

Seeing the Big Picture: Integration of Image Cues in the Primate Visual System

Review

Lisa J. Croner and Thomas D. Albright*
Howard Hughes Medical Institute
The Salk Institute for Biological Studies
10010 North Torrey Pines Road
La Jolla, California 92037

The effortlessness of visual perception belies the complexity of the task. For example, our perceptions of objects depend upon the integration of many different sources of visual information, or visual cues, such as color, motion, texture, and brightness (throughout this review, the term “cue” refers to a light parameter, such as color, texture, etc., that can vary to produce visible contrast; the term “feature” refers to a spatially localized instance of such contrast, possibly in a particular form—e.g., a dot, edge, or corner). Integration of these cues pervades all levels of visual processing in the primate brain, from early detection and representation of image features to object recognition and the transformation of sensory information into motor output. Befitting the central role the integration of visual cues plays, its study has a long and rich history—riddled with controversy. Recently, however, the development of experimental approaches that allow for parallel acquisition of perceptual and neurophysiological data and the emergence of technology for controlling complex visual stimuli have begun to reveal a coherent picture of both the functional significance of cue integration and the underlying neuronal mechanisms. The relationship between the cues of color and motion illustrates the topic and, owing to the availability of new data, serves to introduce a family of concepts, methods, and conclusions.

The predominant view of visual cue integration has roots in two fundamental concepts of nineteenth century neuroscience: (1) elemental qualities of sensation and (2) localization of function within the brain. The former concept, known as “elementism,” holds that perception results from an association of sensations and is reducible to a set of physically independent sensory “elements,” such as color, brightness, and motion (Boring, 1942). Localization of function, which is the more familiar of the two nineteenth century concepts, embraces the now extensively documented fact that specific functional processes can be localized to specific brain regions. In the twilight of the twentieth century, these two concepts have been explicitly joined in the idea that there are specific neuronal representations for the elemental qualities of vision—i.e., some cortical areas code for color processing, others for motion processing, etc. As we shall see, several lines of anatomical and physiological data appear to support this modular view, which leads to an important question: how are categories of image information that are thought to be processed independently linked together to yield a unified perceptual experience?

Our goal in this review is to consider a number of observations that address this question, in the light of a theoretical framework for visual cue integration. In doing so, we lean heavily upon research from our laboratory, where several studies have been directed toward understanding the relationships between the neuronal representations of color and motion in the primate visual system.

Functional Goals and Operational Principles

We can begin to consider how cue integration occurs by first noting that the simple modular processing scheme outlined above is flawed, inasmuch as it considers—as did the proponents of elementism—that information about different cues is combined linearly, with no interactions between cues. As has been documented repeatedly by psychologists of the Gestalt tradition (e.g., Koffka, 1935), the association of a set of sensory cues frequently leads to perceptual consequences that are not reducible to the individual effects of each cue. On the contrary, the perceptual interpretation of a cue is often swayed by the context of other cues with which it appears. Constraints on the nature of the interactions among visual cues arise when one considers how the visual system functions as a whole—as evinced by the rules of retinal image formation, the structure of the visual environment, and the behavioral goals of the observer. For visual motion, these constraints lead to the following operational principles for cue interactions, which suggest, in turn, specific testable hypotheses.

(1) Motion Detection

An observer’s window on the world is the pattern of light cast upon the retinal surface, i.e., the “retinal image.” Motion of the observer, or motion of objects in the observer’s environment, leads to changes in the retinal image, or retinal image motion. But retinal image motion is not sensed directly. Detection of such motion is secondary to the detection of some type of spatial contrast in the pattern of retinal illumination, such as that caused by changes in the intensity, pattern, or spectral composition of light. To perform optimally in a variable environment, a motion detector should generalize across spatial contrast cues, thus encoding the motion of an image feature regardless of the cue that enables it to be seen. For example, Figure 1A illustrates that a thrown ball has the same motion regardless of its surface coloration or patterning; the ideal motion-sensing system would thus give the same signal about motion regardless of the cues that distinguish the ball from the background.

(2) Motion Interpretation

We perceive the motions of objects, not the motions of retinal image features. It is only through retinal image features, however, that we can infer object motion, and that inference is indirect. Specifically, images of our world are formed on the retinal surface by optical projection, which by dimensional reduction precludes a unique relationship between any single retinal image feature and the object that gave rise to it. To construct a neuronal representation of object motion, the visual system must overcome this ambiguity, which generally can be

*To whom correspondence should be addressed (e-mail: tom@salk.edu).

