected by increasing the attentional load of displays or by having subjects attend to display parts instead of integrated display wholes (e. g. , [Ternus, 1938](#)).

[Cavanagh and Mather \(1989\)](#) were skeptical about the contention that apparent-motion perception has a limited capacity because motion can be seen in random-dot kinemategrams that are used to study short-range motion (e. g. , [Braddick, 1974](#)) and that are often defined by several hundred moving dots. However, this argument neglects the fact that correlator models of short-range motion function without individuating stimulus dots as independent elements (see also [Daugman, 1988](#)). These models require as input only the raw gray-level intensities of the proximal stimulus. In other words, although motion detection—modeled by correlator mechanisms—may not be capacity limited, tag assignment—modeled by different mechanisms, such as the network described earlier—may indeed be subject to capacity limitations.

A second implication of the proposed relation between motion correspondence processing and the tag-assignment problem is that the quantitative similarities among visual, auditory, and haptic apparent motion (e. g. , [Perrott, 1982](#); [Sherrick & Rogers, 1966](#)) may exist because all three are mediated by the same principles. For example, [Pylyshyn \(1989\)](#) proposed that the FINST mechanism is involved in coordinating processing across modalities. Specifically, visual locations that have been assigned FINSTs can be cross-bound to similar attentional markers (called ANCHORS) in the frame of reference used by motor processes. This cross-binding accounts for such capacities as being able to point at things being looked at. The proposal that locations available to different modalities must be tagged by different kinds attentional tokens (i. e. , FINSTs for vision; ANCHORS for motor systems; other kinds of tags for other modalities) also implies that the tag-assignment problem must be solved for each coded modality. If attentional tags for each modality have similar properties (i. e. , if they represent only feature locations and not feature properties), then the tag-assignment problem could be solved for each modality by applying the same constraining principles and by using the same kind of processing mechanism.

A third implication of the proposed relation between motion correspondence and tag assignment concerns identifying the neural substrate responsible for such processing. If the assignment of motion correspondence matches is required merely for the perception of motion, strong constraints are not placed on where in the brain it occurs; any physiological structure demonstrating a high degree of sensitivity to stimulus movement is a plausible candidate. However, if motion correspondence processing is also involved in the attentional tracking of elements defined in different modalities, sensitivity to movement is a necessary but not sufficient characteristic of the physiological medium. In addition, the neural mechanisms for motion correspondence must be sensitive to elements defined in different sensory modalities and must also be at least partly responsible for directing attention. As is shown in the next section these additional constraints—caused solely by the attentional tracking hypothesis—are very useful in determining the likely physiological locus of motion correspondence processing.

## Part 4: Toward the Physiology of Motion Correspondence Processing

In this final section, the properties of the network detailed in Part 2 are related to what is known about the physiological mechanisms underlying human motion perception. It is argued that many of these physiological properties are consistent with the motion correspondence model. It is also argued that new types of physiological measurements—measurements of cell responses during the tracking of *multiple* moving elements—are still required to establish how the brain assigns motion correspondence matches.

The goals of this discussion are quite modest. It was noted earlier that the autoassociative network was not a model of performance but instead a model of competence. Thus no attempt is made to identify specific associations between each network component and particular physiological structures. Instead, an attempt is made to establish general links between the model and the information that is explicitly coded by a physiological pathway for motion processing.

This approach is consistent with the connectionist framework in which the model has been formulated. On the

one hand, many connectionist researchers have argued that their networks do not specify underlying neural circuitry but instead represent algorithms that might plausibly be carried out by brain mechanisms (e. g. , [Rumelhart & McClelland, 1985](#); [Smolensky, 1988](#)). On the other hand, there is a strong commitment among connectionist researchers to link their networks to physiological processes. Connectionists bear a particular responsibility for establishing biological constraints on their networks because connectionist models are often described as being much more neuronally inspired than are other models in cognitive science (e. g. , [Clark, 1989](#)). A connectionist committed to determining the biological plausibility of a particular model attempts to show how design decisions that guided a network's construction (e. g. , decisions about what processing units in a network might represent, about the kinds of connections among processing units, and about the kinds of changes that occur in the network over time) can be related to the physical properties of some relevant neural substrate. The sections that follow demonstrate that many of the design decisions underlying the autoassociation network for motion correspondence are consistent with what is known about the physiological processes underlying human motion perception.

## The Motion Pathway in Vision

Evidence from anatomical, physiological, and clinical neuroscience studies has led many researchers to suggest that there exist parallel physiological pathways in the human visual system (e. g. , [Livingstone & Hubel, 1988](#); [Maunsell & Newsome, 1987](#); [Ungerleider & Mishkin, 1982](#)). Each pathway is argued to be responsible for the processing of different kinds of visual information. In particular, one major pathway appears to be specialized for the processing of visual form (i. e. , specifying *what* an object is), and a second major pathway appears to be specialized for the processing of visual motion (i. e. , specifying *where* an object is). These two pathways are typically assumed to be distinct and independent, which suggests that the perception of motion is mediated by physiological processes independent of those that mediate the perception of form. Although such an extreme view is unlikely to be completely correct (see [DeYoe & van Essen, 1988](#)), Livingstone and Hubel demonstrated that it is a powerful heuristic for uniting many results of physiological and psychological studies.

The existence of a distinct pathway for the processing of visual motion is supported by a variety of findings. First, it has been discovered that damage to certain areas of the human brain produces severe deficits in the perception of motion but has little effect on the perception of form (e. g. , [Hess et al. , 1989](#); [Zihl et al. , 1983](#); see [Botez, 1975](#) for examples of lesions that affect form perception but do not affect motion perception). Second, physiological recordings of cell responses to stimuli have revealed many neurons that are highly sensitive to stimulus movement but are not affected by the color, the size, or the shape of stimuli (e. g. , [Albright, 1984](#); [Albright et al. , 1984](#); [Dubner & Zeki, 1971](#); [Maunsell & van Essen, 1983b](#); [Rodman & Albright, 1987](#); [Zeki, 1974](#)). Third, anatomical evidence reveals patterns of neural connectivity that appear to define an anatomical pathway for processing motion. Specifically, there exist rich interconnections between cell regions that are highly sensitive to motion, whereas relatively few connections exist between such cell regions and those that are highly sensitive to other stimulus properties, such as color (e. g. , [Livingstone & Hubel, 1988](#)).

Functional architecture for the motion pathway in vision. (LGN = the magnocellular laminae of the lateral geniculate nucleus. The parvocellular laminae of this structure are part of the form pathway. V1 = the striate visual cortex. Only layers 4Ca and 4B of the striate cortex are thought to be part of this pathway. V2 = Area 2 of visual cortex, of which only the thick stripes are part of the motion pathway. MT = the middle temporal visual area, located within the superior temporal sulcus. Note that this component of the pathway receives direct input from both V1 and V2. MST = the medial superior temporal area, also located within the superior temporal sulcus. Area 7 = part of the parietal association cortex. It is argued in the text that motion correspondence matches are assigned by this component of the pathway on the basis of stimulus properties coded in MT and MST. )

Figure 12 depicts the major components of the motion pathway and illustrates their direct interconnections. This figure is highly simplified; there likely exist many additional connections between this network and other physiological components of the visual system ([Desimone & Ungerleider, 1986](#); [DeYoe & van Essen, 1988](#); [Maunsell & van Essen, 1983a](#)). In addition, actual patterns of connectivity are much more complicated than illustrated; in most cases, direct connections between components are reciprocal (e. g. , [Maunsell & Newsome, 1987](#)). Nevertheless, [Figure 12](#) captures the basic structure of what physiologists describe as a pathway for the processing of motion.

## Where in the Motion Pathway Is Correspondence Computed?

The perception of motion serves as the foundation for many different functions, including the segregation of

figure and ground, the perception of depth, the coordination of eye movements, and locomotion through space (e. g. , [Gibson, 1979](#); [Lisberger, Morris, & Tychsen, 1987](#); [Nakayama, 1985](#); [Regan, 1986](#)). Each of these functions also requires different kinds of motion measurements or computations (e. g. , [Ullman, 1981](#)). That the motion pathway is likely to mediate many (if not all) of these measurements suggests that each component of the pathway is probably responsible in part for many functions. In this section it is argued that one of the several functions likely to be carried out by Area 7 of the posterior parietal cortex is the assignment of correspondence matches. It is also argued that these matches are assigned on the basis of motion measurements that are not coded until relatively late in the pathway; that is, they are coded in the middle temporal (MT) and medial superior temporal (MST) areas of the superior temporal sulcus. These points are made by a consideration of the properties required of a neural substrate if it were to implement correspondence processing of the type described in Part 3.

## **The neural substrate must be sensitive to individuated elements.**

It was argued earlier that motion correspondence matches are assigned after the to-be-tracked moving elements had been individuated (e. g. , [Anstis, 1980](#); [Ramachandran & Anstis, 1986a](#)). In the motion pathway, discrete elements are probably not individuated until the MT stage, which indicates that this stage marks the earliest point in the motion pathway at which correspondence processing could occur. There are two general types of evidence for this claim.

First, [Allman, Miezin, and McGuinness \(1985\)](#) found that the MT area is characterized by very large directionally selective receptive fields that have an antagonistic organization. Thus a cell in the MT area responds vigorously when a small central region of its receptive field is being stimulated by movement in one direction at the same time that the remaining part of the receptive field is being stimulated by coherent movement in a different direction. If the entire receptive field is stimulated by coherent motion in a single direction, the cell will not respond. It appears that one major function of MT cells is to individuate elements from their background on the basis of motion information. Recall that elements defined in terms of relative motion can serve as inputs to human processes of motion correspondence (e. g. , [Pantle & Picciano, 1976](#); [Prazdny, 1986](#)). In the primate visual system, such elements do not appear to be defined before the MT stage.

Second, [Chubb and Sperling \(1988\)](#) showed that the Reichardt class of detector often used to model early motion perception (e. g. , [Dawson & Di Lollo, 1990](#); [Reichardt, 1961](#); [van Santen & Sperling, 1984, 1985](#)) is incapable of detecting the coherent, global movement of certain figural patterns (e. g. , the plaid pattern produced by superimposing two drifting sinusoid gratings). However, human observers can easily detect such motion (e. g. , [Chubb & Sperling, 1988](#); [Lelkens & Koenderink, 1984](#); [Victor & Conte, 1990](#)). [Movshon, Adelson, Gizzi, and Newsome \(1984\)](#) showed that MT neurons respond to pattern motion. Furthermore, [Movshon et al.](#) showed that cells located earlier in the motion pathway (i. e. , in the striate cortex) are not sensitive to the motion of global patterns. On the basis of his own physiological studies of the MT area, [Albright \(1984\)](#) proposed a network for the detection of pattern motion. Albright's model works by tracking small, unoriented subunits of the moving stimulus; thus his mechanism for pattern motion detection in the MT area requires that parts of a moving stimulus be individuated.

## **The neural substrate must be sensitive to element locations.**

A second important property of motion correspondence processing, defended earlier on both empirical and theoretical grounds, is that motion correspondence matches are assigned primarily on the basis of element locations and not on the basis of element appearances. Physiological evidence indicates that the parietal cortex of the human brain is involved primarily in the representation of spatial locations. For example, damage to the posterior parietal lobe produces severe deficits in the ability to accurately reach toward a visual target (e. g. , [Damasio & Benton, 1979](#); [Ratcliff & Davies-Jones, 1972](#)) and can also produce severe deficits in the ability to perceive motion (e. g. , [Hess et al. , 1989](#); [Newsome & Pare, 1988](#); [Zihl et al. , 1983](#)). However, damage localized in the parietal cortex appears to have little effect on the processing of visual form. The latter parts of the motion pathway, beginning with the MT area, are near or in parietal cortex and are thus located in a region of the brain that appears to be responsible for representing and processing location information. For instance, the MT area is a small region on the posterior bank of the superior temporal sulcus (e. g. , [Maunsell & Newsome, 1987, Figure 2](#)). This portion of the superior temporal sulcus is surrounded by the angular gyrus of the inferior parietal lobule ([Barr, 1974, p. 211](#)). The MST area lies immediately adjacent to the MT area in the medial part of the superior temporal sulcus. Area 7 is located in the posterior part of the parietal lobe.

In addition, of course, the physiological evidence used to defend the notion of distinct pathways for processing motion and form emphasizes the role of stimulus location. Specifically, many MT neurons (e. g. , [Albright, 1984](#); [Albright et al. , 1984](#); [Dubner & Zeki, 1971](#); [Maunsell & van Essen, 1983b](#); [Rodman & Albright, 1987](#); [Zeki, 1974](#)) and MST neurons (e. g. , [Saito et al. , 1986](#); [Tanaka et al. , 1986](#)) are insensitive to the figural appearances of stimuli. These neurons respond to element motion and not to element size, color, or shape.

