Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 2433–2440, March 1995

Review

Visual motion perception

Thomas D. Albright and Gene R. Stoner*

The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037

ABSTRACT **The primate visual motion system performs numerous functions essential for survival in a dynamic visual world. Prominent among these functions is the ability to recover and represent the trajectories of objects in a form that facilitates behavioral responses to those movements. The first step toward this goal, which consists of detecting the displacement of retinal image features, has been studied for many years in both psychophysical and neurobiological experiments. Evidence indicates that achievement of this step is computationally straightforward and occurs at the earliest cortical stage. The second step involves the selective integration of retinal motion signals according to the object of origin. Realization of this step is computationally demanding, as the solution is formally underconstrained. It must rely—by definition—upon utilization of retinal cues that are indicative of the spatial relationships within and between objects in the visual scene. Psychophysical experiments have documented this dependence and suggested mechanisms by which it may be achieved. Neurophysiological experiments have provided evidence for a neural substrate that may underlie this selective motion signal integration. Together they paint a coherent portrait of the means by which retinal image motion gives rise to our perceptual experience of moving objects.**

Motion of an image on one's retina reflects a change in one's visual environment. From the point of view of survival, the ability to detect such motion is one of the most important functions performed by the primate visual system. Motion perception has been a subject of study for many decades in psychophysical experiments. More recently, it has also been a central focus of both neurobiological and computational studies. In consequence, it is a system for which we now have a unique and critical convergence of information. Indeed, visual motion processing is arguably the most well-understood sensory subsystem in the primate brain.

By comparison with the state of our knowledge, the scope of this review is fairly circumscribed. For example we will consider only the primate visual system and, specifically, the geniculo–striate–extrastriate pathways. Our discussion will also be limited to but a small sweep of the manifold functions performed by the primate motion processing subsystem. We will begin by briefly reviewing the origin and substance of the contemporary view that visual motion is processed by a basic and well-defined neural system.

The Motion Processing Apparatus

Elementism—the reductive interpretation of perception as merely the association of a set of basic sensory elements—was a dominant theme in mid-19th century perceptual psychology. According to this influential doctrine, motion was not thought to be directly sensed but rather indirectly inferred from the association of light sensations at different points in space and time. That is to say, motion was not thought to be directly sensed but rather indirectly inferred. This view of motion perception began to fade in the latter part of the 19th century as its explanatory limitations became widely recognized. In particular, the immediacy and directness of the motion sense were suggested by early reports of two motion "illusions": (*i*) the nonveridical experience of motion that follows prolonged exposure to real motion, a phenomenon known as the motion after-effect or waterfall illusion (1); and (*ii*) the experience of motion induced by light stimulation at discrete spatial and temporal intervals (2), a phenomenon known as apparent motion. The critical feature of both is the fact that movement can be perceived without the observer having detected changes in the "primary" positional cues, a fact that led Wertheimer (3) to insist on the validity of motion as a sensation unto itself, not uniquely reducible to other sensory events—a view that has carried us through the present day.

Although the primacy of the motion sense was thus commonly accepted by the time neurophysiological techniques became readily available, the foundation for our present understanding was set with the discovery of an explicit neural representation of motion in the form of cells that exhibit selectivity for the direction in which an image feature moves across the retina (Fig. 1). In primates, this property of directional selectivity is first seen at the level of primary visual cortex (area V1) (4, 5). These motion-sensitive V1 neurons constitute part of a larger functional subsystem, as evinced by the fact that they lie within an anatomically segregated pathway extending from retina through several cortical stages (Fig. 2; see ref. 7 for review). This pathway is commonly known as the "magnocellular" or "M" stream and is typically contrasted with the anatomically parallel and functionally complementary "parvocellular" or "P" stream. The M stream originates with a morphologically distinct subset (Pα) of retinal ganglion cells, which project uniquely to the magnocellular laminae of the thalamic lateral geniculate nucleus (LGN). The magnocellular output ascends cortically to area V1 and continues through several successive cortical visual areas including, most notably, the middle temporal visual area, which is commonly known as area MT (or V5).

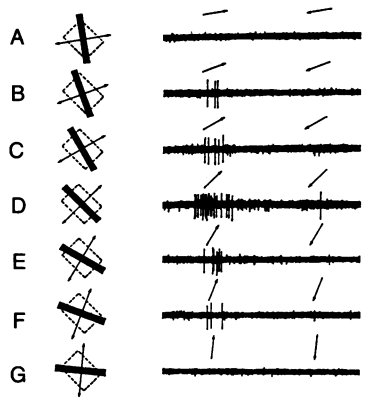

FIG. 1. Neuronal directional selectivity, as first observed by Hubel and Wiesel (4) in primary visual cortex (area V1) of rhesus monkey. (*Left*) The neuronal receptive field is indicated by broken rectangles. The visual stimulus was moved in each of 14 directions (rows A–G; opposing directions indicated by arrows) through receptive field. (*Right*) Recorded traces of cellular activity are shown in which the horizontal axis represents time (2 s per trace), and each vertical line represents an action potential. This neuron responds most strongly to motion up and to the right (row D). Reprinted with permission from ref. 4 (copyright 1968, The Physiological Society, London).

First described in the macaque by Dubner and Zeki (8), area MT is a small visuotopically organized zone (9–12) that

Abbreviations: 1-, 2-, and 3-D, one-, two-, and three-dimensional; area MT, middle temporal visual area; IOC, intersection of constraints; F/B, foreground/background assignment.

*To whom reprint requests should be addressed.

lies along the posterior extent of the lower bank of the superior temporal sulcus (Fig. 2) and is a recipient of ascending projections from areas V1, V2, and V3 (12, 13). Directional selectivity is by far the most distinctive feature of MT and the property that has drawn persistent interest since its discovery. By striking contrast to surrounding cortical areas, including others of the M stream, some 95% of MT neurons exhibit marked directional selectivity of the simplest form (i.e., selectivity for motion along a linear path in the frontal plane) in combination with a conspicuous absence of selectivity for form or color (10, 14, 15). It is the preeminence of these properties that led Zeki and others to the supposition that area MT is a principal component of the neural apparatus for motion processing.

Functions of the Motion Processing Apparatus

Visual motion is a source of information that can serve many functions for a behaving animal. These functions include (*i*) establishing the three-dimensional (3-D) structure of a visual scene (16–18), (*ii*) guiding balance and postural control (19, 20), (*iii*) estimating the direction of the observer's own path of motion (21) and his/her time to collision with objects or surfaces in the environment (22), and (*iv*) parsing retinal image features into the objects that are present in the visual scene ("image segmentation") (18, 23, 24). One of the most important and intuitively obvious goals of the motion processing apparatus is, of course, to recover and represent the trajectories of real-world objects in a form that will facilitate behavioral responses to those movements (25). It is this latter function that we will focus on in the remainder of our review.