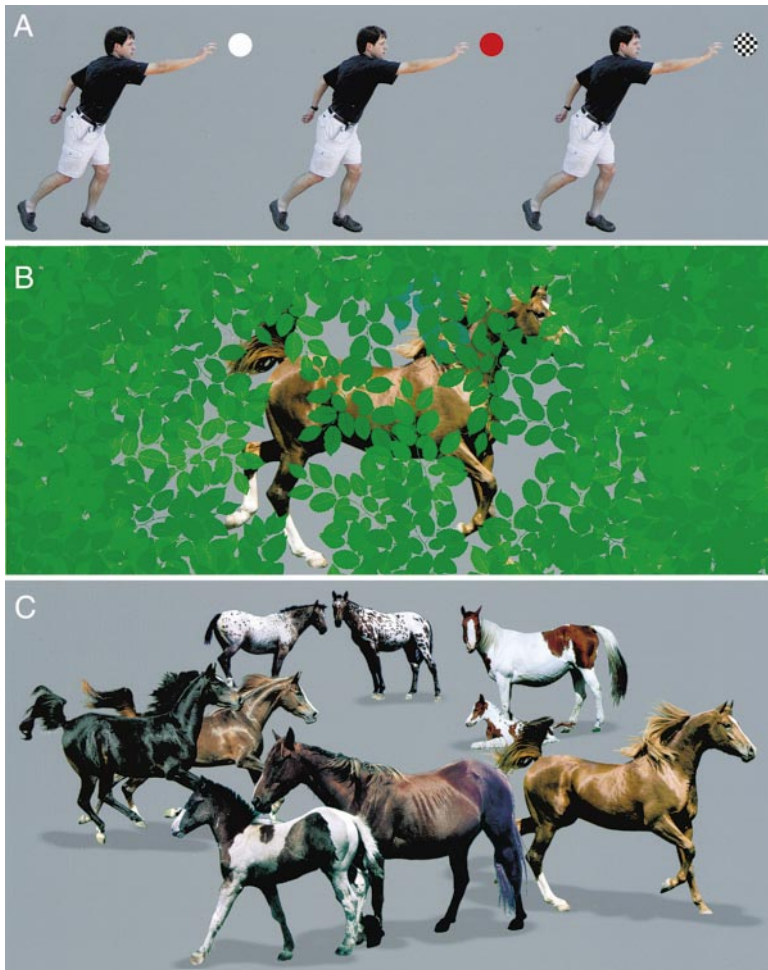


Figure 1. Examples Illustrating Three General Principles that Govern How the Primate Visual System Integrates Motion with Other Visual Cues

(A) Motion detection is secondary to the extraction of spatial contrast in the pattern of retinal illumination. In these three identical examples of a ball being thrown, the visual system must detect spatial contrast on the basis of luminance (left), color (middle), or texture (right) before the motion of the ball can be computed. Neuronal motion detectors operating in a variable visual environment would perform best if they generalized across different spatial contrast cues.

(B) Assignment of retinal image features to particular objects in a scene depends on the interpretation of contextual cues. In this picture of a trotting horse partially obscured by foliage, perception of the animal and its motion requires that the visual system integrate information about regions of the horse segregated by foliage. Here, this integration depends largely on grouping of retinal image features (the separate brown regions) on the basis of color.

(C) Objects are distinguished and recognized by their visual uniqueness, which is often found in the conjunction of different image cues. In this example, individual horses are distinguished by combinations of colors, textures, and motions (here apparent only by shape).

done by evaluating contextual cues—i.e., local patterns of brightness, color, texture, etc.—in which each moving retinal image feature appears. For example, to evaluate the motion of the trotting horse shown in Figure 1B, a motion-sensing system must use visual context to determine that the regions segregated by foliage are parts of the same moving animal.

(3) *Selecting Targets for Action*

We distinguish and recognize objects, and select them as targets for action, on the basis of their visual uniqueness, which can often be found solely in the conjunction of image cues, such as color and motion. Neuronal representations of cues that arise from a common object should be bound together, thereby forming both a representation of the whole object and a neuronal signal to guide actions directed at a target. For example, if we wish to select among the horses shown in Figure 1C, neuronal signals about colors, textures, shapes, and motions must be combined to distinguish the individual animals.

Because they rely upon visual cues in different ways, these three principles are likely manifested as different kinds of interactions between the neuronal representations of distinct cue types. We will address these proposed interactions further below. First, we will briefly describe the anatomical and physiological evidence for modular processing.

Are Different Visual Cues Processed by Independent Pathways?