## **The neural substrate must have very large receptive fields.**

Under the traditionally assumed two-process distinction that has guided motion perception research over the last two decades, the motion correspondence problem is assumed to be solved by the long-range process. This is because classical apparent motion can be perceived between elements separated by several degrees of visual angle in the visual field; the short-range process is assumed to be unable to detect motion over these large distances. The physiological implications of this view are quite straightforward: The neural substrate involved in assigning motion correspondence matches must have very large receptive fields.

[Cavanagh and Mather \(1989\)](#) argued that the long-range motion system should not be interpreted as being responsible for element tracking because there is no physiological evidence for receptive fields large enough to accomplish this task throughout the visual field. However, recent experiments showed that this argument is incorrect. [Allman et al. \(1985\)](#) demonstrated that classical techniques used to map out receptive field sizes (i. e. , the use of small bars or spots as stimuli) vastly underestimate the effective receptive field size of MT cells. This is because these classical techniques do not stimulate the peripheral components of the receptive fields in such a way as to generate maximal responses from the cells (i. e. , classical stimuli do not stimulate the antagonistically organized receptive fields of these cells). When stimuli designed to more adequately stimulate the peripheral parts of the receptive field are used, the size of the receptive field of most MT cells appears to be 50 to 100 times as large as that revealed through the use of classical stimuli. Furthermore, in terms of absolute size, the techniques of Allman et al. reveal that MT receptive fields are enormous. Allman et al. estimated that the smallest receptive field size that they observed was  $1,200^{\circ 2}$ , the largest was estimated to be between  $4,900^{\circ 2}$  and  $7,000^{\circ 2}$ . Because the total visual field for the animals studied was approximately  $20,000^{\circ 2}$ , these estimates suggest that the receptive field of a single MT cell could cover between 6% and 35% of the entire visual field (for similar estimates, see [Saito et al. , 1986](#); [Tanaka et al. , 1986](#)). Because it is commonly assumed that the receptive fields of cells late in visual pathways are constructed by synthesizing or combining receptive fields from cells earlier in the pathway (e. g. , [Kuffler, Nicholls, & Martin, 1984](#), pp. 64-67), this suggestion implies that the receptive field size of cells located in the MST area or in Area 7 are even larger than those found in the MT area. Indeed, [Robinson, Goldberg, and Stanton \(1978, p. 922\)](#) reported that the receptive fields of neurons in Area 7 frequently cover one or two quadrants of the visual field; on occasion, receptive fields have been found to cover the entire visual field (see also [Motter & Mountcastle, 1981](#)).

## **The neural substrate must be involved in tracking objects.**

To convincingly argue that some particular neural substrate is responsible for assigning correspondence matches, one must clearly demonstrate that this substrate is responsible for object tracking. Evidence concerning this object-tracking criterion places correspondence processing later in the motion pathway than in the MT area.

In examining the functional properties of Area 7, many researchers have observed cells that appear to mediate object tracking (e. g. , [Goldberg & Bruce, 1985](#); [Hyvärinen & Poranen, 1974](#); [Lynch, Mountcastle, Talbot, & Yin, 1977](#); [Motter & Mountcastle, 1981](#); [Robinson et al. , 1978](#); [Sakata, Shibutani, Kawano, & Harrington, 1985](#)); such cells are not evident earlier in the motion pathway. *Visual fixation neurons* produce sustained responses when a target is fixated; some of these cells respond only when the target occupies a preferred spatial location. *Visual tracking neurons* respond when a moving stimulus is tracked by smooth-pursuit eye movements. Visual tracking neurons also respond when a stationary object is fixated while another object moves, but they do not respond with stationary fixations alone (e. g. , [Robinson et al. , 1978](#)). Many visual tracking neurons exhibit a preferred direction of object motion. *Saccade neurons* respond up to 150 ms before a saccadic eye movement to a designated target. Saccade neurons do not respond during stationary fixations, during object tracking, or before spontaneous saccades.

## **The neural substrate must track elements defined in different sensory modalities.**

The object-tracking criterion can be elaborated to place additional constraints on localizing motion correspondence processing in a particular physiological structure. In Part 3 it was argued that motion correspondence processing was responsible for the primitive tag-assignment function in [Pylyshyn's \(1988\)](#) model of spatial indexing. One consequence of this argument was that the principles governing correspondence processing in the visual modality should also govern object tracking in other sensory modalities (i. e. , to track the identities of visual FINSTs and of haptic ANCHORS). In physiological terms, this implies that the neural substrate responsible for visual tracking must also be responsible for other kinds of tracking (e. g. , tracking of haptic objects). This hypothesis is supported by experimental studies of Area 7. For example, [Robinson et al. \(1978\)](#) described the properties of *hand projection neurons*. These neurons respond to targets to which hand movements are to be directed. Hand projection neurons do not respond to the reaching movement in the absence of a visual stimulus or to the visual stimulus in the absence of the reach. [Hyvärinen and Poranen \(1974\)](#) also observed many Area 7 cells that responded during manual reaching, tracking, or manipulation and noted that many of these cells had a preferred direction of reaching. It is likely that damage to these types of cells are responsible for deficits in visual localization (i. e. , accurate reaching toward visual targets) observed in humans with parietal lobe lesions (e. g. , [Damasio & Benton, 1979](#); [Ratcliff & Davies-Jones, 1972](#)).

## **The neural substrate must mediate attentional processing of tracked elements.**

One further implication of the putative relation between motion correspondence and tag-assignment processing concerns the involvement of visual attention. [Pylyshyn and Storm \(1988\)](#) showed that human observers are capable of simultaneously tracking several target elements individuated from distractors by attention alone. The hypothesis developed in this article is that this tracking must be mediated by motion correspondence processes (see Part 3). Therefore, the neural substrate responsible for such processing must be capable of distinguishing attended from nonattended objects, even if both types are equally visible. Again, Area 7 exhibits this capability. Without exception, researchers who examine the properties of neurons in Area 7 have noted that these neurons are governed by strong extraretinal influences (e. g. , [Goldberg & Bruce, 1985](#); [Hyvärinen & Poranen, 1974](#); [Lynch et al. , 1977](#); [Robinson et al. , 1978](#); [Sakata et al. , 1985](#); see also [Hurlbert & Poggio, 1985](#)). Specifically, strong responses in all of the different types of Area 7 neurons occur only when the object stimulating their receptive fields is being attended to or is of some interest to the observing animal. If the object is not being attended to, the Area 7 neurons have very weak responses (for a striking example, see [Robinson et al. , 1978](#), Figure 17).

## **Summary.**

This evidence concerning the tracking of attended objects clearly indicates that Area 7 in the posterior parietal lobe is the major locus for motion correspondence processing. However, this does not also indicate that other parts of the motion pathway are unimportant for the assignment of motion correspondence matches. For instance, [Lisberger et al. \(1987, pp. 108-117\)](#) noted that oculomotor pursuit depends on the kind of information represented in the MT area, even though the evidence just given indicates that pursuit processing itself occurs later in the motion pathway. This suggests that motion correspondence matches are assigned by neural mechanisms in Area 7, but this assignment depends on measurements performed earlier in the motion pathway, particularly in the MT and MST areas.

## **Neural Measurements for Motion Correspondence Processing**

In the preceding section it was argued that correspondence matches are assigned in Area 7 by a consideration of general characteristics of motion correspondence processing: characteristics that should be true of any correspondence model. In this section, the biological plausibility of the model detailed in Part 2 is explored by means of determining whether there is any evidence that the specific motion measurements that it requires are made in the MT or the MST area.

## **The nearest neighbor principle.**

According to both the current model and minimal mapping theory (e. g. , [Ullman, 1979](#)), the nearest neighbor principle is required to assign motion correspondence matches. In the current model, this constraint is

implemented by an adjustment of connection weights to favor short motion correspondence matches. Because a motion correspondence match can also be viewed as a directed distance traveled over a sampled time (i. e. , the time separating Frame 1 from Frame 2), this is equivalent to saying that the model prefers slow element velocities. Physiological evidence indicates that the stimulus measurements represented by MT cells are well suited for implementing the nearest neighbor principle.

MT cells appear to encode the velocity of moving elements and thus implement one important measurement required by the model. Many experiments have shown that MT cells have both a preferred direction and a preferred speed of stimulus motion (e. g. , [Albright, 1984](#); [Dubner & Zeki, 1971](#); [Maunsell & van Essen, 1983b](#); [Mikami, Newsome, & Wurtz, 1986](#); [Rodman & Albright, 1987](#)). The range of speeds encoded by these neurons is rather broad, ranging from 2°/s to 256°/s (e. g. , [Maunsell & van Essen, 1983b](#), p. 1137).

At first glance, the broad range of preferred speeds for MT neurons seems inconsistent with the nearest neighbor principle, as is evidence that MT cells allow faster velocities to be coded than do directionally selective cells in Area V1 (e. g. , [Mikami et al. , 1986](#); [Newsome, Mikami, & Wurtz, 1986](#)). However, a closer examination of the *distribution* of speed sensitivities in the MT reveals that this distribution is consistent with a preference for slow velocities. Although the MT area encodes a broad range of speeds, most individual MT cells are most sensitive to slow or intermediate speeds. For example, all of the 70 cells studied by [Dubner and Zeki \(1971\)](#) responded to speeds ranging from 1°/s to 5°/s. Similarly, more than half of the 109 cells examined by [Maunsell and van Essen \(1983b, Figure 7a\)](#) had preferred speeds of 32°/s or less. [Zeki \(1974\)](#) estimated that the optimal speeds for a population of MT cells ranged between 5°/s and 50°/s, although his methods limited the accuracy of this estimate. In sum, the evidence indicates that although MT cells can encode many stimulus velocities, slow velocities are much more likely to elicit responses in MT cells than are fast velocities.

An additional assumption in the current model is that the nearest neighbor principle is defined through the use of two-dimensional coordinates; motion correspondence matches are not assigned in three-dimensional space. Again, the characteristics of neural responses in the MT area are consistent with this assumption. As can be seen in [Figure 12](#), the MT area receives direct input from Area V2 in the visual cortex, which encodes information about the binocular disparity of stimuli (e. g. , [Livingstone & Hubel, 1988](#)). Thus in principle the MT area could represent the three-dimensional velocities of elements. However, this does not appear to be done in practice. Instead, MT neurons appear to encode the two-dimensional velocity of elements at fixed horizontal disparities (i. e. , element movement in fixed frontoparallel planes). For example, [Maunsell and van Essen \(1983c\)](#) reported that some MT neurons code the velocity of an element moving in a plane that is far from the observer, whereas others code the velocity of an element moving in a plane that is near to the observer. "To our surprise, no neurons in our sample from MT were truly selective for motion in depth in the sense of responding maximally to stimuli that simulated movement with components toward or away from the animal" (p. 1149). This property, although surprising to Maunsell and van Essen, is assumed in the current model and in minimal mapping theory (e. g. , [Ullman, 1979](#)) and is consistent with psychophysical evidence that the human visual system does not assign correspondence matches in three-dimensional space (e. g. , [Dawson & Wright, 1989](#); [Ullman, 1978](#)).

## The relative velocity principle.

The current motion correspondence model can be differentiated from [Ullman's \(1979\)](#) minimal mapping theory in exploiting a relative velocity principle, which encourages the assignment of similar motion correspondence matches (or, equivalently, similar velocities) to elements near one another in Frame 1. With respect to establishing the biological plausibility of the relative velocity principle, two questions must be considered: First, do there exist neurons that are explicitly sensitive to relative velocity information? Second, is there any physiological evidence that measurements of relative velocity are integrated in such a way that the relative velocity principle is directly implemented?

With respect to the first question, cells sensitive to the relative movements of stimuli have been found in many areas of the brain, including the superior colliculus ([Mandl, 1985](#)), Area V1 ([Bridgeman, 1972](#); [Hammond, Ahmed, & Smith, 1986](#); [Hammond, Pomfrett, & Ahmed, 1989](#); [Hammond & Smith, 1984](#); [Kaji & Kawabata, 1985](#); [Orban, Gulyas, & Vogels, 1987](#)), Area V2 ([Orban, Gulyas, & Spileers, 1988](#)), the MT area (e. g. , [Allman et al. , 1985](#)), the MST area ([Saito et al. , 1986](#); [Tanaka et al. , 1986](#)), and Area 7 (e. g. , [Motter & Mountcastle, 1981](#); [Sakata, Shibutani, & Tsurugai, 1986](#)). The seemingly ubiquitous presence of such detectors is consistent with the importance of relative motion information for many visual functions (for reviews, see [Nakayama, 1985](#); [Regan, 1986](#)) and with the results of many psychophysical experiments showing that human observers are highly sensitive to such motion (e. g. , [Cutting & Proffitt, 1982](#); [Dawson, 1987](#); [Gogel, 1974](#); [Johansson, 1950](#); [Proffitt &](#)

Cutting, 1979, 1980; Ramachandran & Anstis, 1985, 1986b).