The problem of representing object motion might, upon first consideration, seem to be among the most direct and accessible of motion-related functions, possibly because it is a salient, common, and apparently effortless part of conscious experience. As we shall see, however, it is normally a difficult and underconstrained computation and one that is necessarily dependent upon other sources of information about the structure of the visual scene. This problem of motion processing can be conveniently divided into two subproblems, which we will term "detection" and "interpretation." Solutions to these problems are thought to be computed at sequential stages in a hierarchical process, and they have been tentatively identified with specific neuronal populations in the primate brain.

Motion Detection

The problem of visual motion detection, which is common to all of the motion-related functions listed above has traditionally been cast in terms of the properties of retinal image features. As the pattern of light falling upon the retina undergoes displacement in space and time, the task of the motion detector is to detect the spatio-temporal continuity of image features. Posed in this manner, there are really two facets to the problem. The first pertains to the definition of "features" and the second addresses the mechanism responsible for "matching" these features in space-time.

Input to Motion Detectors: Retinal Image Cues and Motion Correspondence Tokens. What is it, precisely, that is matched over space and time? The question highlights the fact that motion detection is, by its very nature, dependent upon prior or coincident detection of contrast in the retinal image. The contrast elements potentially available as matching primitives are of various physical types (luminance, chrominance, texture, etc.) and levels of abstraction (e.g., local gradients, edges, and more complex object characteristics). Historically, most attempts to identify motion detection mechanisms have blurred issues of matching primitives with the matching operation itself. As a result, many of the subtypes of motion detectors that have been proposed to exist (see below) are defined by contrast type or level of abstraction rather than on the basis of the matching algorithm.

We will first consider more closely the level of abstraction at which spatio-temporal matches are made. At the simplest level, matching could occur prior to the encoding of any of the complex features—such as edges—that characterize our perception of natural scenes. The fact that a strong motion percept can arise from stimuli lacking any coherent spatial structure [such as random-dot cinematograms (26)] appears to support the idea (24, 27). On the other hand, the demonstration that motion is more likely to be seen between spatio-temporally displaced edges of the same orientation than between edges of differing orientation (28) suggests that more complex image "tokens" are matched by the motion system. The apparent contradiction between these psychophysical findings has traditionally been held as evidence for two motion detection subsystems: (*i*) "short-range" system that simply detects spatio-temporal continuity of local luminant energy over small displacements and (*ii*) a "long-range" system that is sensitive to higher-order stimulus attributes and operates over a larger spatio-temporal range.

The neurophysiological data bearing on the issue of matching primitives are less dichotomous. For most directionally selective neurons, these primitives are usefully defined by the other stimulus attributes for which cells express sensitivity. In area V1, this means that the matching

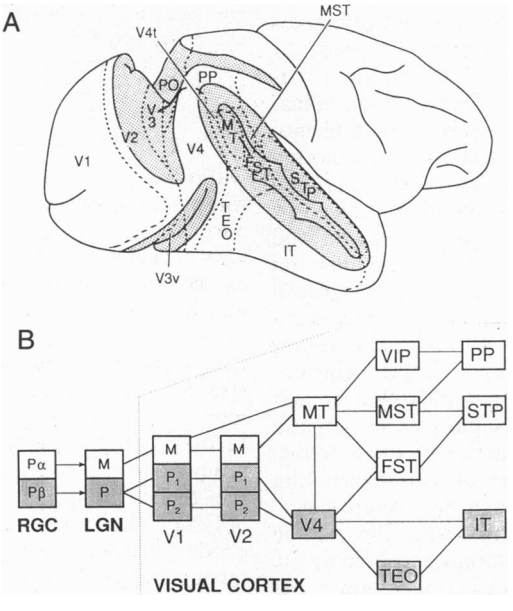

FIG. 2. (*A*) Lateral view of macaque brain showing location of striate cortex (V1) and some extrastriate visual areas. Sulci have been partially opened (shaded regions). Indicated borders of visual areas (dashed lines) are approximate. EC, external calcarine sulcus; IO, inferior occipital sulcus; IP, intraparietal sulcus; LA, lateral sulcus; LU, lunate sulcus; PO, parieto-occipital sulcus; ST, superior temporal sulcus. (*B*) Anatomical connectivity diagram emphasizing hierarchical organization and parallel processing streams along the geniculo–striate–extrastriate pathway. Except where indicated by arrows, connections are known to be reciprocal. Not all known components of magnocellular (unshaded) and parvocellular (shaded) pathways are shown. RGC, retinal ganglion cell layer; LGN, lateral geniculate nucleus of the thalamus; M, magnocellular subdivisions; P_1 and P_2, parvocellular subdivisions; MT, middle temporal; MST, medial superior temporal; FST, fundus superior temporal; PP, posterior parietal cortex; VIP, ventral intraparietal; STP, superior temporal polysensory. Reprinted with permission from ref. 6 (copyright 1993, Elsevier Science, Amsterdam).

token typically consists of a one-dimensional (1-D) image contour of a particular orientation (4, 5). These early motion detectors thus employ a set of matching primitives that is richer than that commonly associated with the "low-level" short-range system—a conclusion that seems at variance with the traditional short-range/long-range dichotomy. Additional cause to question this view comes from recent psychophysical experiments that fail to find any sharp distinction between the matching primitives important for small vs. large spatio-temporal displacements (29).

Cavanagh and Mather (29) offer an alternative dichotomy of motion-detection subsystems founded on the type of image contrast (rather than feature complexity *per se*) that defines a moving feature. Specifically, they suggest a division between mechanisms sensitive to "first-order" vs. "second-order" image variation. First-order refers to variation along the primary dimensions of luminance or color, whereas second-order image contrast includes variation along "secondary" dimensions, such as texture, binocular disparity, or luminance contrast modulation. Support for this dichotomy comes from psychophysical (e.g., ref. 30) and neurophysiological (e.g., ref. 31) evidence suggesting that the motion of first-order image features is detected independently of that for features defined by second-order cues.

Regardless of whether different types of image variation are extracted by independent mechanisms, we predict that at some level there should exist motion-sensitive neurons that represent the displacement of a feature, whether it is defined by color, luminance, or texture. This prediction of form-cue invariance (32, 33) is based on the observation that the velocity of an object is physically unrelated to the cue that distinguishes that object in the retinal image. Perceptual form-cue invariance for moving features has been demonstrated for luminance and chrominance (34, 35) as well as a variety of second-order cues, such as texture (36, 37) and stereoscopic disparity (38). A potential neuronal substrate for form-cue invariant perception has been found in area MT. In addition to its well-known sensitivity to conventional luminance-defined motion (10, 14, 15), many cells in this cortical area also respond to the motion of chromatically defined stimuli (39–41) as well as motion of second-order cues (32). The functional utility of such an arrangement is quite clear: the system gains more uniform sensitivity to motion over the broad spectrum of cues that are characteristic of our visual world.