Considerable anatomical and physiological evidence supports the claim that different image cues—particularly color and motion—are processed independently in the primate visual system. These data and their implications have been thoroughly reviewed elsewhere (e.g., Livingstone and Hubel, 1988; Schiller and Logothetis, 1990; Merigan and Maunsell, 1993; Stoner and Albright, 1993; Dobkins and Albright, 1998); here, we summarize a few key points and introduce some relevant reservations.

Parallel processing for color and motion is thought to originate in the retina. In primates, two morphologically distinct classes of retinal ganglion cells, termed parasol and midget ganglion cells, project exclusively to two different sets of layers in the lateral geniculate nucleus (LGN) of the thalamus, known as the magnocellular (M) and parvocellular (P) layers. Cells of the M and P layers project, in turn, to two distinct sublayers of the primary visual cortex (Figure 2A). Physiological studies at each processing stage have shown that, beginning in the retina, neurons in the M and P pathways possess characteristic responses to stimulus features: M pathway neurons exhibit properties indicative of a role in motion processing; whereas P pathway neurons, by contrast, exhibit properties that suggest contributions to both color and form processing.

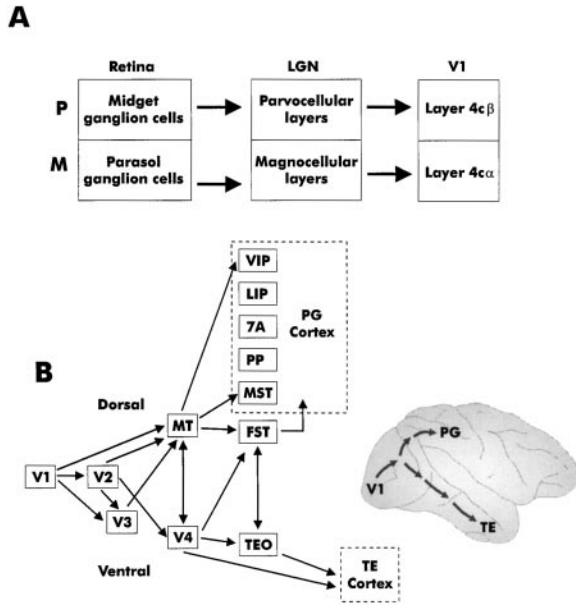


Figure 2. Diagrams Illustrating the Two Proposed Functional and Anatomical Dichotomies in the Primate Visual System

(A) The segregated M and P pathways begin in the retina with the parasol and midget ganglion cells, which project to the magnocellular (M) and parvocellular (P) layers, respectively, in the lateral geniculate nucleus (LGN) of the thalamus. From the LGN, M and P layer neurons project to two distinct sublayers in the primary visual cortex (V1).

(B) The partially segregated dorsal (“where”) and ventral (“what”) streams are apparent in projections from V1, which take either a dorsal route to parietal cortex or a ventral route to temporal cortex. Arrows indicate dominant feedforward projections between visual areas. V1 indicates primary, or striate, visual cortex. All other cortical areas indicated are extrastriate visual areas. Abbreviations: MT, middle temporal; VIP, ventral intraparietal; LIP, lateral intraparietal; PP, posterior parietal; MST, medial superior temporal; FST, fundus superior temporal; PG, inferior parietal cortex; TE, inferior temporal cortex. Based on Ungerleider and Haxby (1994).

A second avenue of research has focused on the connections and functional contributions of cortical visual areas. Anatomical studies have revealed two principal processing streams ascending from primary visual cortex (V1): a “dorsal stream” that extends from V1 into the parietal lobe, and a “ventral stream” that terminates in the temporal lobe (Figure 2B). One view, promoted by Ungerleider and Mishkin (1982) and expanded upon by Milner and Goodale (1995) and others, holds that the dorsal and ventral streams serve visual–spatial (“where”) and object recognition (“what”) functions, respectively.

Thus, two separate lines of research each led to proposed functional dichotomies—the first present at relatively early stages of visual processing and consisting of segregated color/form versus motion pathways, the second present at later (cortical) processing stages and manifested as separate representations of the “what” and “where” properties of objects. It was tempting to conclude that these two dichotomies were related; not surprisingly, a proposal of this sort by Livingstone and Hubel (1987) attracted considerable attention. The proposal was based, in part, upon two types of findings: (1) evidence for anatomical continuity between components of early- (M versus P) and late-stage (dorsal versus

ventral) dichotomies, and (2) physiological and neuropsychological evidence suggesting that the dorsal and ventral cortical streams are specialized—mirroring the early M and P pathways—for motion and color processing, respectively. Not only does this view offer an attractive conjunction of elementism and modularity, but it provides a seamless extension of M and P channels through the highest levels of visual cortex. This proposal has stimulated a plethora of studies attempting to uncover perceptual consequences of segregated color and motion processing. Despite such efforts, however, there is a striking lack of consensus in the field. We argue that this lack of consensus has arisen, at least in part, because the proposal is based upon faulty principles of visual processing, as described above. The constraints identified above, as well as the recent findings summarized below, suggest that the truth is more complex.