With respect to the second question, it must be determined whether relative velocity measurements are integrated in a manner consistent with the relative velocity principle: Do there exist any cells that have a preference for several elements moving with similar velocities (i. e. , cells that prefer small relative velocities)? As argued earlier, the MT and MST areas and Area 7 were the most likely sites for motion correspondence processing; thus in consideration of this issue, attention is restricted to these locations.

The relative velocity measurements instantiated in the MT area do not appear to be consistent with the relative velocity principle. Allman et al. (1985) showed that directionally selective MT neurons have an antagonistic organization to their receptive fields (see also Tanaka et al. , 1986). MT cells have a strong preference for stimuli in which a small central region is moving in one direction and in which the background is moving coherently in some other direction (

Examples of relative motion stimuli used to study the response properties of neurons in the motion pathway. (a) An antagonistic stimulus in which a background pattern moves coherently in a direction that is different from that moved by a central figure. (b) A coherently moving texture that elicits strong responses from field cells in medial superior temporal (MST) area. (c) The movement of a single element elicits a strong response in figure cells in the MST areas. (d) A proposed stimulus for the study of the relative velocity principle. (If this principle is indeed instantiated physiologically, then there should exist cells that have much stronger responses to this stimulus than to the stimuli depicted in Parts a and c. )

Figure 13a). These cells do not respond when the entire receptive field is stimulated by coherent motion in a single direction (Figure 13b). Thus these cells prefer nonzero relative velocities, in contrast to the processing units in the computer simulation. However, this is not particularly surprising, for two reasons. First, motion correspondence matches are assumed to be assigned to individuated elements, and it was argued earlier that the antagonistic receptive fields in the MT area provide an important mechanism for this individuation. Second, the preceding section indicated that the nearest neighbor principle is applied to velocity measurements instantiated in the MT area. Because the relative velocity principle requires measurements of the differences between element velocities, it is plausible to expect that this principle is implemented *after* the MT stage in the motion pathway.

Indeed, there is evidence indicating that some of the directionally selective MST neurons behave in a manner consistent with the relative velocity principle. Tanaka et al. (1986) examined 519 MST neurons, of which the majority (285) were classified as being directionally selective. These directionally selective cells were assigned to three different categories: *Figure cells* had strong responses to antagonistic patterns of motion (Figure 13a) and to the motion of single elements (Figure 13c) but did not respond to the coherent motion of entire fields (Figure 13b). *Nonselective cells* had equally strong responses to the motion of single elements and to field motion, provided that the motion was in a preferred direction. *Field cells* had strong responses to the motion of coherent fields but did not respond to the motion of single elements. It is the behavior of field cells that is consistent with the relative velocity principles: These cells show little response to a single element moving in a preferred direction, but they show a strong response when several elements move in this direction. Furthermore, these cells are relatively common. The ratio of occurrence of the three types of cells is 2:3:2, respectively, and so it can be estimated that field cells constitute more than 15% of the entire population of MST neurons. In addition, field neurons are unique to the MST areas they are not found in the MT area.

Although the existence of field neurons in the MST area is consistent with the relative velocity principle, their behavior does not provide conclusive evidence for the implementation of the principle. This is because it is unclear as to whether field cells are responding strongly to the small relative velocities of several individuated elements or are instead merely responding to a coherent flow of motion that does not require any individuation of elements at all. Weak arguments can be made in favor of the former case. First, sensitivity to pattern motion exists at this later stage in the motion pathway (Movshon et al. , 1984), and this sensitivity may depend on the individuation of pattern parts (Albright, 1984). Second, Tanaka et al. (1986, p. 141) reported that there is some tendency for field cells to prefer patterns composed of large dots ( $2^{\circ}$ - $4^{\circ}$  in diameter), which suggests that these cells are sensitive to the component elements of the pattern and not just to the unindividuated luminance profile of the pattern.

However, conclusive support for the biological plausibility of the relative velocity principle requires that the responses of directionally selective cells be tested with a slightly more complicated stimulus pattern than has been used previously. This stimulus would consist of at least two moving elements, individuated from the background in terms of relative motion (see Figure 13d). Cells that implement the relative velocity constraint should exhibit stronger responses to this type of stimulus than to the single element stimulus depicted in Figure

13a. Furthermore, the response should be weakened if the two individuated elements in [Figure 13d](#) move at different speeds and should be strengthened if the two similarly moving elements are placed close together in the visual field (e. g. , [Dawson, 1987](#)). These types of responses, if they were indeed found in the MST area, could be implemented in two ways: in a single field neuron, whose response became stronger as more similarly moving elements were added to a display, or in a figure neuron, through excitatory and inhibitory connections to other figure neurons.

One possible reason why stimuli of the type depicted in [Figure 13d](#) have not been used to study cell responses is that physiological researchers have, quite naturally, assumed that the motion pathway is designed to create rich representations of object movement in three-dimensional space. As a result, they have examined stimuli whose properties are important determinants of such rich representations (see [Nakayama, 1985](#)). Although there is little doubt that the motion pathway is indeed well-suited for coding sophisticated properties of object movement, the preceding arguments indicate that it also serves a related but distinct function: the tracking of the identities of multiple moving targets. Presumably, one would consider studying only the physiological effects of stimuli like [Figure 13d](#) after realizing that such a tracking function is also mediated by the motion pathway.

## Field effects and relative motion.

Using the terminology applied by [Julesz \(1981\)](#) to texture perception, one can classify motion measurements in terms of their order of complexity. Measurements that require only consideration of the properties of single points on an image (e. g. , measuring the velocity of each image element) are called *first-order* measurements. Minimal mapping theory, in focusing on the nearest neighbor principle, depends on only such measurements. Measurements that require consideration of the properties of pairs of image elements (e. g. , measuring differences between the velocities of two elements) are called second-order measurements. The model detailed in Part 2 differs from minimal mapping theory by exploiting a second-order measurement, relative velocity.

A *field effect* is a principle for the assignment of motion correspondence matches that depends on motion measurements of second or higher order. A field effect results when the motion correspondence matches assigned in one part of the visual field affect which matches are assigned in other parts of the visual field. Field effects provide additional experimental evidence against the independence assumption in minimal mapping theory (i. e. , against the assumption that the correspondence problem can be solved through the use of only firstorder measurements). The relative velocity principle defines a second-order field effect: Similar (and, in many cases, identical) motion correspondence matches are assigned to neighboring display elements. The implementation of this second-order field effect in the model allows it to account for the effect of an unambiguous moving context on correspondence match assignment in a motion competition display ([Dawson, 1987](#); see [Figure 7f](#)).

However, correspondence processing can be affected by field effects of higher order: effects that require the comparison of complex *patterns* of motion. For example, [Ramachandran and Anstis \(1985\)](#) used an ambiguous apparent motion display like that depicted in

Higher order field effects in apparent motion. (a) Stimulus Pattern i can give rise to either the set of correspondence matches depicted in Pattern ii or the set depicted in Pattern iii. (b) An example of a third-order field effect. The correspondence matches assigned to Pattern i are identical to those assigned to Pattern ii when both are included in a single stimulus. (c) A violation of the third-order field effect. (When the stationary spots from Part b are removed, human subjects often report percepts involving coherent motion between patterns. )

Figure 14a. When a display is constructed from multiple instances of the [Figure 14a](#) display (see [Figure 14b](#)), each instance of the display gives rise to the same interpretation (i. e. , the same set of motion correspondence matches). It is not the case that one instance of the display is given one interpretation and another instance of the display is given a different interpretation (as in [Figure 14c](#)). The current model is not capable of generating this type of field effect because it requires third-order motion measurements (i. e. , it depends on minimizing differences between patterns of relative velocities).

Physiological evidence indicates that there exists cells, late in the motion pathway, that are sensitive to higher order properties of motion. In particular, cells that respond to changing size, to pattern rotation, and to the flow of motion from or to a central fixation point have been found in the MST area and in Area 7 (e. g. , [Motter & Mountcastle, 1981](#); [Saito et al. , 1986](#); [Sakata et al. , 1986](#)). Field effects could be mediated by patterns of excitatory and inhibitory connections between such cells. For example, the type of display depicted in [Figure 14a](#) is quite capable of eliciting response from a rotation-sensitive cell in the MST area (see [Saito et al. , 1986](#), [Figure 8](#)). An excitatory connection between one such cell detecting the motion of one of the patterns in 14b and

a similar cell detecting the motion of the other pattern could prevent the two patterns of motion from being assigned different interpretations.

Of importance, however, that the higher order field effects reported by [Ramachandran and Anstis \(1985\)](#) are not powerful determinants of motion correspondence matches; for this reason, such effects have not been incorporated into the current model. [Dawson and Nevin-Meadows \(1990\)](#) showed that field effects are easily mitigated by minor changes to the display in [Figure 14b](#). Consider the version of the field effect display in [Figure 14c](#). The only difference between it and [Figure 14b](#) is the absence of a stationary dot in the center of each pattern. However, this minor change in the stimulus often leads to the removal of the field effect; in many cases, subjects report seeing three dots arranged on the vertices of an imaginary triangle move back and forth while two other dots flash on and off. The correspondence matches consistent with this interpretation are also depicted in [Figure 14c](#); these matches are also generated by the correspondence network through the use of the standard settings. The presence of the stationary dots in [Figure 14b](#) is clearly a strong determinant of the higher order field effect, providing powerful cues that direct attention to specific parts of the display (i. e. , to two individuated patterns) and away from other areas (i. e. , from the area between the two patterns).

This is not to say that the [Figure 14c](#) display does not lead to higher order field effects, which the current model does not predict. By paying attention to different locations in the stimulus (e. g. , to where a dot would move back and forth in [Figure 14b](#) or to two dots that would move back and forth as the base of a triangle as in [Figure 14c](#)), one can intentionally switch between the field effect interpretation and the triangle interpretation. My colleagues and I are currently attempting to produce higher order field effects from the brainstate-in-a-box version of the current model ([Dawson, 1988](#)) by providing very strong initial biases for subsets of potential motion correspondence matches (i. e. , directing the model's attention to these matches).

## The element integrity principle.

Both the current model and [Ullman's \(1979\)](#) modified minimal mapping theory (see also [Grzywacz & Yuille, 1988](#)) apply an element integrity constraint to the assignment of motion correspondence matches. This constraint is used to prevent the splitting or the fusing of moving elements, and it causes the computer simulations to prefer one-to-one matches between frames of view. One method for implementing this constraint is to abandon the cover principle used in minimal mapping theory ([Ullman, 1979](#); see also Part 3). Recall that the cover principle prevents elements from being interpreted as suddenly appearing or disappearing. As a result, solutions like the two depicted in [Figures 2a](#) and [2b](#) are impossible when the cover principle is implemented: it forces both to be perceived as the split depicted in [Figure 2c](#). If a physiological substrate solves the motion correspondence by using rules that do not require the cover principle (i. e. , by applying rules as specified in the current model), then it should be able to respond to the sudden appearance or disappearance of elements instead of being able to respond only to element motion. [Motter and Mountcastle \(1981\)](#) observed some cells in Area 7 that respond in this fashion. In 357 light-sensitive neurons that they studied, 52% (185) produced a transient response after the sudden appearance of a stationary stimulus, 17% (51) produced sustained responses to its appearance, and a small number of cells (3%, or 12) generated transient responses at the stimulus's sudden appearance and disappearance. Thus many cells in Area 7—the likely site for the actual assignment of identity matches—can clearly signal that a new element has been added to the display.

Unfortunately, very little additional physiological evidence supports or disconfirms the implementation of the element integrity principle in any other manner, such as the inhibitory connections used in the current model. Again, this is because element splitting or fusing is not an important property of motion perception in general but is specifically related to element tracking, which has not been considered by physiologists to be an important function served by the motion pathway. As a result, stimuli specifically designed to test the element integrity principle (e. g. , motion competition displays, moving elements that actually break into two) have simply not been studied in physiological experiments. This again points to the need for using multiple element displays to increase the understanding of the physiological processes responsible for the assignment of correspondence matches.