Computations Underlying Motion Detection. Spatio-temporal comparison is fundamental to the detection of motion whatever the type of image cue or level of abstraction that happens to define a moving image feature. It thus seems plausible that the neural circuitry underlying motion detection may—like perception and the responses of some neurons—be form-cue invariant. While verification of this prediction must await more detailed analyses of neuronal circuitry, the computational principles outlined below should be considered in this broad context.

Several detailed models have been proposed to account for the computations carried out by motion detectors (e.g., refs. 42–47). The earliest complete model was developed nearly 40 years ago by Hassenstein and Reichardt (42) to explain the behavioral sensitivity to visual motion exhibited by winged arthropods. According to this "correlation model," motion is simply sensed by computing the product of the outputs of two receptors possessing luminance sensitivity profiles that are displaced in space and time. Subsequent neurophysiological studies of motion detection in the rabbit retina (48) demonstrated the existence of a similar correlation-type operation, based upon delayed inhibition extending laterally from one receptor's output to that of a spatially adjacent receptor (Fig. 3). When a visual stimulus moves in the appropriate direction, the flow of inhibition through the circuit parallels the spatial displacement of the stimulus, thereby nulling any potential receptor output. For other directions, the inhibition is less marked or nonexistent. The result is neuronal activity that varies as a function of direction of stimulus motion (i.e., directional selectivity). Similar evidence applies to the mechanisms for motion detection in other mammalian systems (e.g., ref. 49).

"Motion energy" models (44, 45) emphasize the processing of motion in spatio-temporal frequency domain and offer an alternative to correlation models. Although the computations that underlie motion energy and correlation models are formally equivalent to one another (43), the two strategies do suggest somewhat different neural implementations. The results of recent neurophysiological experiments appear to support the motion energy form of implementation (50).

1-D vs. Two-Dimensional (2-D) Motion Signals. The retinal surface is 2-D. Accordingly, the trajectory of a moving retinal image (cast by a moving object) requires a 2-D vector for adequate description. The motion signals rendered by the detection processes we have described are, however, inherently 1-D. Specifically, the V1 neurons that are thought to constitute the motion-detection stage are individually sensitive to image-contrast gradients along a single dimension—i.e., they are orientation selective. Each neuron of this sort can only detect image motion along the dimension for which it possesses contrast gradient sensitivity—along the axis perpendicular to the preferred orientation. All other image motion is effectively invisible. Because it is 1-D, the output provided by each motion detector is thus ambiguous with respect to the 2-D motions of the retinal image. A secondary processing stage is needed to integrate 1-D motion signals and thereby interpret the visual scene events that gave rise to them.

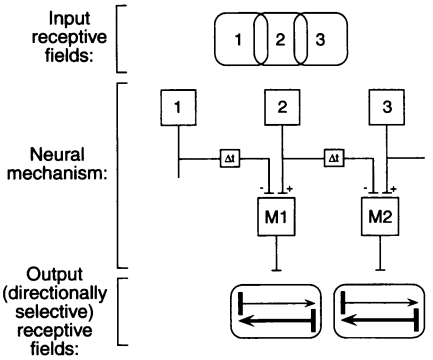

FIG. 3. Simple spatio-temporal comparator for directional motion detection by neurons, based upon computational principles of Hassenstein and Reichardt (42) and neurophysiological observations of Barlow and Levick (48). The neural circuit (*Middle*) consists of a repeating sequence of "input" neurons 1–3 with spatially adjacent receptive fields (*Top*). Each input neuron possesses both an excitatory connection (+) that extends directly and a time-delayed (△t) inhibitory connection (−) that extends laterally and asymmetrically to neurons at the motion detection stage (M1, M2). These motion-sensitive neurons operate as NOT-AND gates, and the lateral temporally delayed connection confers upon them the property of directional selectivity. Thus, leftward motion (receptive field stimulation sequence 3,2,1) yields a direct excitatory input to each motion detector in turn. Under these conditions, the lateral inhibitory connections have no functional effect. By contrast, rightward motion triggers an inhibitory signal that propagates laterally in parallel with the displacement of the input activity. Each motion detector thus responds more strongly to leftward than to rightward motion across its receptive field (*Bottom*).

Motion Interpretation

The problem of visual motion interpretation is that of recovering the trajectories of visual scene elements (objects) from the retinal image information provided by the motion detection stage—the sort of representation that both forms our conscious experience of motion and enables us to interact effectively with a dynamic environment. While we have seen that it is a computationally straightforward matter to detect the displacement of retinal image features, the subsequent representation of real object motions is greatly inconvenienced by the fact that such motions are not uniquely evident from the dynamic patterns of light in the retinal image. The reason for this is quite clear

and not unexpected: Determination of object motion in a 3-D world from a 2-D projected image is a facet of the inverse problem of optics, for which no unique solution exists (51).

In principle, knowledge of the physical rules by which reflected lights mix in the formation of the retinal image should facilitate recovery of the scene elements that gave rise to the image. That the primate visual system has "knowledge" of such rules is implied by the very fact that we are able to segment complex retinal images into component objects. Retinal image features that are, according to these rules, indicative of specific object interrelationships are appropriately termed image segmentation cues. While the importance of such cues for object recognition has long been appreciated (52), recent evidence suggests that they are also crucial to the interpretation of motion (33, 52–54).

As illustration of this point, consider the image in Fig. 4. That is, of course, a single frame from a dynamic sequence, which is meant to convey some of the complexities of natural scenes. The goal of this motion subsystem is to represent the independent trajectories of the scene elements: the monkey, the occluding branch, and the shadow. As we have seen, however, each primary motion detector merely renders a signal indicating the presence of a moving feature (an oriented contour, for example) at a particular location in the retinal image. Moving objects are typically, as in this example, formed from an amalgam of such motion correspondence tokens, which implies that motion interpretation depends upon integration of primary motion signals (55). Moreover, the integration process must be selective, utilizing image segmentation cues to link retinal motion signals according to the object of origin (33).

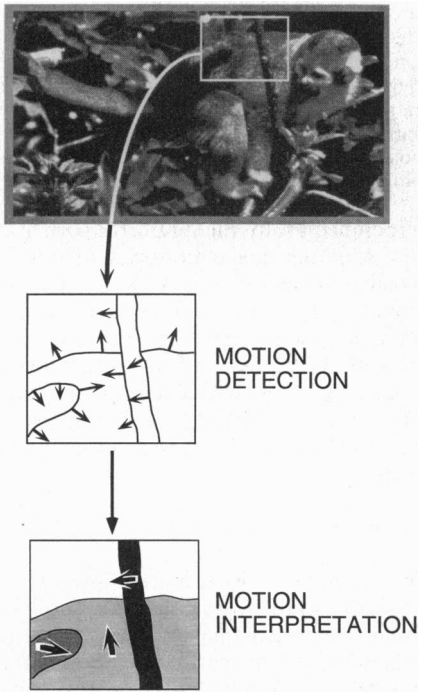

FIG. 4. (*Top*) Single frame from dynamic sequence illustrating complexity of motion interpretation in natural scenes. (*Middle*) The motion detection stage extracts local 1-D components of velocity for retinal image features. Individually, these motion measurements are ambiguous with respect to motions of the objects from which they arise. (*Bottom*) The task of the motion interpretation stage is to selectively combine detection-stage measurements according to the object of origin. The desired outcome is a representation of the independent motions of objects.