Neuronal Substrates for Motion Detection

Most studies of the interactions between visual cues have focused on the contributions of nonmotion cues (e.g., color) to motion processing. Hence, the discussion here will concentrate on neuronal populations that are specialized for motion processing and the routes by which nonmotion information may be accessible to those neurons. In primates, motion direction is first signaled by neurons in V1. It is, however, at a subsequent cortical stage, known as the middle temporal visual area (area MT), that motion processing comes into its own: ~90% of MT neurons exhibit pronounced selectivity for the direction and speed of a stimulus moved through the receptive field (Zeki, 1974; Van Essen et al., 1981; Maunsell and Van Essen, 1983; Albright, 1984; Albright et al., 1984). The responses of a typical MT neuron to moving stimuli are shown in Figure 3. Motion up and to the right evoked the largest responses from this neuron and would be referred to as the neuron’s “preferred” direction; motion in other directions either evoked smaller responses or suppressed the neuron’s spontaneous (unstimulated) activity. Different MT neurons have different preferred directions, and the range of preferred directions spans 360° without bias.

As diagramed in Figure 2B, MT receives direct cortical input from areas V1, V2, and V3 and also has connections with areas thought to lie at the same level of the visual hierarchy, such as area V4. MT neurons integrate information over a larger region of visual space than do their cortical afferents; on average, the receptive field diameter of an MT neuron is equal to its distance from the center of gaze (Albright and Desimone, 1987). While the layout of receptive fields across MT preserves the neighbor relations of visual space, a much larger portion of MT is devoted to processing information arising from the center, rather than the periphery, of the visual field (Gattass and Gross, 1981; Van Essen et al., 1981; Albright and Desimone, 1987). MT neurons project to higher cortical areas that appear to encode more complex forms of motion.

A variety of findings have supported a role for MT in the analysis and perception of motion. As described above, the vast majority of MT neurons respond selectively to the direction and speed of a moving stimulus.

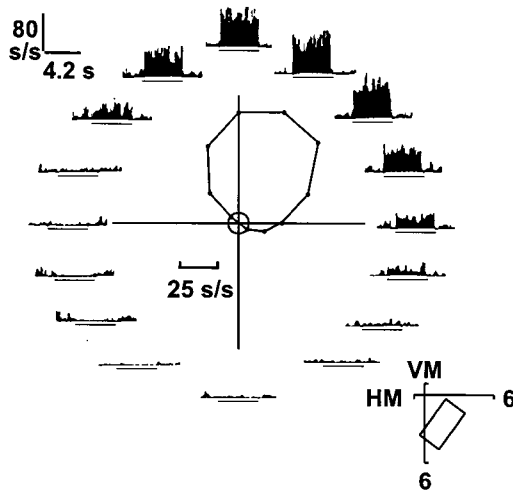


Figure 3. Direction Tuning of a Typical MT Neuron

The visual stimulus was a textured pattern moved continuously through the receptive field of the neuron (shown at lower right) for 4.2 s. Individual histograms represent responses summed over five trials to each of 16 directions of motion. The line beneath each histogram indicates the period of time during which the moving stimulus was presented. At center, the response for each direction is plotted on a polar graph. The radial axis represents response (spikes per second), the polar axis represents direction of stimulus motion, and the small circle represents level of spontaneous activity. Marked suppression of activity in the direction 180° away from the preferred direction is characteristic of many MT neurons. Abbreviations: VM, vertical meridian; HM, horizontal meridian. Modified from Albright (1984) with permission from the American Physiological Society.

In addition, MT is organized such that motion direction is represented in a columnar fashion, with all directions represented for each location in visual space (Albright et al., 1984; Albright and Desimone, 1987). The importance of MT to motion perception is further supported by the fact that its damage in either human or animal observers severely compromises motion perception and may even lead to motion blindness (Zihl et al., 1983; Newsome and Paré, 1988). Finally, electrical microstimulation alters perceived motion in a predictable manner (Salzman et al., 1990, 1992). The full range of MT response characteristics, anatomical organization, and suggested contributions to perception are reviewed elsewhere (Albright, 1993).