## Alternative Structures for the Motion Correspondence Network

The preceding sections have shown that Area 7 in the parietal cortex is a plausible candidate for the physiological site of motion correspondence and tag-assignment processing. The types of motion measurements coded by areas in the motion pathway just before Area 7 are also consistent with the constraints exploited by the current model. Although it is clear that further experimentation is required to gain a more complete

understanding of the neural mechanisms underlying the assignment of motion correspondence matches, the evidence cited indicates that the major assumptions underlying the network described in Part 3 are biologically plausible.

Nevertheless, some attributes of the network are clearly not biologically plausible. This is because specific characteristics of the model were not designed to be consistent with neural circuitry but rather were designed to provide a convenient, effective procedure in which to test the utility of the three constraining principles. In this section, some of these implausible characteristics are briefly considered. Proposals for altering the network's structure to eliminate these characteristics, without changing the nature of the network's computations, are also described.

## Designing a fixed processing network.

A major problem with the biological plausibility of the current model is that it does not use a fixed architecture. Instead, a new network is constructed for each problem presented to the model. A physiological implementation of the model would require the design of a fixed network (i. e. , a set of processing units with fixed patterns of interconnectivity) capable of dealing with any presented problem.

A fixed motion correspondence network is possible in principle and could be constructed in such a way that it performs the same computations as the model described in Part 2. A fixed network would have a set of input units (i. e. , a two-dimensional retina) that code the location of elements. Temporal filtering would be used to differentiate Frame 1 elements from Frame 2 elements. As a result of this filtering, the processors coding the location of Frame 1 elements would have less activation because of the temporal decay. These input units would be connected to a layer of match processors that would represent possible motion correspondence matches. Each match processor would be connected to only two of the input units, one representing the starting position of the match in Frame 1 and the other representing the end of the match in Frame 2. A match processor would become active, and thus would become capable of assigning a correspondence match, only when both of its input units were activated. The physical connections between match processors would be as defined in the equations in Part 2, and would thus implement the three constraining properties. The weights of these connections can be defined in advance and thus integrated into the system because they depend only on the start and end locations of a particular motion correspondence match, which in turn are fixed by the connections to the processors in the input array.

A fixed network of this type may indeed be instantiated in the brain. [Albright et al. \(1984\)](#) demonstrated that the MT area exhibits a columnar organization of directionally selective cells. Each column of cells detects motion for a particular region of the visual field. Within each column exist cells sensitive to different directions of motion (and presumably different speeds as well). Because motion correspondence matches can be interpreted as being element velocities, the cells within each MT column may be serving a function similar to that of the match processors in the fixed network just described.

## Abandoning vector normalization.

A second biologically implausible property of the algorithm involves the normalization of the activity vector  $a$  after each iteration, producing the rotating brainstate-in-a-sphere. This is implausible for two reasons. First, the neural circuitry required to implement such a normalization is unlikely to be feasible because it would involve massively parallel connections between processors (see the next section). Second, physiological studies of neural systems that appear to directly implement brainstate "rotations" suggest that normalization is not performed. For instance, [Georgopoulos, Lurito, Petrides, Schwartz, and Massey \(1989\)](#) described how a population of neurons (i. e. , a "neuronal vector") in the motor cortex represents an upcoming movement in space. In a task requiring the mental rotation of a planned movement, it was discovered that activation levels in the neuronal vector changed in such a way that the vector itself could be interpreted as rotating. In addition, however, the length of the neuronal vector was changed. Its length was not normalized to a constant value.

Vector normalization is not required for the current model to assign motion correspondence matches. [Dawson \(1988\)](#) showed that it can be replaced with a brainstate-in-a-box restriction on processing (e. g. , [Anderson et al. , 1977](#)) in which individual processors are driven at convergence to either a maximal or a minimal activation level. This latter type of processing is clearly more biologically plausible because the maximal or minimal activation level of a processor can easily be interpreted in terms of the maximal or minimal spike frequency of a neuron.

However, replacing the brainstate-in-a-sphere model with a brainstate-in-a-box model may result in slight (but empirically significant) changes in what the network is actually computing. The current model was formulated as a brainstate-in-a-sphere because this resulted in the proof that the network would converge to a uniquely definable least-energy state. This same proof cannot be generated for the brainstate-in-a-box version of the model, as was noted in Part 2, because it cannot be assumed that the eigenvalues of the connection matrix are all positive. Thus in some cases, the two networks may converge to very different solutions, and for some displays, it is possible (though unlikely) that the brainstate-in-a-box fails to converge at all. In this case, the price of increasing biological plausibility may be a less certain computational understanding of what the model is doing.

## Eliminating massively parallel connections.

A third implausible characteristic of the current model is that it is massively parallel. In other words, every processor in the network is connected to every other processor. This type of interconnectivity is not characteristic of the human cortex, and thus the biological plausibility of connectionist models that assume massive parallelism has been questioned (e. g. , [Crick & Asanuma, 1986](#), p. 370).

The massive parallelism in the model speeds up processing considerably (see [Grzywacz & Yuille, 1988](#)) and also permits the modeling of the mutual influence of stimuli relatively far apart in the visual field on each other's processing. The large size of the receptive fields in the MT and MST areas and in Area 7 indicates that such distant influences need to be modeled. However, in principle these influences can be simulated without massive parallelism at the expense of slower processing. Two processors need not be directly connected to affect each other's performance (e. g. , [Dawson & Pylyshyn, 1986](#)). They can instead affect each other through indirect connections with intermediate processors.

One can easily eliminate massive parallelism from the network by placing some restrictions on which processors can be connected to one another. For instance, if two possible motion correspondence matches were beyond some critical distance apart in the visual field, then a connection between the processors representing them would not be established. This would not affect the formal properties of the network (i. e. , the convergence proof). One could "disconnect" two processors by creating a connection between them with zero weight. As a result, a symmetric connection matrix for the network could still be defined, and the conditions required for proving that the network converges would still remain.

## Limiting the number of processing units.

One final problem for the biological plausibility of the model concerns the number of processing units required to solve arbitrary correspondence problems. When there are  $N$  elements in Frame 1 and  $M$  elements in Frame 2, the network requires  $N \times M$  processing units to solve the correspondence problem for that display. Thus as input displays include larger numbers of elements, the size of the required network grows at an alarming rate. This suggests that at some point, for very large problems, the network will face [Ballard's \(1986\)](#) packing problem: It will require more processing units than there are neurons in the substrate presumed to assign motion correspondence matches.

Whether the network does indeed face the packing problem is largely an empirical question. At issue is whether motion correspondence processing is mediated by a limited capacity mechanism, as assumed by [Cavanagh and Mather \(1989\)](#). If it is not, then motion correspondence matches must be assigned to every element in a display (e. g. , every dot in a random-dot kinematogram), and the packing problem provides severe constraints on the model. However, if motion correspondence processing is indeed the mechanism for tag assignment, then it is quite likely a limited capacity process; very few elements can be processed at one time. For instance, [Pylyshyn and Storm \(1988\)](#) demonstrated that human observers have difficulty tracking more than five or six independently moving objects simultaneously. If only a small number of elements can be tracked at one time, the tracking mechanism could easily be modeled in a very small network.

## Conclusion: The Two-Process Distinction Revisited

A model that is capable of maintaining the identities of individuated elements as they move has been described. It solves a particular problem of underdetermination, the motion correspondence problem, by simultaneously applying three different constraints: the nearest neighbor principle, the relative velocity principle, and the element integrity principle. The model generates the same correspondence solutions as does the human visual system

for a variety of displays, and many of its properties are consistent with what is known about the physiological mechanisms underlying human motion perception. The model can also be viewed as a proposal of how the identities of attentional tags are maintained by visual cognition, and thus it can be differentiated from a system that serves merely to detect movement.

Several other researchers have hypothesized that long-range motion perception is involved in the tracking of visual elements. For instance, [Marr \(1982, pp. 202-204\)](#) suggested that there may be two correspondence problems for motion perception, one of which was maintaining the identity of elements that change appearance over time. [Petersik \(1989\)](#) also adopted this position in considering why the visual system might generate apparent motion at all. Unfortunately, this view appears to have had little impact on motion perception research.

The classic spatiotemporal differences between short- and long-range motion appear to have been incorrect (e. g. , [Cavanagh & Mather, 1989](#)). However, this does not mean that there is no merit in distinguishing between the low- and high-level processing of moving displays. [Anstis's \(1980\)](#) figure-based distinction still appears to be fundamentally sound. It is likely that low-level systems, perhaps best viewed as correlator models, exist to detect motion without requiring visual elements to be individuated. Nevertheless, although correlator models can detect motion and bypass the motion correspondence problem, they cannot function as element trackers (e. g. , [Adelson & Bergen, 1985](#); [Cavanagh & Mather, 1989](#)). Furthermore, the capacity to maintain the identities of elements appears to be a fundamental property of visual cognition ([Pylyshyn, 1989](#)). The implication is that high-level systems, like the token-matching scheme described in this article, exist to track elements after they have been individuated.

## Footnotes

1

The problem of underdetermination is not universally accepted. For instance, researchers who adopt the approach of direct perception argue that the visual system faces very different information-processing problems (e. g. , [Gibson, 1979](#)), and they assume that the information provided to perception is extremely rich. The differences between the views of direct and of indirect perception have been vigorously discussed in recent years (e. g. , [Cutting, 1986](#); [Fodor & Pylyshyn, 1981](#); [Turvey, Shaw, Reed, & Mace, 1981](#); [Ullman, 1980a](#)). For the purposes of this article, the existence of the problem of underdetermination is assumed.

2

Apparent motion can be viewed as the limiting case for real motion. Indeed, under the assumption that the temporal rates of frame presentation, apparent motion, and real motion are formally equivalent (e. g. , [Watson, Ahumada, & Farrell, 1986](#)), the physiological mechanisms that respond to real motion respond equally as well to apparent motion (e. g. , [Manning, Finlay, & Fenelon, 1988](#); [Newsome, Mikami, & Wurtz, 1986](#)). As [Newsome et al.](#) pointed out, the attraction of studying apparent motion is that the stimuli that produce this phenomenon contain only a few critical features and thus are likely to lead to an understanding of the properties that are necessary for motion perception.

3

This is not to say that motion perception mechanisms are completely insensitive to element appearances. Several psychophysical experiments have shown that correspondence matches can be affected by some aspects of element appearance (e. g. , [Chen, 1985](#); [Green, 1986](#); [Mack, Klein, Hill, & Palumbo, 1989](#); [Prazdny, 1986](#); [Ramachandran, Ginsburg, & Anstis, 1983](#); [Schechter, Hochstein, & Hillman, 1988](#); [Ullman, 1980b](#); [Watson, 1986](#)). However, the evidence suggests that such principles are much weaker than those based on measurements of element positions. First, conflicting results exist for many of these studies (compare [Chen, 1985](#), with [Dawson, 1990a](#); [Schechter et al. , 1988](#), with [Navon, 1976](#); and [Watson, 1986](#), with [Baro & Levinson, 1988](#)). Second, the quantitative effects of figural appearance on motion correspondence matching is quite weak. For example, [Schechter et al.](#) used changes in geometric shape to affect thresholds in a motion competition task by 3'-4' of visual angle. In contrast, [Dawson \(1987\)](#) demonstrated that manipulations of element positions could change thresholds in a related task by 30'-60'. The possible role of figural properties is considered in detail later in the article, and some proposals for potentially extending the model to be sensitive to such properties are considered at that point.

4

Technically speaking, the system as defined is not completely linear because the normalization defined in Equation 6 is applied in every iteration. In principle, however, this is not required. An alternative completely linear algorithm would not apply Equation 6 during the iterations. Instead, the network would be preset to process for a specified number of iterations (e. g. , 2,000 iterations). At the end of these iterations, the normalization defined in Equation 6 would be applied, followed by the evaluation of the thresholds. This completely linear algorithm would lead to exactly the same results as the one described (see Hall, 1963, pp. 63-66). However, it would be less efficient in the sense that it would process many simple displays for more iterations than would ordinarily be required.

5

A great deal of research in my laboratory has involved exploring variations of these degrees of freedom. For example, the performance of the model has been examined with a very wide variety of parameter settings in the equations given (for one example, see Dawson, 1988). Alternative definitions of constraints have also been considered. For instance, some of the simulations have defined relative velocity weights as the cosine of the angle between two motion vectors. Models defined by different parameters, or by different versions of the constraints, perform in different ways; some displays are problematic for one model but not for another. Nevertheless, each model is capable of correctly solving a wide range of displays, providing that some version of each constraint is applied.