The mechanics of integration and the means by which selectivity is imposed have become amenable to study using visual stimuli known as "plaid patterns" (56, 57), which are formed by spatial superposition of two drifting periodic gratings (Fig. 5). Plaids provide a simple laboratory analogue to natural conditions that give rise to overlapping moving contours in the retinal image; as happens, for example, when one moving object passes in front of another. Their utility derives from the fact that under some conditions the two component gratings are seen to drift "noncoherently" past one another, whereas under other conditions, the gratings form a rigid 2-D pattern that moves "coherently" in a single direction. These robust perceptual phenomena thus serve as a model for the selective integration of motion signals. By examining the character of the coherent motion percept, it may be possible to infer something about the integration mechanism itself. Likewise, by exploring the image conditions that lead to the vastly dissimilar coherent and noncoherent motion percepts, it may be possible to discover the mechanisms that govern selectivity. Finally, by employing plaid stimuli in neurophysiological experiments, it should be possible to reveal the neuronal structures and events responsible for selective integration of motion signals.

The Mechanics of Integration. *Computational considerations.* Consider the simple case of a single moving object with a many-faceted boundary contour (Fig. 6). According to the preceding arguments, early detectors represent, in a piece-wise fashion, the motion of retinal features along limited regions of the contour. Because the motion signal rendered by each is inherently 1-D, the true 2-D trajectory of the pattern is only knowable by integrating information from multiple detectors. Formally speaking, true pattern motion can be determined from the intersection of constraints provided by two or more 1-D motion signals (Fig. 6)—that is, the single 2-D direction and speed that is consistent with the evidence provided by each 1-D signal (47, 57). Although the mechanism remains to be determined, one can readily envision a circuit whereby a neuron representing a specific 2-D pattern trajectory is supported by an ensemble of neurons providing appropriate 1-D motion signals (15, 55, 57). Accordingly, simultaneous activation of two or more inputs would engage a unique pattern motion detector.

Psychophysics. The hypothesis that the integration mechanism actually exploits the IOC rule has been lent support by several psychophysical studies employing plaid patterns (e.g., refs. 57–59). Wilson and colleagues (e.g., ref. 60) have observed, however, that some plaid configurations ("type II" plaids, which unlike the example in Fig. 5 possess component directions that both lie to one side of the pattern direction) lead to a percept that differs from the IOC prediction but is approximated by a vector-sum of 1-D component motions (see ref. 61 for review). *A priori*, the general utility of vector summation as an approach to the motion signal integration problem would appear questionable, as (in contrast to IOC) it

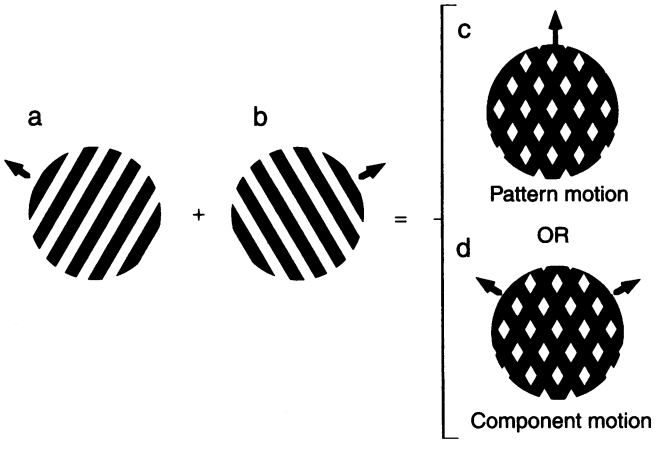

FIG. 5. Moving plaid patterns are produced by superimposition of two drifting periodic gratings (*a* and *b*) (56, 57). The resultant percept is either that of a coherently moving 2-D pattern (*c*) or two 1-D gratings sliding past one another (*d*), depending upon a variety of stimulus parameters.

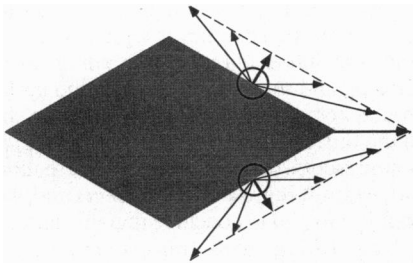

FIG. 6. Intersection of constraints (IOC) solution to motion signal integration. When examined locally (as implied by circles), the apparent motion of a 1-D image feature embedded in a 2-D pattern is inherently ambiguous but physically constrained to a "family" of 2-D image motions (as suggested by the vectors that terminate on the broken constraint lines). The IOC from two or more 1-D motion measurements defines the 2-D velocity of the pattern (bold rightward arrow).

does not guarantee a veridical solution. The nonveridical percept associated with type II plaids may be rare, however, when viewing natural scenes because of the more uniform distributions of components possessed by most natural moving objects. Moreover, as Wilson (61) and others (62) have argued, this limitation can be corrected by utilizing information given by second-order motion cues present in plaid patterns. In any event, the psychophysical evidence accumulated thus far can arguably be accommodated by integration schemes based on either IOC or vector-sum architectures. It thus seems that final resolution of this controversy must await anatomical or physiological studies of functional connectivity between motion processing stages.

Neurophysiology. We have seen that the motion detection stage can be identified with directionally selective neurons in area V1. A goal of recent neurophysiological experiments has been to identify the level of the cortical hierarchy at which the integration of these primary motion signals takes place (55, 63, 64). It was predicted that this secondary stage would be recognizable by the emergence of neurons that exhibit selectivity for the direction of motion of 2-D patterns rather than for the motion of the oriented features that make up those patterns (55). This prediction can be tested by comparing the directional tuning elicited by a 1-D stimulus, such as a single grating, with that elicited by a plaid pattern, composed from two such gratings. The preferred direction of motion is determined for each neuron in a conventional fashion by using the single drifting grating. When stimulated with plaid patterns, neurons that respond best when either component moves in the preferred direction are by definition preintegration and are termed "component neurons." This is the expected behavior of the motion detection stage, which we considered above. By contrast, neurons that respond best when the entire plaid pattern moves in the preferred direction reflect integration of component motions and are termed "pattern neurons."

As anticipated, neurons in area V1 are of component type by this criterion (55), as are the majority of neurons (≈60%) in area MT. A subset of MT neurons (≈25%), however, are of the more interesting pattern type (Fig. 7; refs. 55, 63, and 64). Neurophysiological analysis has thus identified specific neural events that correspond to the outcome of the predicted computational process and may underlie the perceptual integration of motion signals.