MT's dominant role in motion processing would seem to place it solidly in the so-called M pathway through visual cortex. As such, MT is an ideal cortical location for investigating the contributions of nonmotion cues to motion processing in terms of the three principles outlined above. In the remainder of this review, we will address how each of these principles is manifested in the response properties of MT neurons.

Principle 1: Motion Detection is Largely Invariant Across Spatial Contrast Cues

Most retinal image cues, such as luminance (a measure related to brightness), chrominance (related to hue), and texture, arise from spatial variations in the light cast upon the retinae. Motion, of course, takes place over time as well as space and is thus a secondary cue that depends upon some primary form of spatial definition. Under most circumstances, that spatial definition is the

result of multiple coincident cues—the light reflected from a journal, for example, may be brighter, redder, and less textured than that reflected from the desk on which it lies—but the primate visual system can detect spatial contrast arising from any one of these cues alone. For them to be broadly utilitarian, the responses of neurons sensitive to motion must likewise be independent of the particular spatial cues present in the image.

Luminance and Chrominance

The sensitivity of the primate visual system to the motion of patterns defined by luminance contrast with the background is well established, both at perceptual and neuronal levels (reviewed by Albright, 1993; Dobkins and Albright, 1998). By contrast, much has been made of the fact that moving stimuli defined solely by chromatic contrast (such as a pattern of red and green stripes of the same luminance) sometimes appear to move more slowly than their luminance-defined counterparts (e.g., Cavanagh et al., 1984; Livingstone and Hubel, 1987). This result has been hailed as evidence that chromatic contrast does not serve as a reliable spatial cue for motion processing, thus indicating a lack of interaction between the neuronal representations of color and motion. The direction of motion of chrominance-defined stimuli is generally perceived veridically, however, particularly if the stimuli are designed to elicit sufficiently contrasting responses from different cone photoreceptor populations (Dobkins and Albright, 1993; Palmer et al., 1993; Hawken et al., 1994; Stromeyer et al., 1995).

Neurophysiological studies of the contribution of chromatic cues to motion processing have focused largely on responses in area MT. In seeming support of MT's proposed position in the "color blind" M pathway, early studies found that MT neurons have little selectivity for either the chrominance or form of moving stimuli (Zeki, 1974; Baker et al., 1981; Maunsell and Van Essen, 1983; Albright, 1984). However, more recent studies have shown that some MT neurons continue to signal the direction of moving patterns defined only by chromatic contrast (Saito et al., 1989; Gegenfurtner et al., 1994). Thus, MT apparently receives chromatic information that, according to the segregated pathways hypothesis, should be available only to the P stream.

To determine whether this chromatic information reaches MT via the P pathway, Dobkins and Albright (1993, 1994) studied whether motion is processed in such a way as to strongly preserve the chromatic identity of moving features. Specifically, if the motion of a border defined by chromatic contrast is visible only when the colors on either side of the border stay the same, this would support P pathway input to MT. However, if a chromatic border is seen to move even when the colors change, then M pathway input suffices (see below). To study this, Dobkins and Albright used visual stimuli known as chromatic "sinusoidal gratings" (Figure 4) that could be configured to convey two opposing motion signals simultaneously—one direction visible if motion correspondence (the pairing of two spatially and temporally disparate stimuli to form a motion cue) preserves chromatic identity, and the opposite direction visible if motion correspondence ignores chromatic identity (Figure 4C). The perceptual reports of human observers revealed that the motion system generally disregards chromatic identity; i.e., it utilizes image features defined by chromatic contrast, independent of the specific colors involved.

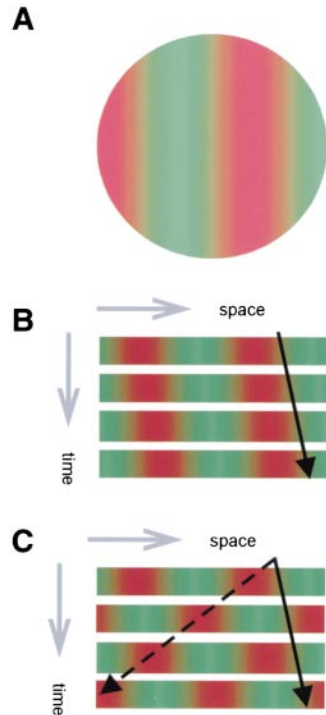


Figure 4. Schematic Diagram of the Stimuli Used to Characterize How Motion Correspondence Utilizes Chromatic Borders