## References:

1. Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2A, 284-299.
2. Aggarwal, J. K. & Duda, R. O. (1975). Computer analysis of moving polygonal images. *IEEE Transactions on Computers*, C-24, 966-976.
3. Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52, 1106-1130.
4. Albright, T. D., Desimone, R. & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology*, 51, 16-31.
5. Allman, J., Miezin, F. & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14, 105-126.
6. Anderson, J. A. & Mozer, M. C. (1981). Categorization and selective neurons. In G. E. Hinton & J. A. Anderson (Eds. ), *Parallel models of associative memory* (chap. 8). Hillsdale, NJ: Erlbaum.
7. Anderson, J. A., Silverstein, J. W., Ritz, S. A. & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 84, 413-451.
8. Anstis, S. M. (1978). Apparent movement. In R. Held, H. W. Leibowitz, & H. -L. Teuber (Eds. ), *Handbook of sensory physiology* (chap. 21). New York: Springer-Verlag.
9. Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London*, 290B, 153-168.
10. Anstis, S. M. (1986). Motion perception in the frontoparallel plane: Sensory aspects. In K. R. Boff, L. Kauffman, & J. P. Thomas (Eds. ), *Handbook of perception and human performance: Sensory processes and perception* (Vol. 1, chap. 16). New York: Wiley.
11. Attneave, F. (1974). Apparent movement and the what-where connection. *Psychologia*, 17, 108-120.
12. Attneave, F. & Block, G. (1973). Apparent movement in tridimensional space. *Perception & Psychophysics*, 13, 301-307.
13. Ballard, D. H. (1986). Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, 9, 67-120.
14. Barnard, S. T. & Thompson, W. B. (1980). Disparity analysis of images. *IEEE Transactions on Pattern*

15. Baro, J. A. & Levinson, E. (1988). Apparent motion can be perceived between patterns with dissimilar spatial frequencies. *Vision Research*, 28, 1311-1313.
16. Barr, M. L. (1974). *The human nervous system (2nd ed. )*. New York: Harper & Row.
17. Blake, A. & Zisserman, A. (1987). *Visual reconstruction*. Cambridge, MA: MIT Press.
18. Botez, M. I. (1975). Two visual systems in clinical neurology: Readaptive role of the primitive system in visual agnosis patients. *European Neurology*, 13, 101-122.
19. Braddick, O. J. (1974). A short range process in apparent movement. *Vision Research*, 14, 519-528.
20. Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London*, 290B, 137-151.
21. Braddick, O. J. & Adlard, A. (1978). Apparent motion and the motion detector. In J. Armington, J. Krauskopf, & B. R. Wooten (Eds. ), *Visual psychophysics and physiology*(chap. 33). New York: Academic Press.
22. Breitmeyer, B. G. & Ritter, A. (1986). Visual persistence and the effect of eccentric viewing, element size, and frame duration on bistable stroboscopic motion percepts. *Perception & Psychophysics*, 39, 275-280.
23. Bridgeman, B. (1972). Visual receptive fields sensitive to absolute and relative motion during tracking. *Science*, 178, 1106-1108.
24. Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London*, 227B, 249-265.
25. Burt, P. & Sperling, G. (1981). Time, distance and feature trade-offs in visual apparent motion. *Psychological Review*, 88, 137-151.
26. Caelli, T. & Dodwell, P. (1980). On the contours of apparent motion: A new perspective on visual space-time. *Biological Cybernetics*, 39, 27-35.
27. Cavanagh, P., Arguin, M. & von Grunau, M. (1989). Interattribute apparent motion. *Vision Research*, 29, 1197-1204.
28. Cavanagh, P. & Mather, G. (1989). Motion: The long and short of it. *Spatial Vision*, 4, 103-129.
29. Chen, L. (1985). Topological structure in the perception of apparent motion. *Perception*, 14, 197-208.
30. Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America*, 5A, 1986-2007.
31. Clark, A. (1989). *Microcognition*. Cambridge, MA: MIT Press.
32. Corbin, H. (1942). The perception of grouping and apparent movement in visual depth. *Archives of Psychology (Abstract No. 769, Whole No. 273)*
33. Crick, F. & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In J. McClelland, D. Rumelhart, & PDP Group (Eds. ), *Parallel distributed processing, Vol. 2. Psychological and biological models*(chap. 20), Cambridge, MA: MIT Press.
34. Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.
35. Cutting, J. E. & Proffitt, D. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology*, 14, 211-246.
36. Damasio, A. R. & Benton, A. L. (1979). Impairments of hand movements under visual guidance. *Neurology*, 29, 170-178.
37. Daugman, J. G. (1988). Pattern and motion vision without Laplacian zero crossings. *Journal of the Optical Society of America*, 5A, 1142-1148.

38. Dawson, M. R. W. (1986). *Using relative velocity as a natural constraint for the motion correspondence problem*. London, Canada: University of Western Ontario Centre for Cognitive Science Technical Memorandum No. 27.
39. Dawson, M. R. W. (1987). Moving contexts do affect the perceived direction of apparent motion in motion competition displays. *Vision Research*, 27, 799-809.
40. Dawson, M. R. W. (1988). The cooperative application of multiple natural constraints to the motion correspondence problem. In R. Goebel (Ed. ), *Proceedings of the Seventh Canadian Conference On Artificial Intelligence*. Edmonton, Alberta, Canada: University of Alberta Printing Services.
41. Dawson, M. R. W. (1989). Constraining tag-assignment from above and below. *Behavioral and Brain Sciences*, 12, 400-402.
42. Dawson, M. R. W. (1990a). Apparent motion and element connectedness. *Spatial Vision*, 4, 241-251.
43. Dawson, M. R. W. (1990b). Empirical issues in theoretical psychology: Comment on Kukla. *American Psychologist*, 45, 778-780.
44. Dawson, M. R. W. & Di Lollo, V. (1990). Effects of adapting luminance and stimulus contrast on the temporal and spatial limits of short-range motion. *Vision Research*, 30, 415-429.
45. Dawson, M. R. W. & Harder, B. (1989, September). *Testing the necessity of constraints on motion correspondence*. Poster presented at the 4th annual Joseph R. Royce Research Conference, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.
46. Dawson, M. R. W. & Nevin-Meadows, N. (1990). [The effect of individuating locations on "field effects" in apparent motion]. Unpublished raw data.
47. Dawson, M. R. W. & Pylyshyn, Z. W. (1986). Using relative velocity information to constrain the motion correspondence problem: Psychophysical data and a computer model. In *Proceedings of the Sixth Canadian Conference on Artificial Intelligence* (pp. 117-123). Montréal, Canada: École Polytechnique of Montréal.
48. Dawson, M. R. W. & Pylyshyn, Z. W. (1988). Natural constraints on apparent motion. In Z. W. Pylyshyn (Ed. ), *Computational processes in human vision* (chap. 5). Norwood, NJ: Ablex.
49. Dawson, M. R. W. & Wright, R. D. (1989). The consistency of element motion affects the visibility but not the direction of apparent movement. *Spatial Vision*, 4, 17-29.
50. Desimone, R. & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Comparative Neurology*, 248, 164-189.
51. DeYoe, E. A. & van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience*, 11, 219-226.
52. Di Lollo, V. & Hogben, J. H. (1987). Suppression of visible persistence as a function of spatial separation between inducing stimuli. *Perception & Psychophysics*, 41, 345-354.
53. Dubner, R. & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, 35, 528-532.
54. Farrell, J. E. & Kessler, E. J. (1988). Visible persistence is the result of spatiotemporal filtering. *Perception & Psychophysics*, 43, 304-306.
55. Farrell, J. E. & Shepard, R. N. (1981). Shape, orientation, and apparent rotational movement. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 477-486.
56. Ferrie, F. P., Levine, M. D. & Zucker, S. W. (1982). Cell tracking: A modeling and minimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4, 277-291.
57. Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
58. Fodor, J. A. & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's

"Ecological Approach. " *Cognition*, 9, 139-196.

59. Foster, D. H. (1978). Visual apparent motion and the calculus of variations. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception* (chap. 3). New York: Wiley.
60. Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B. & Massey, J. T. (1989). Mental rotation of the neuronal population vector. *Science*, 243, 234-236.
61. Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
62. Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical optimization*. New York: Academic Press.
63. Gogel, W. (1974). The adjacency principle in visual perception. *Quarterly Journal of Experimental Psychology*, 26, 425-437.
64. Goldberg, M. E. & Bruce, C. J. (1985). Cerebral cortical activity associated with the orientation of visual attention in the rhesus monkey. *Vision Research*, 25, 471-481.
65. Goodman, N. (1978). *Ways of worldmaking*. Indianapolis, IN: Hackett.
66. Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research*, 26, 599-607.
67. Green, M. & Odom, J. V. (1986). Correspondence matching in apparent motion: Evidence for three-dimensional spatial representation. *Science*, 233, 1427-1429.
68. Gregory, R. L. (1970). *The intelligent eye*. London: Weidenfeld & Nicolson.
69. Grimson, E. (1981). *From images to surfaces*. Cambridge, MA: MIT Press.
70. Grzywacz, N. M. & Yuille, A. L. (1988). Massively parallel implementations of theories for apparent motion. *Spatial Vision*, 3, 15-44.
71. Hall, G. G. (1963). *Matrices and tensors*. New York: Pergamon Press.
72. Hammond, P., Ahmed, B. & Smith, A. T. (1986). Relative motion sensitivity in cat cortex as a function of stimulus direction. *Brain Research*, 386, 93-104.
73. Hammond, P., Pomfrett, C. J. D. & Ahmed, B. (1989). Neural motion aftereffects in the cat's striate cortex: Orientation selectivity. *Vision Research*, 29, 1671-1683.
74. Hammond, P. & Smith, A. T. (1984). Sensitivity of complex cells in cat striate cortex to relative motion. *Brain Research*, 301, 287-298.
75. Hess, R. H., Baker, C. L. & Zihl, J. (1989). The "motion-blind" patient: Low-level spatial and temporal filters. *Journal of Neuroscience*, 9, 1628-1640.
76. Hildreth, E. C. (1983). *The measurement of visual motion*. Cambridge, MA: MIT Press.
77. Hinton, G. E. & Sejnowski, T. J. (1986). Learning and relearning in Boltzman machines. In D. E. Rumelhart, J. L. McClelland, & PDP Group (Eds.), *Parallel distributed processing: Vol. 1. Foundations* (chap. 7). Cambridge, MA: MIT Press.
78. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, U. S. A*, 79, 2554-2558.
79. Horn, B. K. P. (1986). *Robot vision*. Cambridge, MA: MIT Press.
80. Horn, B. K. P. & Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185-203.
81. Hurlbert, A. & Poggio, T. (1985). Spotlight on attention. *Trends in Neurosciences*, 8, 309-311.
82. Hyvärinen, J. & Poranen, A. (1974). Function of parietal associative area 7 as revealed from cellular discharges in alert monkeys. *Brain*, 97, 673-692.

83. Jain, R., Martin, W. N. & Aggarwal, J. K. (1979a). Extraction of moving object images through change detection. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (pp. 425-428). Los Altos, CA: Morgan Kaufmann.
84. Jain, R., Martin, W. N. & Aggarwal, J. K. (1979b). Segmentation through the detection of changes due to motion. *Computer Graphics and Image Processing*, 11, 13-34.
85. Jain, R., Militzer, D. & Nagel, H. (1977). Separating non-stationary from stationary scene components in a sequence of real-world TV-images. *Proceedings of the Fifth Joint International Conference on Artificial Intelligence* (pp. 612-618). Los Altos, CA: Morgan Kaufmann.
86. Johansson, G. (1950). *Configurations in event perception*. Uppsala, Sweden: Almqvist & Wiksell.
87. Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
88. Jones, E. & Bruner, J. (1954). Expectancy in apparent visual movement. *British Journal of Psychology*, 45, 157-165.
89. Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & PDP Group (Eds.), *Parallel distributed processing: Vol. 1. Foundations* (chap. 9). Cambridge, MA: MIT Press.
90. Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290, 91-97.
91. Kaji, S. & Kawabata, N. (1985). Neural interactions of two moving patterns in the direction and orientation domain in the complex cells of cat's visual cortex. *Vision Research*, 25, 749-753.
92. Köhler, W. (1975). *Gestalt psychology*. New York: New American Library. (Original work published 1947)
93. Kolers, P. (1972). *Aspects of motion perception*. New York: Pergamon Press.
94. Kolers, P. (1983). Some features of visual form. *Computer Vision, Graphics, and Image Processing*, 23, 15-41.
95. Kolers, P. & Green, M. (1984). Color logic of apparent motion. *Perception*, 13, 149-154.
96. Kolers, P. & Pomerantz, J. R. (1971). Figural change in apparent motion. *Journal of Experimental Psychology*, 87, 99-108.
97. Kolers, P. & von Grunau, M. (1976). Shape and colour in apparent motion. *Vision Research*, 16, 329-335.
98. Krumhansl, C. L. (1984). Independent processing of visual form and motion. *Perception*, 13, 535-546.
99. Kuffler, S. W., Nicholls, J. G. & Martin, A. R. (1984). *From neuron to brain (2nd ed.)*. Sunderland, MA: Sinauer.
100. Lelkins, A. & Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research*, 24, 1083-1090.
101. Lisberger, S., Morris, E. J. & Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience*, 10, 97-129.
102. Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement and depth: Anatomy, physiology, and perception. *Science*, 240, 740-750.
103. Lynch, J. C., Mountcastle, V. B., Talbot, W. H. & Yin, T. C. T. (1977). Parietal lobe mechanisms for directed visual attention. *Journal of Neurophysiology*, 40, 362-389.
104. Mack, A., Klein, L., Hill, J. & Palumbo, D. (1989). Apparent motion: Evidence of the influence of shape, slant and size on the correspondence process. *Perception & Psychophysics*, 46, 201-206.
105. Mandl, G. (1985). Responses of visual cells in cat superior coliculus to relative pattern movement. *Vision Research*, 25, 267-281.