Selectivity of Integration. *Computational considerations.* We come closer to the problem of motion interpretation when we evaluate the complexities of selective motion signal integration. Consider the case in which two moving objects have trajectories such that their projections overlap in the retinal image. Here, contrary to the single object case, motion interpretation depends not simply upon integrating motion signals. Rather, it is also conditional upon independently integrating only those signals common to each object (such that, for example, motion signals arising from the shadow in Fig. 4 are not pooled with those arising from the monkey). Posed as a generic matter of combining retinal motion signals, the problem has no unique solution. This notwithstanding, the primate visual system clearly does pretty well. To achieve the selective motion integration suggested by our experience of complex dynamic scenes, the underlying mechanism necessarily exploits retinal image segmentation cues, which by definition reflect the spatial interrelationships (adjacency or superposition vs. contiguity) between surfaces in the visual scene (33). By simulating the optical conditions of a ubiquitous real-world scene, two overlapping moving surfaces, moving plaid patterns have afforded a simple and direct means to evaluate the expected role of image segmentation cues in motion integration.

As indicated previously (Fig. 5), there are conditions under which the components of a plaid pattern are perceived to slide noncoherently across one another. For example, noncoherence becomes more likely if the components differ significantly along a particular stimulus dimension, such as luminance contrast, spatial frequency (55, 57), or color (65, 66). These results traditionally have been interpreted as manifestations of channel-specific integration mechanisms, in which dissimilar components are processed by independent neural channels (55). While such proposals are valid and focus our attention on the mechanisms potentially responsible for selective integration, they generally fail to address the functional utility of this process. We therefore advocate a complementary view in which selective integration is interpreted in a broad functional light: featural dissimilarity may be regarded as a source of information that fosters image segmentation, with the corresponding effects on motion

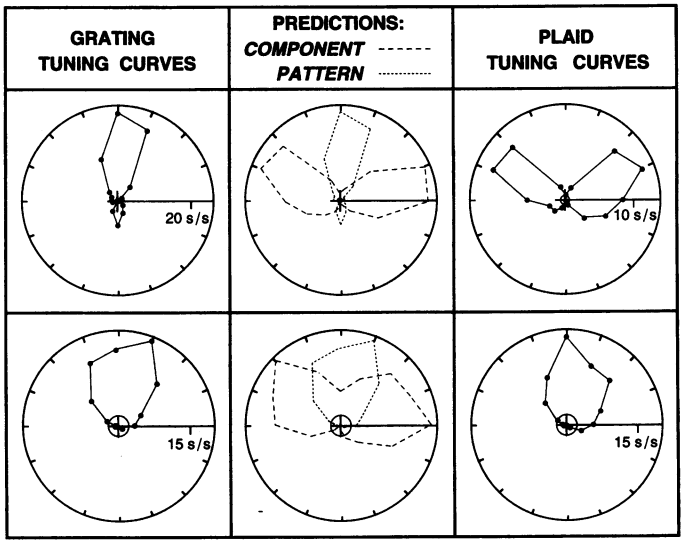

FIG. 7. Data from two MT neurons representing "component" (*Upper*) and "pattern" (*Lower*) stages of motion processing. Direction tuning curves were acquired using a drifting sine-wave grating (*Left*) or a perceptually coherent plaid pattern (*Right*). Responses elicited by each stimulus type, moving in each of 16 directions, are plotted in polar format. The radial axis represents response amplitude [spikes per second (s/s)], the polar angle represents direction of motion, and the small central circle represents spontaneous activity level. Both cells exhibit a single peak in grating tuning curve. From these curves, responses to plaid patterns were predicted (55) (*Center*). Component predictions reflect sensitivity to both oriented gratings in the plaid pattern. Pattern predictions reflect sensitivity to composite appearance of the plaid. By definition, behavior of component neuron conforms to component prediction, while that of pattern neuron conforms to pattern prediction. Adapted from ref. 63; printed with permission (copyright 1989, Springer, Berlin).

coherence born out of this process (33). This "image segmentation hypothesis" has a far-reaching implications, which can be explored empirically through the introduction of image segmentation cues to plaid patterns.

Psychophysics. There are many manipulations that can be used to elicit image segmentation. Of particular interest in this context are cues that promote perceptual depth ordering of spatially overlapping features. Perhaps the most obvious cue in this class is binocular disparity. As predicted by the image segmentation hypothesis, plaids composed of two gratings that lie in different stereoscopically defined depth planes are perceived to move noncoherently (54, 67).

It is also well known that depth ordering can be elicited by luminance-based cues for transparency and opaque occlusion. The "rules" by which these cues operate have been extensively documented (e.g., ref. 68), and they are largely dictated by the physics of light mixture in optical projection. Briefly, the intensity of light corresponding to the region at which two surfaces overlap will normally be equal to the sum of (*i*) light reflected directly from the foreground surface and (*ii*) light reflected from the background surface and subsequently attenuated as it passes through the foreground surface. If, for example, foreground attenuation is complete, the background luminance contribution is nil and the "intersection luminance" is simply equal to the foreground luminance. This is the unique case of occlusion. Alternatively, the foreground surface may attenuate incompletely and possess no reflectance of its own. The latter conditions of transparency are characteristic of shadows. Because the reflectances of both surfaces can vary independently, as can the foreground attenuation factor, there exists a broad range of luminance configurations compatible with transparency/occlusion. The primate visual system typically recognizes—instantly and effortlessly—configurations within this range as having resulted from the depth-ordering conditions that would normally give rise to them.

Stoner *et al.* (53) incorporated these rules for transparency/occlusion in the construction of plaid patterns (Fig. 8). By varying the luminance of a single (repeating) subregion of the plaid, it was possible to form retinal image conditions that were either physically consistent or inconsistent, with one component grating overlapping another. As predicted, human subjects were very likely to perceive noncoherent motion of the individual components when the plaid was configured for transparency or occlusion. By contrast, coherent motion of the plaid became the dominant percept when the luminance relationships were physically incompatible with transparency/occlusion.

Stoner and Albright (unpublished data) extended this result with a demonstration of the additional constraints on motion coherence contributed by pictorial cues for image segmentation. Specifically, they exploited the fact that perceptual transparency/occlusion, as manipulated by luminance relationships, is inherently dependent upon establishing which image features are interpreted by the observer as foreground and which are interpreted as background. For example, the plaid unit illustrated in Fig. 9*a* is only physically consistent with transparency if region A represents the intersection of two foreground features, and region D represents the background across which the features move. If, on the other hand, we were told that region D represents the foreground intersection and region A represents the background, we would be forced to conclude that the pattern is incompatible with transparency/occlusion because the intersection luminance is brighter than is physically possible.