(A) Stimuli were striped red/green patterns produced by varying chromaticity sinusoidally across space (chromatic “sinusoidal gratings”). The circular viewing aperture remained stationary, while the grating moved perpendicular to the stripes.
 (B and C) Space versus time plots show the chromatic profiles across a moving grating at different temporal frames.
 (B) Conventional moving grating. Rightward motion is detectable from correspondence pairs formed between the nearest chromatic borders (solid arrow), which preserve color identity.
 (C) Grating undergoing contrast sign reversal with each temporal frame. Motion of nearest chromatic cues involves a border that does not preserve chromatic identity moving rightward (solid arrow), while motion of a border that preserves chromatic identity is leftward (dashed arrow). When the red and green stripes had the same luminance, motion was generally perceived in the direction of the nearest chromatic border, even when color identity across the border changed.

When individual MT neurons in monkeys were presented with the stimuli diagramed in Figure 4, the pattern of responses mirrored those of human observers (Dobkins and Albright, 1994). Dobkins and Albright argued that these results could be accounted for by signals traveling to MT via the M layers of the LGN. Specifically, the observed lack of sensitivity to chromatic identity is readily explained by the fact that many M cells encode the presence of chromatic contrast but not its sign (Schiller and Colby, 1983; Lee et al., 1988, 1989a, 1989b, 1989c; Logothetis et al., 1990). Thus, signals traveling through the M layers of the LGN and extending to area MT provide a potential route by which chromatic information can influence detection of motion.

While these findings confirm a contribution of chromatic cues to motion processing and suggest underlying mechanisms, recent debate has centered on the strength of sensitivity to chromatic contrast. To address this issue, Thiele et al. (1999) adopted an established paradigm (Cavanagh and Anstis, 1991) that enables one

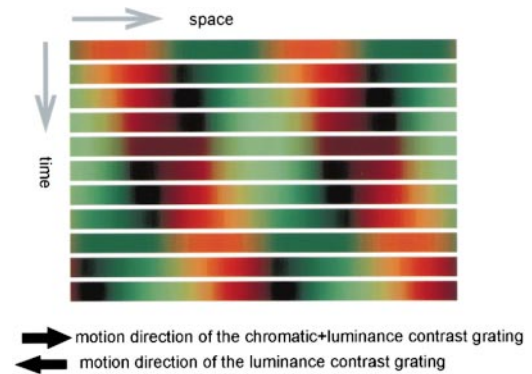


Figure 5. Space versus Time Plot of the Opposed Motion Stimulus Used to Measure the “Equivalent Luminance Contrast” of a Chromatic Signal Influencing Motion Processing in MT

Stimuli were composed of two sinusoidal gratings, one a combined chromatic + luminance contrast grating and the other a pure luminance contrast grating, superimposed and moving in opposite directions. Each stimulus appeared in an aperture overlying the receptive field of an MT neuron. Each horizontal slice in the figure shows the chromatic and luminance profiles across the stimulus for a single stimulus frame. In this example, the chromatic + luminance grating moves rightward, while the luminance grating moves leftward as time progresses. Modified from Thiele et al. (1999) with permission from the Society for Neuroscience.

to calibrate the sensitivity of motion detectors to chrominance-defined stimuli, relative to sensitivity to a luminance-defined standard. The stimulus used consisted of two superimposed sinusoidal gratings moving in opposite directions (Figure 5). One grating was defined solely by luminance contrast (achromatic), while the other had both chromatic contrast (alternating stripes were red and green) and varying amounts of luminance contrast (red stripes could be brighter than green or vice versa). In general, the neuronal responses elicited by the achromatic and chromatic gratings moving simultaneously in opposite directions were dominated by the more salient of the two gratings. For example, if the stimulus consisted of a salient achromatic grating moving in a neuron’s preferred direction and a weak chromatic grating (with low luminance contrast) moving in the nonpreferred direction, then the neuron’s response was relatively large. If, on the other hand, the preferred-direction achromatic grating was less salient than the nonpreferred-direction chromatic grating (with high luminance contrast), then the response was relatively small. By systematically varying the luminance contrast of the chromatic grating, it was possible to determine the contrast value for which responses to the two oppositely moving components were equivalent. When chromatic and achromatic inputs were thus balanced, the difference in their luminance contrasts provided a precise measure of strength of sensitivity to the chromatically defined stimulus, in units of luminance contrast.