106. Manning, M. L., Finlay, D. C. & Fenelon, B. (1988). Visual evoked potentials to stimuli in apparent motion. *Vision Research*, 28, 965-974.
107. Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London*, 275, 483-524.
108. Marr, D. (1982). *Vision*. San Francisco: Freeman.
109. Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194, 283-287.
110. Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London*, B211, 151-180.
111. Mather, G. (1984). Luminance change generates apparent movement: Implications for models of directional specificity in the human visual system. *Vision Research*, 24, 1399-1405.
112. Mather, G. (1988). Temporal properties of apparent motion in subjective figures. *Perception*, 17, 729-736.
113. Maunsell, J. H. R. & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10, 363-401.
114. Maunsell, J. H. R. & van Essen, D. C. (1983a). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3, 2563-2586.
115. Maunsell, J. H. R. & van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey: I. Selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology*, 49, 1127-1147.
116. Maunsell, J. H. R. & van Essen, D. C. (1983c). Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49, 1148-1167.
117. Mikami, A., Newsome, W. T. & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex: II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, 55, 1328-1339.
118. Morgan, M. J. & Watt, R. J. (1983). On the failure of spatiotemporal interpolation: A filtering model. *Vision Research*, 23, 997-1004.
119. Mori, T. (1982). Apparent motion path composed of a serial concatenation of translations and rotations. *Biological Cybernetics*, 44, 31-34.
120. Motter, B. C. & Mountcastle, V. B. (1981). The functional properties of the light-sensitive neurons of the posterior parietal cortex studied in waking monkeys: Foveal sparing and opponent vector organization. *Journal of Neuroscience*, 1, 3-26.
121. Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1984). The analysis of moving visual patterns. In C. Chagas, R. Tattass, & C. G. Gross (Eds. ), *Study group on pattern recognition mechanisms*(pp. 117-151). Vatican City: Pontifica Academia Scientiarium.
122. Mutch, K., Smith, I. & Yonas, A. (1983). The effect of two-dimensional and three-dimensional distance on apparent motion. *Perception*, 12, 305-312.
123. Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research*, 25, 625-660.
124. Navon, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 130-138.
125. Newsome, W. T., Mikami, A. & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex: III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology*, 55, 1340-1351.
126. Newsome, W. T. & Pare, E. B. (1988). A selective impairment of motion processing following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8, 2201-2211.
127. Newsome, W. T., Wurtz, R. H., Dursteler, M. R. & Mikami, A. (1985). Deficits in visual motion processing

following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience*, 5, 825-840.

128. Orban, G. A., Gulyas, B. & Spileers, W. (1988). Influence of moving textured backgrounds on responses of cat area 18 cells to moving bars. *Progress in Brain Research*, 75, 137-145.

129. Orban, G. A., Gulyas, B. & Vogels, R. (1987). Influence of a moving textured background on directional selectivity of cat striate neurons. *Journal of Neurophysiology*, 57, 1792-1812.

130. Pantle, A. & Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science*, 193, 500-502.

131. Perrott, D. R. (1982). Studies in the perception of auditory motion. In R. W. Gatehouse (Ed. ), *Localization of sound: Theory and applications*. Groton, CT: Amphora Press.

132. Petersik, J. T. (1986). Group movement produced by the short-range process. *Perception & Psychophysics*, 39, 445-446.

133. Petersik, J. T. (1987). Dependence of apparent movement of a subjective figure on the perceptual fate of inducing elements. *Perception*, 16, 453-459.

134. Petersik, J. T. (1989). The two-process distinction in apparent motion. *Psychological Bulletin*, 106, 107-127.

135. Prazdny, K. (1986). What variables control (long-range) apparent motion? *Perception*, 15, 37-40.

136. Price, K. E. (1985). Relaxation matching techniques—A comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7, 617-623.

137. Price, K. E. & Reddy, R. (1977). Change detection and analysis in multi-spectral images. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (pp. 619-625)*. Los Altos, CA: Morgan Kaufmann.

138. Proffitt, D. & Cutting, J. E. (1979). Perceiving the centroid of configurations on a rolling wheel. *Perception & Psychophysics*, 25, 389-398.

139. Proffitt, D. & Cutting, J. E. (1980). An invariant for wheel-generated motions and the logic of its determination. *Perception*, 9, 435-449.

140. Pylyshyn, Z. W. (1980). Cognition and computation: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 3, 111-132.

141. Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.

142. Pylyshyn, Z. W. (1988). Here and there in the visual field. In Z. W. Pylyshyn (Ed. ), *Computational processes in human vision* (chap. 9). Norwood, NJ: Ablex.

143. Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32, 65-97.

144. Pylyshyn, Z. W. & Storm, R. (1988). Tracking of multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 1-19.

145. Ramachandran, V. S. (1985). Apparent motion of subjective surfaces. *Perception*, 14, 127-134.

146. Ramachandran, V. S. & Anstis, S. M. (1985). Perceptual organization in multistable apparent motion. *Perception*, 14, 135-143.

147. Ramachandran, V. S. & Anstis, S. M. (1986a). Figure-ground segregation modulates apparent motion. *Vision Research*, 26, 1969-1975.

148. Ramachandran, V. S. & Anstis, S. M. (1986b). The perception of apparent motion. *Scientific American*, 254 (6) , 102-109.

149. Ramachandran, V. S. & Cavanagh, P. (1987). Motion capture anisotropy. *Vision Research*, 27, 97-106.

150. Ramachandran, V. S., Ginsburg, A. P. & Anstis, S. (1983). Low spatial frequencies dominate apparent motion. *Perception*, 12, 457-461.
151. Ramachandran, V. S., Inada, V. & Kiama, G. (1986). Perception of illusory occlusion in apparent motion. *Vision Research*, 26, 1741-1749.
152. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research*, 13, 1399-1401.
153. Ratcliff, G. & Davies-Jones, G. A. B. (1972). Defective visual localization in focal brain wounds. *Brain*, 95, 49-60.
154. Regan, D. (1986). Visual processing of four kinds of relative motion. *Vision Research*, 26, 127-145.
155. Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information. In W. A. Rosenblith (Ed. ), *Sensory communication*(pp. 303-317). Cambridge, MA: MIT Press.
156. Restle, F. (1979). Coding theory of motion configurations. *Psychological Review*, 86, 1-24.
157. Richards, W. (1988). The approach. In W. Richards (Ed. ), *Natural computation*(chap. 1). Cambridge, MA: MIT Press.
158. Robinson, D. L., Goldberg, M. E. & Stanton, G. B. (1978). Parietal association cortex in the primate: Sensory mechanisms and behavioral modulations. *Journal of Neurophysiology*, 41, 910-932.
159. Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
160. Rodman, H. R. & Albright, T. D. (1987). Coding of visual stimulus velocity in area MT of the macaque. *Vision Research*, 27, 2035-2048.
161. Rumelhart, D. E., Hinton, G. & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & PDP Group (Eds. ), *Parallel distributed processing: Vol. 1. Foundations*(chap. 2). Cambridge, MA: MIT Press.
162. Rumelhart, D. E. & McClelland, J. L. (1985). Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General*, 114, 193-197.
163. Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y. & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6, 145-157.
164. Sakata, H., Shibutani, H., Kawano, K. & Harrington, T. L. (1985). Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Research*, 25, 453-463.
165. Sakata, H., Shibutani, H. & Tsurugai, K. (1986). Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Experimental Brain Research*, 61, 658-663.
166. Sekuler, R., Anstis, S., Braddick, O. J., Brandt, T., Movshon, J. A. & Orban, G. (1990). The perception of motion. In L. Spillman & J. S. Werner (Eds. ), *Visual perception: The neurophysiological foundations*(chap. 9). San Diego, CA: Academic Press.
167. Shechter, S., Hochstein, S. & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, 28, 1013-1021.
168. Shepard, R. N. (1978). The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed. ), *Theory construction and data analysis in the behavioral sciences*(chap. 2). San Francisco: Jossey-Bass.
169. Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds. ), *Perceptual organization*(chap. 10). Hillsdale, NJ: Erlbaum.
170. Shepard, R. N. (1982). Perceptual and analogical bases of cognition. In J. Mehler, E. Walker, & M. Garrett (Eds. ), *Perspectives on mental representation*(pp. 49-67). Hillsdale, NJ: Erlbaum.
171. Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving,

imaging, thinking, and dreaming. *Psychological Review*, 91, 417-447.

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

173. Sherrick, C. E. & Rogers, R. (1966). Apparent haptic movement. *Perception & Psychophysics*, 1, 175-180.

174. Sigman, E. & Rock, I. (1974). Stroboscopic movement based on perceptual intelligence. *Perception*, 3, 9-28.

175. Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.

176. Strong, G. W. & Whitehead, B. A. (1989). A solution to the tag-assignment problem for neural networks. *Behavioral and Brain Sciences*, 12, 381-433.

177. Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, 6, 134-144.

178. Tarr, M. J. & Pinker, S. (1985, November). *Nearest neighbors in apparent motion: Two or three dimensions?* Paper presented at the annual meeting of the Psychonomic Society, Boston.

179. Ternus, J. (1938). The problem of phenomenal identity. In W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 149-160). New York: Humanities Press.

180. Tsuji, S., Osada, M. & Yachida, M. (1979). Three-dimensional movement analysis of dynamic line images. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (pp. 896-901). Los Altos, CA: Morgan Kaufmann.

181. Tsuji, S., Osada, M. & Yachida, M. (1980). Tracking and segmentation of moving objects in dynamic line images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2, 516-522.

182. Turvey, M. T., Shaw, R. E., Reed, E. S. & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, 9, 237-304.

183. Ullman, S. (1978). Two-dimensionality of the correspondence process in apparent motion. *Perception*, 7, 683-693.

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

185. Ullman, S. (1980a). Against direct perception. *Behavioral and Brain Sciences*, 3, 373-381.

186. Ullman, S. (1980b). The effect of similarity between line segments on the correspondence strength in apparent motion. *Perception*, 9, 617-626.

187. Ullman, S. (1981, August). Analysis of visual motion by biological and computer systems. *IEEE Computer*, 14, pp. 57-69.

188. Ullman, S. (1984). Visual routines. *Cognition*, 18, 97-159.

189. Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (chap. 18). Cambridge, MA: MIT Press.

190. van Santen, J. P. H. & Sperling, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America*, 1A, 451-473.

191. van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America*, 2A, 300-320.

192. Victor, J. D. & Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research*, 30, 289-301.

193. Watson, A. B. (1986). Apparent motion occurs only between similar spatial frequencies. *Vision Research*, 26, 1727-1730.

194. Watson, A. B. & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America*, 2A, 322-341.
195. Watson, A. B., Ahumada, A. J. & Farrell, J. (1986). The window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America*, 3A, 300-307.
196. Yalamanchili, S., Martin, W. N. & Aggarwal, J. K. (1982). Extraction of moving object descriptions via differencing. *Computer Graphics and Image Processing*, 18, 188-201.
197. Yuille, A. L. (1983). *The smoothest velocity field and token matching schemes (Massachusetts Institute of Technology AI Memo No. 724)*. Cambridge, MA: MIT.
198. Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology*, 236, 549-573.
199. Zihl, J., von Cramon, D. & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, 106, 313-340.
200. Zucker, S. (1976). Relaxation labeling and the reduction of local ambiguities. In C. H. Chen (Ed.), *Pattern recognition and artificial intelligence* (pp. 852-861). New York: Academic Press.

This appendix shows that (except in a small number of special cases) the vector  $\mathbf{a}$ , representing the activation values of processing units, converges to  $\pm \mathbf{e}_1$ , whereby  $\mathbf{e}_1$  is the most dominant eigenvector of the connection matrix  $\mathbf{W}$ .