Whether region A or D happens to be foreground in the real world is, of course, impossible to determine from these retinal images. Fortunately the primate visual system furnishes default perceptual interpretations by using other sources of information and probabilistic rules. For example, smaller image features are generally interpreted (all else being equal) as foreground (23). Thus, the default interpretation for the image in Fig. 9*a* holds region A to be foreground. By simply reversing the relative sizes of regions A and D, the opposite percept ensues (Fig. 9*b*). Stoner and Albright (unpublished data) used these and other means to manipulate F/B assignment in plaid patterns. Because perceptual transparency depends upon F/B, they predicted that motion coherence would be markedly influenced by F/B manipulations. In accordance with these predictions, human subjects were found most likely to report noncoherent motion when the presumptive foreground intersection was compatible with transparency/occlusion.

These results and others of a similar nature (54, 66, 71, 72) are consistent with the proposal that image segmentation cues influence the perceptual integration of motion signals. The functional utility (indeed, necessity) of such a scheme is clear: Survival in a dynamic environment is critically dependent upon the ability to distinguish motions of occlusive or transparent objects, as well as shadows, from the motions of the surfaces over which they move. However, in the light of current theories of motion detection, these transparency phenomena are truly puzzling, and the mechanism behind them has become a subject of much debate.

Mechanism. One type of mechanism that has merited serious consideration is derived from the fact that transparency manipulations introduce moving luminance gradients to which primary motion detectors are known to be sensitive (refs. 62 and 73; unpublished data). It is thus conceivable that transparent plaid patterns simply elicit greater activity from primary motion detectors tuned to the pattern direction than do nontransparent plaids. Indeed, if one accepts the existence of certain nonlinearities in the encoding of retinal image contrast (70), it can be shown that the magnitude of the luminance gradient moving in the pattern di-

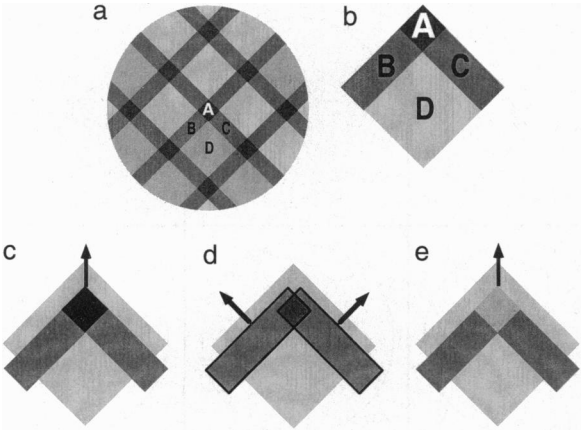

FIG. 8. Procedures for creating transparent plaids are derived from rules of light mixture in formation of retinal image (68). (*Upper*) Each plaid (*a*) can be viewed as a tesselated image composed of four repeating subregions (*b*) identified as A, B, C, and D. Region D is normally seen as background (because of its larger size). Regions B and C are seen as narrow overlapping surfaces and remaining region D as their intersection. Perceptual transparency can be manipulated by varying the luminance of region A, while luminances of regions B, C, and D remain constant. (*Lower*) Icons depict luminance configuration for three examples, along with an indication of dominant percept. For one of these conditions (*d*), luminance of region A is chosen to be consistent with transparency, yielding a percept of noncoherent motion. For the other two conditions illustrated, luminance of region A is either too dark (*c*) or too bright (*e*) to be consistent with transparency. These nontransparent plaids generally elicit a percept of coherent pattern motion. Adapted from ref. 64; printed with permission (copyright 1992, Macmillan, London).

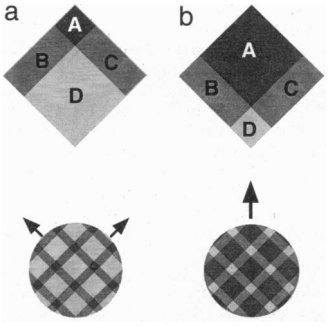

FIG. 9. Schematic depiction of one of several means used by Stoner and Albright (unpublished data) to manipulate percept of foreground and background in plaid patterns. Plaids were constructed as tessellated versions of these basic units (see Fig. 8). (*a Upper*) The largest region D is typically seen as background, and with that constraint, luminance configuration is compatible with transparency. (*a Lower*) As predicted, human subjects report a percept of component motion. (*b Upper*) By simply reversing the relative sizes of regions A and D, while leaving the luminance of each region unaltered, the dominant percept becomes such that region D is now seen as foreground. With that constraint, region D is not consistent with transparency (it is too bright). (*b Lower*) Predictably, subjects report a percept of 2-D pattern motion. These foreground/background (F/B) manipulations do not significantly alter the relative magnitude of luminance gradients moving in pattern and component directions (70).

rection is approximately minimal when the plaid in Fig. 8*a* is configured for transparency (refs. 62 and 73; unpublished data). In other words, the strength of the coherent pattern motion percept can be roughly accounted for by the magnitude of the luminance gradient moving in the pattern direction.

The computational appeal of this simple explanation is undercut, however, by the lack of generality it affords. Consider, for example, the aforementioned interactive effects of F/B and transparency on perceptual motion coherence. In this case, altering the relative sizes of pattern subregions can trigger a radical change in the observer's state of perceptual coherence. However, the relative size manipulations do not significantly alter the fraction of luminant energy drifting in the pattern direction (ref. 62; unpublished data). In other words, the result is incompatible with the predictions of the luminance gradient hypothesis. Similarly, the reported conjoint influences of binocular disparity and luminance cues for transparency (54) expose a dissociation between the presence of luminance gradients and motion coherence. Finally, and perhaps most revealingly, perceptual transparency and motion coherence also parallel spontaneous ("metastable") reversals of F/B that occur in the absence of any modifications to the retinal stimulus (unpublished data).

It would thus seem that the only common determinant of motion coherence is perceptual assignment of the component intersection feature (region A in Fig. 8*a*) as either "intrinsic" (a variation in surface reflectance) or "extrinsic" (an incidental consequence of overlap in the formation of the retinal image) to the plaid pattern (71). While this conclusion carries important functional implications, it begs the question of mechanism. One promising approach to this problem involves the use of a "feature selection" module (74). The function of such a mechanism is to choose from among the available motion signals those that are most likely to be indicative of true pattern motion. When implemented as a neural network, this selective faculty can be acquired through training, and it effectively enables the motion integration stage to distinguish between motion signals arising from intrinsic vs. extrinsic sources (provided that the model has "learned" the relevant set of "rules"). Indeed, simulations of perceptual motion coherence performed with a feature selection model (74) yield results that are strikingly consistent with the original transparency results of Stoner *et al.* (53). Whether this approach can be made to generalize across the range of image factors known to influence perceptual motion coherence remains to be seen.