The measure obtained, termed “equivalent luminance contrast,” revealed that chromatic contrast has a powerful influence over the responses of MT neurons when luminance contrast in the moving stimulus is low (<5%–8% contrast). Increasing the luminance contrast of the moving stimulus, however, subordinated the effects of the chromatic component. Thus, the chromatic properties of an object convey important information

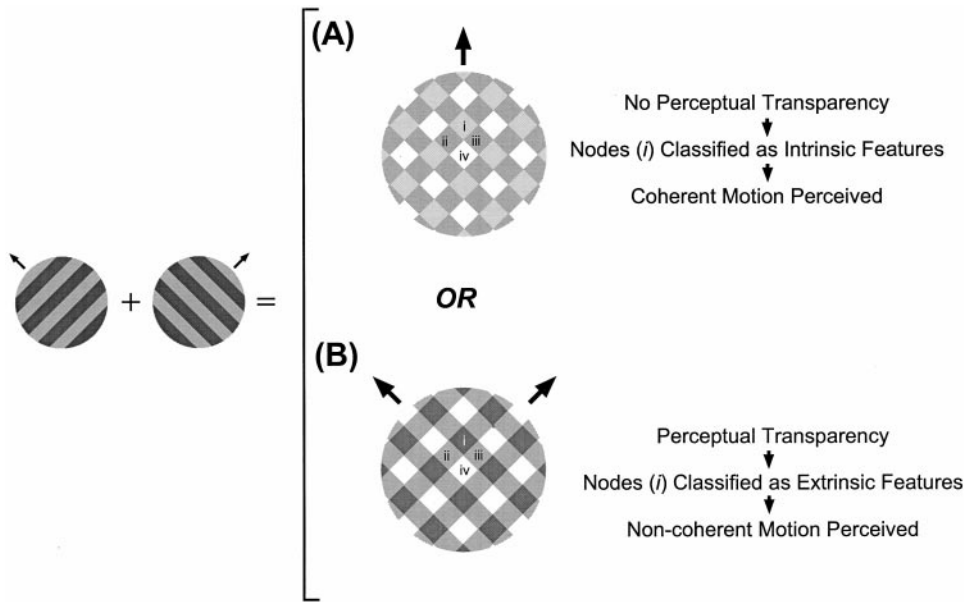


Figure 6. Plaid Patterns Were Produced by Spatial Superimposition of Two Light/Dark Striped Patterns (“Squarewave Gratings”)

The resulting images (A and B) were composed of four repeating subregions, identified as i, ii, iii, and iv. Drawing upon the rules of light mixture in the formation of the retinal image, one can identify specific luminance relationships between these four subregions consistent with transparent or opaque occlusion of one grating by the other (Metelli, 1974; Beck et al., 1984; Stoner et al., 1990). When these luminance relationships were physically inconsistent with transparent or opaque occlusion (A), human observers perceived the components to be at the same depth and to make up a single surface. By contrast, when the luminance relationships mimicked the retinal images formed by transparent or opaque occlusion (B), depth ordering of the gratings was generally seen (Stoner et al., 1990; Stoner and Albright, 1996). Under the latter conditions (B), one of the plaid subregions (i) appeared as regions of overlap. These “nodes” were thus extrinsic features, because they appeared to result from an interaction between two distinct surfaces. When the same nodes were not configured to be consistent with occlusion (A), they were intrinsic features seen to result from variations in the reflectance of a single surface. Stoner, Albright, and colleagues hypothesized that, although the nodes moved identically in both cases, they were differentially utilized by the motion system depending upon their feature classification. Specifically, intrinsic features were utilized for motion interpretation, and extrinsic features were disregarded. Perceived motion of plaid patterns (indicated by arrows in [A] and [B]) upheld this interpretation (Stoner et al., 1990; Stoner and Albright, 1996).

about motion direction to MT neurons only when insufficient luminance contrast exists; otherwise, motion of luminance-defined components drives the neuronal responses. In further support of the proposal of Dobkins and Albright (1994) summarized above, these results can be accounted for by simple pooling of the responses of LGN M cells (Thiele et al., 1999).

Other Cues

Although chromatic cues have received the greatest attention, evidence indicates that motion direction can also be perceived in a qualitatively invariant manner across a range of other contrast cues, such as spatial texture, temporal texture (flicker), and binocular disparity (e.g., Julesz and Payne, 1968; Ramachandran et al., 1973; Sperling, 1976; Petersik et al., 1978; Anstis, 1980; Lelkens and Koenderink, 1984; Chubb and Sperling, 1988; Stoner and Albright, 1992b; Lu and Sperling, 1995). To explore the neural basis of this perceptual invariance, the responses of directionally selective MT neurons have been compared using moving stimuli defined by luminance, spatial texture, and temporal texture. Many neurons exhibited a form-cue-invariant response, in which the preferred direction of motion remained constant across cue conditions (Albright, 1992; O’Keefe and Movshon, 1998). The fact that such neurons reliably encode direction despite marked changes in the way moving stimuli are spatially defined suggests that they

may contribute to the corresponding perceptual invariance (Albright, 1992; Stoner and Albright, 1993).