[Equation A1](#) defines the change in network activation values from iteration  $k - 1$  to iteration  $k$ : In order to determine the state to which the network converges, it is necessary to determine how activation state  $\mathbf{a}_k$  is related to the initial state  $\mathbf{a}_0$ . Consider the following network transitions, which are derived from [Equation A1](#): It is clear that the relation between the activation state on iteration  $k$  and the initial activation state is where  $\mathbf{W}^k$  is the connection matrix  $\mathbf{W}$  raised to the power of  $k$ . Recall from the main text that  $\mathbf{W}$  is a symmetric square matrix with  $N \times M$  rows and columns. The symmetry of  $\mathbf{W}$  ensures that it has  $N \times M$  linearly independent eigenvectors, each associated with a real eigenvalue. As a result, [Equation A2](#) defines what has been called the *power method* for extracting the most dominant eigenvector of a matrix (e. g. , [Hall, 1963](#), pp. 63-66). Specifically, the eigenvectors of  $\mathbf{W}$  span the  $N \times M$  dimensional space in which vector  $\mathbf{a}_0$  is defined. Let an eigenvector of  $\mathbf{W}$  be represented as  $\mathbf{e}_i$ , and let this eigenvector be associated with the eigenvalue  $\lambda_i$ . [Equation A2](#) can be rewritten by representing  $\mathbf{a}_0$  as a linear combination (with weights equal to  $a_i$ ) of  $\mathbf{W}$ 's eigenstructure: As  $k$  becomes sufficiently large, the summation term in the part of [Equation A3](#) vanishes, provided that  $|\lambda_1| > |\lambda_2|$ ; if  $\mathbf{a}_k$  is being normalized, this will cause the network to converge because, without the summation term, [Equation A3](#) simplifies to The effect of normalization is to divide [Equation A4](#) by the value  $|a_1 \cdot (\lambda_1)^k|$ . When  $a_1$  is positive, this results in  $\mathbf{a}_k = \mathbf{e}_1$ . When  $a_1$  is negative, this results in  $\mathbf{a}_k = -\mathbf{e}_1$

. Thus, except for the two rare special cases to be discussed, the iterative processing of the motion correspondence network results in its convergence to the most dominant eigenvector of  $\mathbf{W}$ , multiplied by  $\pm 1$ .

One special case is when  $a_1 = 0$ , which occurs when  $\mathbf{a}_0$  is either the zero vector or one of the other eigenvectors of  $\mathbf{W}$ , which are orthogonal to  $\mathbf{e}_1$ . In these cases the network converges immediately to the zero vector. However, this occurrence is very rare because it can be caused by only  $N \times M$  of the infinite number of possible  $\mathbf{a}_0$  vectors. The second special case occurs when  $\mathbf{W}$  does not have a single dominant eigenvector (i. e. , when  $|\lambda_1| = |\lambda_2|$

). In this instance, the network converges to a stable state, but it is a linear combination of  $\mathbf{e}_1$  and  $\mathbf{e}_2$ . As a result, the stable state is not unique; the network converges to many solutions for the same motion correspondence problem. In practice, however, this special case is likely to be a very rare occurrence and has not yet been encountered in the displays used to study the model.

This appendix modifies a measure of network energy proposed by [Hopfield \(1982\)](#) and shows that the most dominant eigenvector of  $\mathbf{W}$  represents a network state that minimizes this measure.

[Hopfield \(1982\)](#) proposed the following energy metric for an autoassociative network with connection weights represented in the symmetric matrix  $\mathbf{W}$  and unit activation values represented in the column vector  $\mathbf{a}$  whose  $j$ th entry is  $a_j$ : The processing in a Hopfield net minimizes this measure. In addition, in a Hopfield net,  $w_{ii} = 0$ .

Therefore Equation B1 can be rewritten without any restrictions on the double summation: This allows the energy equation to be expressed in linear algebra as where  $\mathbf{a}^T$  is the transposition of  $\mathbf{a}$ . In the correspondence network, the length of the vector  $\mathbf{a}$  is always normalized to unit length, and as a result the dot product  $\mathbf{a}^T \cdot \mathbf{a}$

must equal 1. Therefore, for the current model, the energy equation in Equation B3 can be rewritten as Equation B4 represents an expression of network energy for the motion correspondence model. It is very similar, but not identical, to Hopfield's (1982) metric defined in Equation B1. When applied to a standard Hopfield net, both equations produce the same result because in such a network the diagonal components of  $\mathbf{W}$  are equal to zero. However, when applied to the motion correspondence network, the two energy metrics produce different results because in this model  $\mathbf{W}$  does not have a zero diagonal.

The iterative processing of the correspondence model should proceed until network energy is minimized. Under what conditions does this occur?  $E$  is minimized by the activity pattern  $\mathbf{a}$  that produces the maximal value for the term: which is called the *Rayleigh ratio* (e. g. , Hall, 1963, p. 67). The maximal value for the Rayleigh ratio is  $\lambda_{\max}$ , the largest eigenvalue of  $\mathbf{W}$  (e. g. , Gill, Murray, & Wright, 1981, p. 25). This maximal value is obtained when  $\mathbf{a} = \pm \mathbf{e}_1$ , where  $\mathbf{e}_1$  is the eigenvector of  $\mathbf{W}$  associated with  $\lambda_{\max}$ . Thus when the autoassociative network converges to the activity pattern  $\pm \mathbf{e}_1$  (as proved in Appendix A), it is converging to a state that minimizes the energy metric defined in Equation B4.

Received: November 20, 1990. Revised: March 4, 1991. Accepted: May 5, 1991.

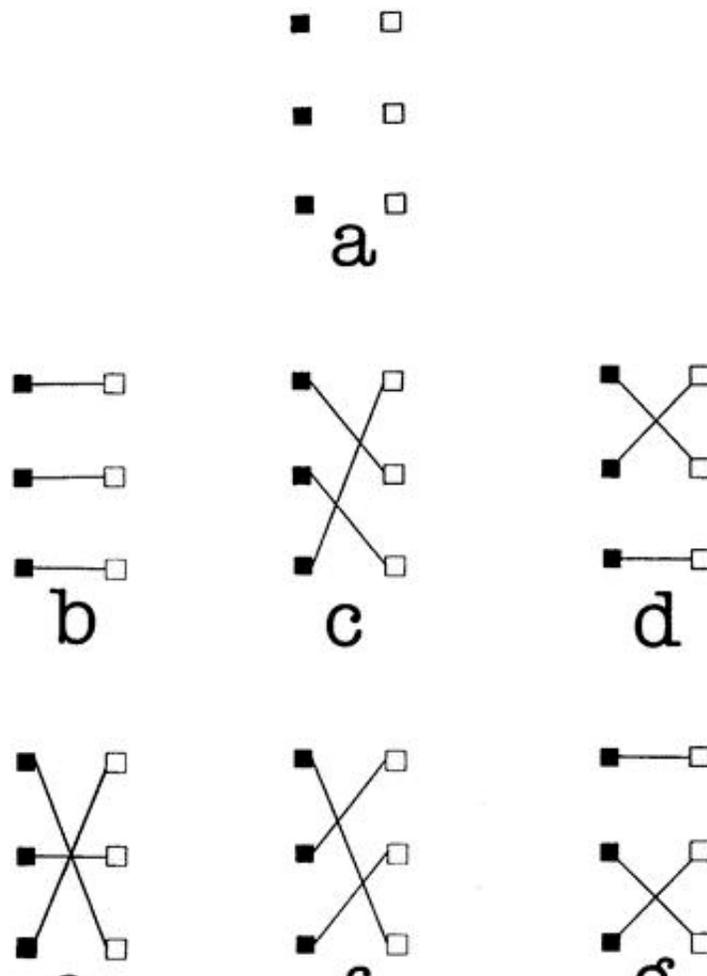
Copyright 1991 by the American Psychological Association, Inc.

For personal use only--not for distribution

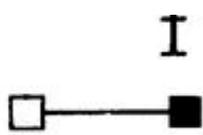
This publication is protected by US and international copyright laws and its content may not be copied without the copyright holder's express written permission except for the print or download capabilities of the retrieval software used for access. This content is intended solely for the use of the individual user.

Source: Psychological Review. Vol. 98 (4) October 1991, pp. 569-603

Accession Number: rev984569 Digital Object Identifier: 10.1037//0033-295X.98.4.569



e



a

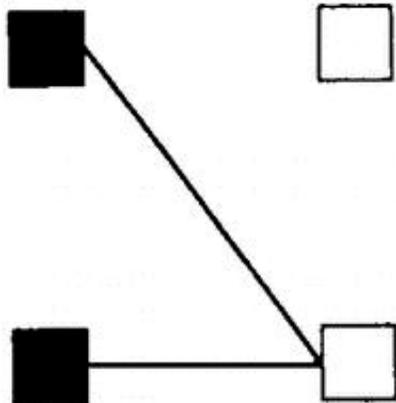
$\sigma$



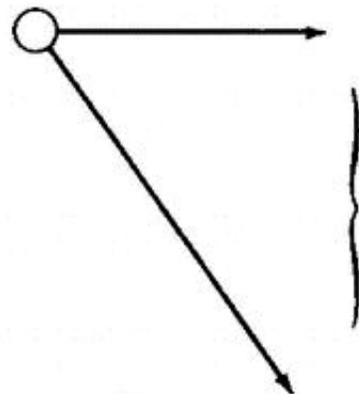
b



c



a



b



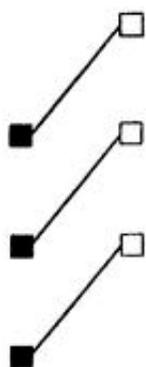
a



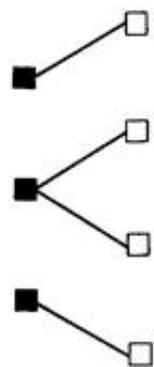
**b**



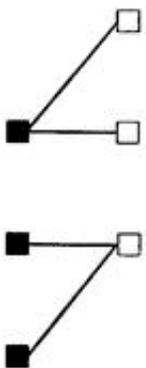
**c**



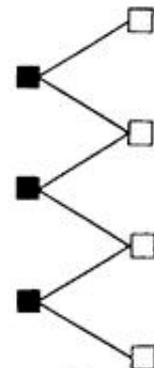
**a**



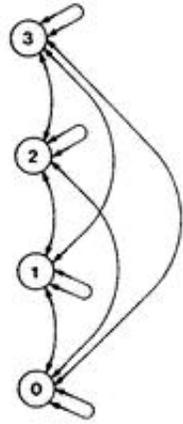
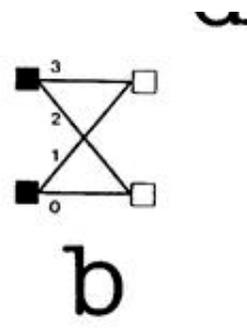
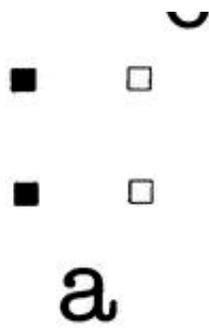
**b**



**c**



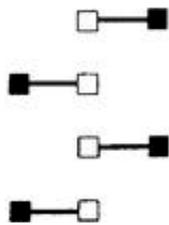
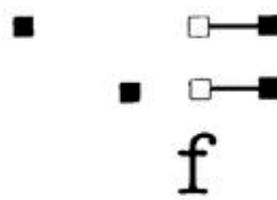
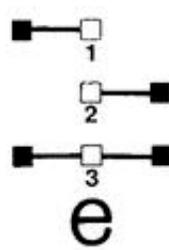
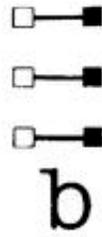
**d**



$$\begin{pmatrix} .5 \\ .5 \\ .5 \\ .5 \end{pmatrix} \quad \begin{pmatrix} .03 & -.10 & -.12 & .05 \\ -.10 & .02 & -.04 & -.12 \\ -.12 & -.04 & .02 & -.10 \\ .05 & -.12 & -.10 & .03 \end{pmatrix}$$