Neurophysiology. As we have seen, studies of cellular response properties have revealed a population of neurons in cortical visual area MT that may be responsible for perceptual motion coherence (55, 63). Because these pattern-type neurons are believed to play a significant role in the motion signal integration process, Stoner and Albright (64) hypothesized that neural responsivity would be altered by the same stimulus attributes known to influence the selectivity of perceptual motion coherence. To test this hypothesis, perceptual motion coherence was manipulated in a manner identical to the earlier psychophysical study (53), such that plaid patterns were either compatible or incompatible with transparency. Data obtained from a typical directionally selective MT neuron are shown in Fig. 10. When stimulated using nontransparent/perceptually coherent plaids, responses were strongest when the plaid moved in the cell's preferred direction. As described above (Fig. 7*Lower*), this style of selectivity is characteristic of pattern-type neurons. When stimulated using the transparent and perceptually noncoherent plaid, however, this cell's behavior underwent a marked transformation: The pattern response dropped while component responses became elevated. The resultant bilobed directional tuning curve is characteristic of component type neurons (Fig. 7*Upper*). In other words, when the stimulus was configured to elicit a percept of coherent pattern motion, the cell appeared to represent that motion. By contrast, when the stimulus was designed to elicit a percept of two components sliding past one another, the cell appeared to represent those independent motions. Similar behavior was observed for the majority of MT neurons studied.

The plasticity expressed by such neurons stands as a potential neural substrate for the selective quality of the perceptual experience. The circuitry responsible for this neuronal phenomenon is yet unknown. Possibilities include enhanced synaptic efficacy for inputs to the pattern motion stage that reflect motion of reliable or intrinsic features. Such proposals appear to be compatible with the various mechanisms that have been proposed for integration itself (i.e., IOC, vector summation). They are, moreover, in line with the feature selection model mentioned above (74), and they could be implemented by altering the temporal synchrony of active component inputs (75). Much more neurophysiological data are needed, however, to weigh these various possibilities.

Concluding Remarks

Nearly 50 years ago, and not long after the advent of neurophysiological techniques,

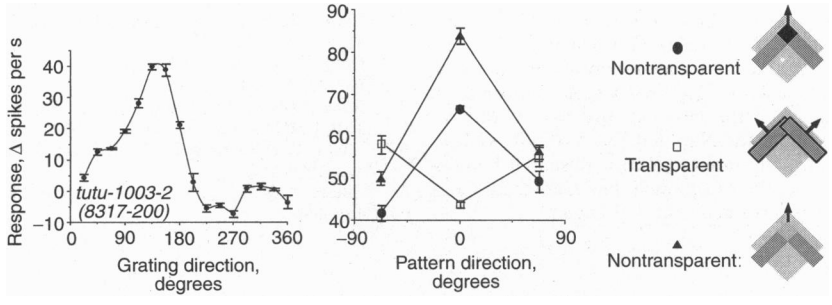

FIG. 10. Neural correlates of perceptual motion signal integration in cortical visual area MT. (*Left*) Directional tuning for single drifting grating. (*Center* and *Right*) Responses to coherent (nontransparent) and noncoherent (transparent) plaids. When stimulated with coherent plaid patterns, response (●, ▲) was maximal when the pattern moved in the neuron's preferred direction (0° in *Center*). When stimulated with noncoherent plaid patterns, however, responses (□) were maximal when either component moved in the preferred direction (±67.5° in *Center*). Adapted from ref. 64; printed with permission (copyright 1992, Macmillan, London).

Donald Hebb (69) observed that "the psychologist and neurophysiologist chart the same bay." While one sometimes has the feeling that they sail silently past one another in the night, scarcely has Hebb's dictum proven more profitable than in the experimental study of the neural bases of visual motion perception. As we have tried to convey, attempts to understand motion perception necessarily involve the inseparable issues of phenomenology, function, and mechanism, and they benefit most when guided by the convergent lights of computation, physiology, anatomy, perception, and behavior. In measure of success, we have identified computational goals of the system, linked them to specific loci in a distributed and hierarchically organized neural system, and documented their functional significance in a real-world sensory/behavioral context. Nonetheless, it should be evident that many basic questions remain unanswered, posing formidable technical and conceptual challenges to modern neuroscience.

Finally, one of the most important lessons from this analysis of visual motion perception concerns the critical role of context in neurophysiological investigations. To see the motion of a retinal image feature (in the sense that we have considered it) means to interpret the real-world events that gave rise to it, and that interpretation is quite impossible without context. The point hardly needs definition, as it has been a linchpin of experimental psychology for well over 100 years (e.g., ref. 23). Nonetheless, for a variety of reasons, it is a point that has been largely neglected in neurophysiological studies of vision. The reviewed studies of motion signal integration suggest that neuronal as well as perceptual response to a moving retinal image feature is critically dependent upon contextual factors that influence image interpretation and are unrelated to motion *per se*. Only by adopting a contextual approach will it be possible for neurophysiologists and psychologists to jointly chart the many murky waters of perception.

We thank J. Costanza for superb technical assistance. G. Buračas, G. Carman, L. Croner, and R. Duncan provided helpful comments on the manuscript. This work was supported by grants from the National Institute of Mental Health and the National Eye Institute. G.R.S. was partially supported by a Research Fellowship from the McDonnell–Pew Center for Cognitive Neuroscience at San Diego.