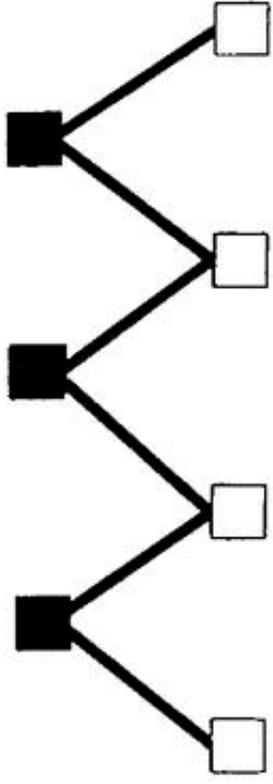
**c**

**d**

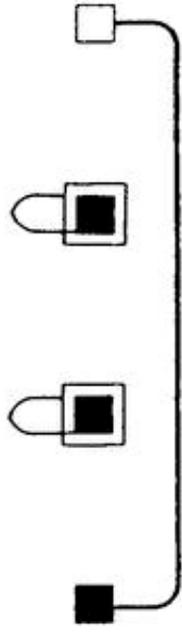




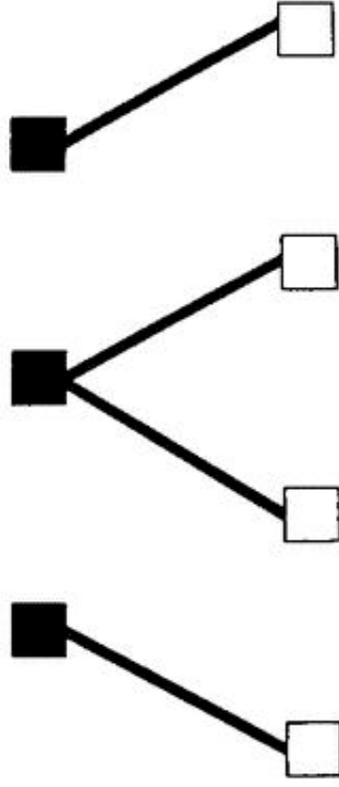
a



$\sigma_3$



b



$h$



a

b

c

$$nn_{ii} = \exp(-\alpha \|m_i\|).$$

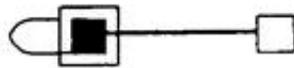
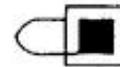
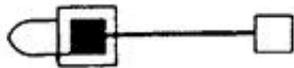
$$rv_{ij} = \Psi_{ij} [2 \exp(-\beta \|m_i - m_j\|) - 1].$$

$$\Psi_{ij} = \exp(-\epsilon \|x - y\|).$$

$$C = \delta(\omega_1 \mathbf{NN} + \omega_2 \mathbf{RV} + \omega_3 \mathbf{EI}),$$

$$\mathbf{a}^{k+1} = \mathbf{a}^k + \mathbf{C}\mathbf{a}^k = (\mathbf{I} + \mathbf{C})\mathbf{a}^k = \mathbf{W}\mathbf{a}^k.$$

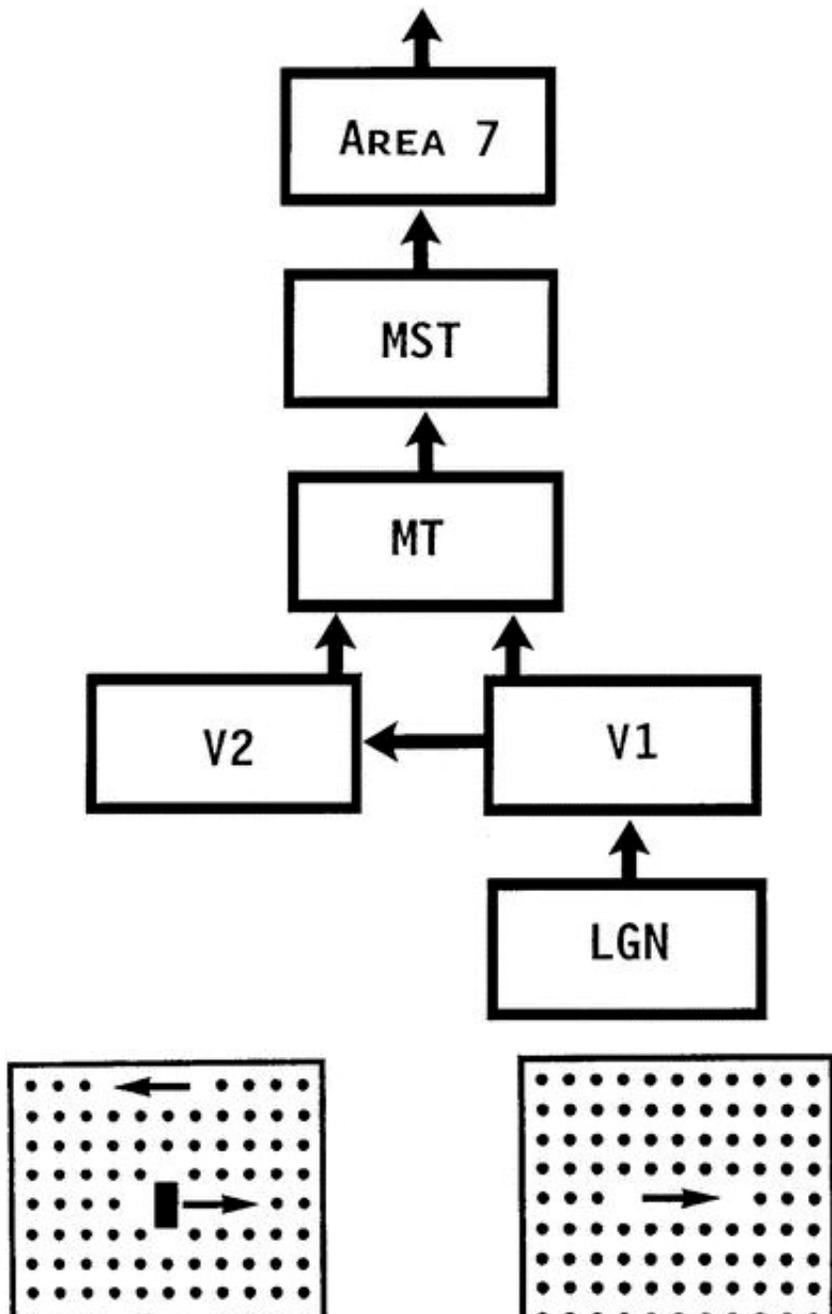
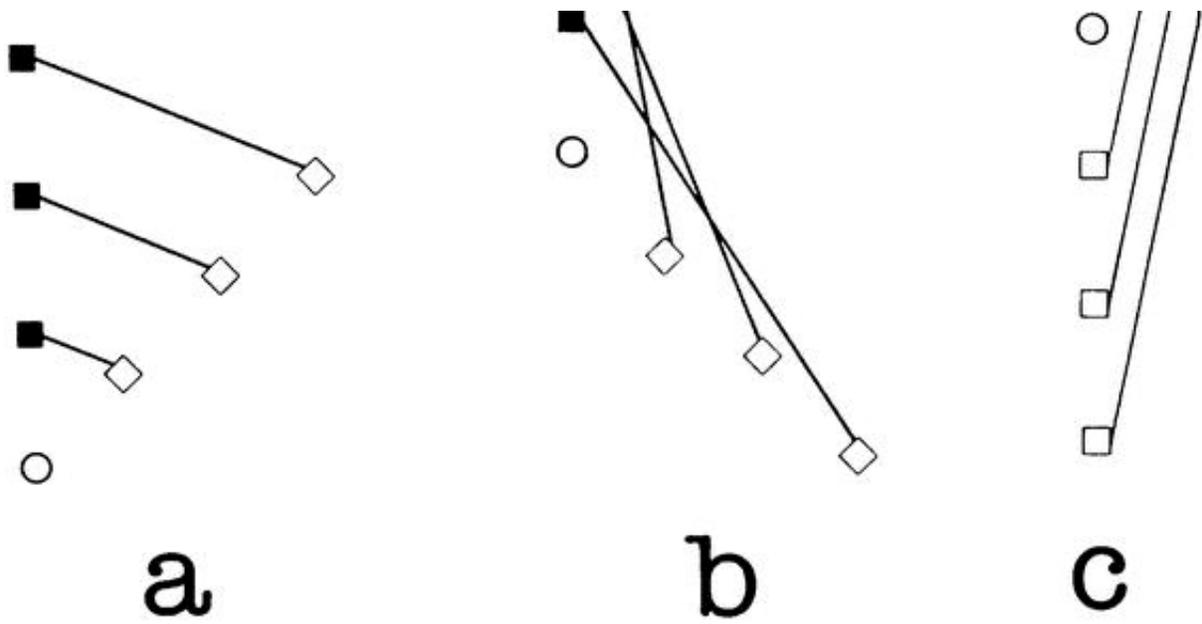
$$\mathbf{a}^{k+1} = \frac{1}{\|\mathbf{a}^{k+1}\|} \mathbf{a}^{k+1}.$$



a

b



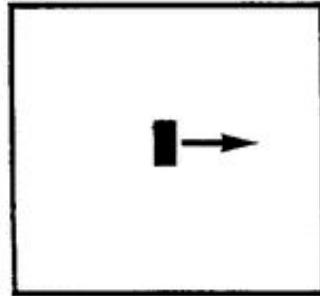




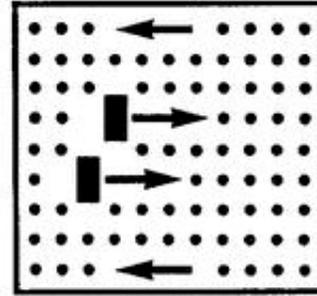
a



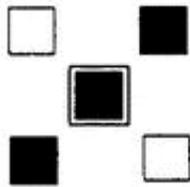
b



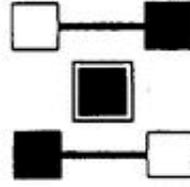
c



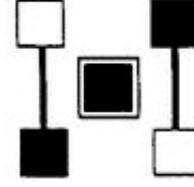
d



i

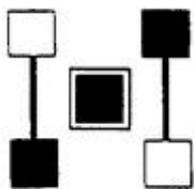


ii

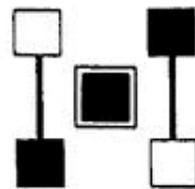


iii

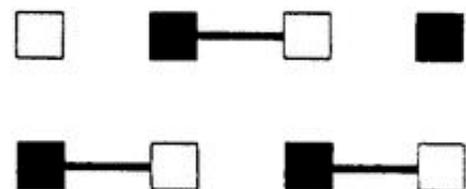
a



i



ii



c

b

$$\mathbf{a}^k = \mathbf{W} \cdot \mathbf{a}^{k-1}$$

$$\mathbf{a}^1 = \mathbf{W} \cdot \mathbf{a}^0.$$

$$\mathbf{a} = \mathbf{W} \cdot \mathbf{a},$$

$$\mathbf{a}^2 = \mathbf{W} \cdot \mathbf{a}^1 = \mathbf{W} \cdot (\mathbf{W} \cdot \mathbf{a}^0) = \mathbf{W}^2 \cdot \mathbf{a}^0;$$

$$\mathbf{a}^3 = \mathbf{W} \cdot \mathbf{a}^2 = \mathbf{W} \cdot (\mathbf{W}^2 \cdot \mathbf{a}^0) = \mathbf{W}^3 \cdot \mathbf{a}^0.$$

$$\mathbf{a}^k = \mathbf{W}^k \cdot \mathbf{a}^0,$$

$$\alpha^k = \mathbf{W}^k \cdot \alpha^0$$

$$= \mathbf{W}^k \sum (\alpha_i \cdot \lambda_i \cdot \mathbf{e}_i)$$

$$= \sum \alpha_i \cdot (\lambda_i)^k \cdot \mathbf{e}_i$$

$$= (\lambda_1)^k [\alpha_1 \mathbf{e}_1 + \sum_{i=2} \alpha_i \left[ \frac{\lambda_i}{\lambda_1} \right]^k \mathbf{e}_i].$$

$$\mathbf{a}^k = \alpha_1 \cdot (\lambda_1)^k \cdot \mathbf{e}_1$$

$$\mathbf{E} = -\frac{1}{2} \sum \sum w_{ij} \cdot \mathbf{a}_i \cdot \mathbf{a}_j, \quad i \neq j.$$

$$\mathbf{E} = -\frac{1}{2} \sum \sum w_{ij} \cdot \mathbf{a}_i \cdot \mathbf{a}_j.$$

$$\mathbf{E} = -\frac{1}{2} \mathbf{a}^T \cdot \mathbf{W} \cdot \mathbf{a},$$

$$\mathbf{E} = -\frac{1}{2} \cdot \frac{\mathbf{a}^T \cdot \mathbf{W} \cdot \mathbf{a}}{\mathbf{a}^T \cdot \mathbf{a}}.$$

$$\frac{\alpha^T \cdot \mathbf{W} \cdot \alpha}{\alpha^T \cdot \alpha},$$