1. Addams, R. (1834) *Philos. Mag.* **5,** 373.
2. Exner, S. (1875) *Arch. Ges. Physiol.* **11,** 403–432.
3. Wertheimer, M. (1961) in *Classics in Psychology*, ed. Shipley, T. (Philosophical Library, New York), pp. 1032–1089 (originally published in 1912).
4. Hubel, D. H. & Wiesel, T. N. (1968) *J. Physiol. (London)* **195,** 215–243.
5. Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976) *J. Neurophysiol.* **39,** 1288–1319.
6. Albright, T. D. (1993) in *Visual Motion and Its Role in the Stabilization of Gaze*, eds. Miles, F. A. & Wallman, J. (Elsevier, Amsterdam), pp. 177–201.
7. Livingstone, M. S. & Hubel, D. H. (1988) *Science* **240,** 740–749.
8. Dubner, R. & Zeki, S. M. (1971) *Brain Res.* **35,** 528–532.
9. Allman, J. M. & Kaas, J. H. (1971) *Brain Res.* **31,** 85–105.
10. Zeki, S. M. (1974) *J. Physiol. (London)* **236,** 549–573.
11. Gattass, R. & Gross, C. G. (1981) *J. Neurophysiol.* **46,** 621–638.
12. VanEssen, D. C., Maunsell, J. H. R. & Bixby, J. L. (1981) *J. Comp. Neurol.* **199,** 293–326.
13. Ungerleider, L. G. & Mishkin, M. (1979) *J. Comp. Neurol.* **188,** 347–366.
14. Maunsell, J. H. R. & Van Essen, D. C. (1983) *J. Neurophysiol.* **49,** 1127–1147.
15. Albright, T. D. (1984) *J. Neurophysiol.* **52,** 1106–1130.
16. von Helmholtz, H. (1924) *Physiological Optics*, Vol. 3 [English translation by Southall, J. P. C. from (1909) *Handbuch der Physiologischen Optik* (Voss, Hamburg, Germany), 3rd. Ed.].
17. Wallach, H. & O'Connell, D. N. (1953) *J. Exp. Psychol.* **45,** 205–217.
18. Nakayama, K. & Loomis, J. M. (1974) *Perception* **3,** 63–80.
19. Gibson, J. J. (1950) *The Perception of the Visual World* (Houghton Mifflin, Boston).
20. Lee, D. N. (1980) *Philos. Trans. R. Soc. London B* **290,** 169–179.
21. Koenderink, J. J. (1986) *Vision Res.* **26,** 161–180.
22. Lee, D. N. (1976) *Perception* **5,** 437–457.
23. Koffka, K. (1935) *Principles of Gestalt Psychology* (Routledge & Kegan Paul, London).
24. Braddick, O. (1974) *Vision Res.* **14,** 519–527.
25. Gibson, J. J. (1979) *The Ecological Approach to Visual Perception* (Houghton Mifflin, Boston).
26. Julesz, B. (1971) *Foundations of Cyclopean Perception* (Univ. of Chicago Press, Chicago).
27. Anstis, S. M. (1970) *Vision Res.* **10,** 1411–1430.
28. Ullman, S. (1979) *The Interpretation of Visual Motion* (MIT Press, Cambridge, MA).
29. Cavanagh, P. & Mather, G. (1989) *Spatial Vision* **4,** 103–129.
30. Ledgeway, T. & Smith, A. T. (1994) *Vision Res.* **34,** 2727–2740.
31. Zhou, Y. X. & Baker, C. L. Jr. (1993) *Science* **261,** 98–101.
32. Albright, T. D. (1992) *Science* **255,** 1141–1143.
33. Stoner, G. R. & Albright, T. D. (1993) *J. Cognit. Neurosci.* **5,** 129–149.
34. Cavanagh, P. & Anstis, S. M. (1991) *Vision Res.* **31,** 2109–2148.
35. Dobkins, K. R. & Albright, T. D. (1993) *Vision Res.* **33,** 1019–1036.
36. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973) *Vision Res.* **13,** 1399–1401.
37. Chubb, C. & Sperling, G. (1988) *J. Opt. Soc. Am. A* **5,** 1986–2006.
38. Julesz, B. & Payne, R. A. (1968) *Vision Res.* **8,** 433–444.
39. Dobkins, K. R. & Albright, T. D. (1994) *J. Neurosci.* **14,** 4854–4870.
40. Saito, H., Tanaka, K., Isono, H., Yasuda, M. & Mikami, A. (1989) *Exp. Brain Res.* **75,** 1–14.
41. Gegenfurtner, K. R., Kiper, D. C., Beusmans, J. M. H., Carandini, M., Zaidi, Q. & Movshon, J. A. (1994) *Visual Neurosci.* **11,** 455–466.
42. Hassenstein, B. & Reichardt, W. (1956) *Chlorophanus Z. Naturforsch Teil B* **11,** 513–524.
43. Van Santen, J. P. H. & Sperling, G. (1985) *J. Opt. Soc. Am. A* **2,** 300–320.
44. Adelson, E. H. & Bergen, J. R. (1985) *J. Opt. Soc. Am. A* **2,** 284–299.
45. Watson, A. B. & Ahumada, A. J. (1985) *J. Opt. Soc. Am. A* **2,** 322–342.
46. Marr, D. C. & Ullman, S. (1981) *Proc. R. Soc. London B* **211,** 151–180.
47. Fennema, C. L. & Thompson, W. B. (1979) *Comput. Graphics Image Process.* **9,** 301–315.
48. Barlow, H. B. & Levick, R. W. (1965) *J. Physiol. (London)* **173,** 477–504.
49. Ganz, L. & Felder, L. (1984) *J. Neurophysiol.* **51,** 294–324.
50. Emerson, R. C., Bergen, J. R. & Adelson, E. H. (1992) *Vision Res.* **32,** 203–218.
51. Horn, B. K. P. (1986) *Robot Vision* (MIT Press, Cambridge, MA).
52. Nakayama, K. & Shimojo, S. (1990) *Cold Spring Harbor Symp. Quant. Biol.* **55,** 911–924.
53. Stoner, G. R., Albright, T. D. & Ramachandran, V. S. (1990) *Nature (London)* **344,** 153–155.
54. Trueswell, J. C. & Hayhoe, M. M. (1993) *Vision Res.* **33,** 313–328.
55. Movshon, J. A., Adelson, E. H., Gizzi, M. & Newsome, W. T. (1985) in *Study Group on Pattern Recognition Mechanisms*, eds. Chagas, C., Gattass, R. & Gross, C. G. (Pontifica Academia Scientiarum, Vatican City), pp. 117–151.
56. De Valois, K. K., De Valois, R. L. & Yund, E. W. (1979) *J. Physiol. (London)* **291,** 483–505.
57. Adelson, E. H. & Movshon, J. A. (1982) *Nature (London)* **300,** 523–525.
58. Welch, L. (1989) *Nature (London)* **337,** 734–736.
59. Burke, D. & Wenderoth, P. (1993) *Vision Res.* **33,** 343–350.
60. Ferrera, V. P. & Wilson, H. R. (1990) *Vision Res.* **30,** 273–287.
61. Wilson, H. R. (1994) in *Visual Detection of Motion*, eds. Smith, A. T. & Snowden, R. J. (Academic, London), in press.
62. Stoner, G. R. & Albright, T. D. (1994) in *Visual Detection of Motion*, eds. Smith, A. T. G. & Snowdon, R. J. (Academic, London), in press.
63. Rodman, H. R. & Albright, T. D. (1989) *Exp. Brain Res.* **75,** 53–64.
64. Stoner, G. R. & Albright, T. D. (1992) *Nature (London)* **358,** 412–414.
65. Krauskopf, J. & Farell, B. (1990) *Nature (London)* **348,** 328–331.
66. Kooi, F. L., De Valois, K. K. & Switkes, E. (1992) *Perception* **21,** 583–598.
67. Adelson, E. H. & Movshon, J. A. (1984) *J. Opt. Soc. Am. A* **1,** 1266.
68. Metelli, F. (1974) *Sci. Am.* **230,** 91–95.
69. Hebb, D. O. (1949) *The Organization of Behavior* (Wiley, New York).
70. MacLeod, D. I. A., Williams, D. R. & Makous, W. (1992) *Vision Res.* **32,** 347–363.
71. Shimojo, S., Silverman, G. H. & Nakayama, K. (1989) *Vision Res.* **29,** 619–626.
72. Vallortigara, G. & Bressan, P. (1991) *Vision Res.* **31,** 1967–1978.
73. Noest, A. J. & van den Berg, A. V. (1993) *Spatial Vision* **7,** 125–147.
74. Nowlan, S. J. & Sejnowski, T. J. (1994) *J. Neurosci.*, in press.
75. Gray, C. M., Koenig, P., Engel, A. K. & Singer, W. (1989) *Nature (London)* **338,** 334–337.