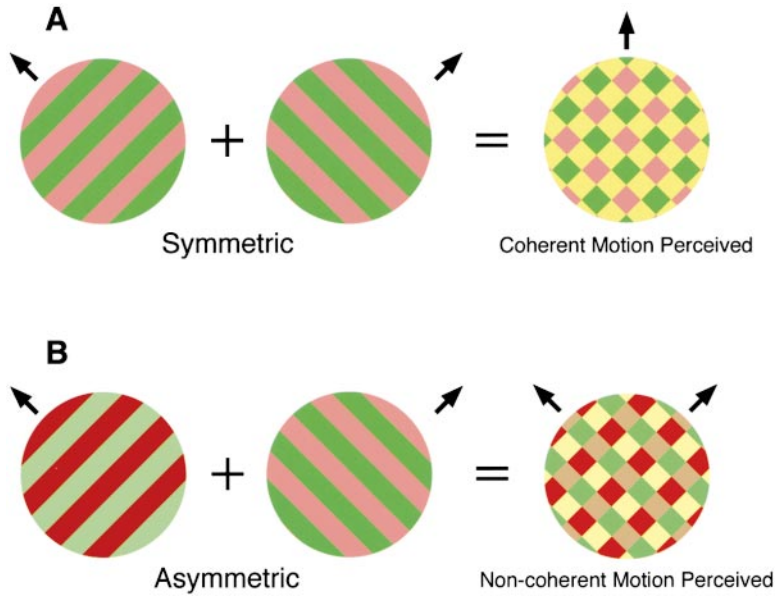
Summary

The studies described above show that the primate motion-processing subsystem receives and utilizes information about multiple spatial cues that can be used for motion detection. These findings raise many questions, of course, about the routes through which such cues reach motion detectors. As regards the contribution of chromatic cues, the existing data suggest that their influences can be accounted for by simple pooling of responses arising from the so-called M pathway.

In the next section, we summarize experiments that reveal a very different type of contribution from nonmotion cues to motion processing—one in which nonmotion cues, such as chromatic contrast, form the context in which moving features appear and thereby influence the interpretation of motion rather than its detection.

Principle 2: Motion Interpretation—Separating the Wheat from the Chaff in Moving Retinal Images

We perceive the motions of objects, yet it is only the motions of retinal image features that are directly accessible to the visual system. Because information about object identity is lost in the formation of the retinal image, the problem of recovering object motion from retinal



chromatic surfaces, and thus more likely to elicit a percept of chromatic transparency. Transparency should lead, in turn, to classification of plaid nodes as extrinsic features and a percept of noncoherent motion. Conversely, Dobkins et al. hypothesized that the symmetric configuration was more likely to support a percept of nontransparency, intrinsic classification of the plaid nodes, and a percept of coherent motion. See also Figure 6.

Figure 7. Illustration of a Manipulation of Chromatic Plaids that Had a Dramatic Effect on Perceptual Motion Coherence

Plaids were constructed using gratings possessing both chromatic and luminance contrast (red/green, R/G; bright/dark, +/−). Shown are example component gratings and their intersections. These chromatic plaids were of two basic types: “symmetric” plaids (A) were constructed from two gratings with identical spatial relationships between chrominance and luminance (e.g., both R+/G−). “Asymmetric” plaids (B) were constructed from two gratings with reciprocal spatial relationships between chrominance and luminance (R+/G− and R−/G+). When moved, symmetric plaids were generally perceived as a single coherently moving surface (A), whereas asymmetric plaids appeared as two gratings moving across one another (B) (Kooi et al., 1992; Dobkins et al., 1998). Dobkins et al. hypothesized that an asymmetric, rather than symmetric, relationship between chrominance and luminance contrast was more consistent with transparent occlusion by

image motion never presents a unique solution. Some solutions are better than others, however, and contextual cues enable two important steps in the recovery of object motion: feature classification and feature grouping. Consider, for example, the dynamic retinal images caused by two overlapping objects moving past one another. The problems produced by this commonplace situation are those of (1) classifying moving image features according to whether they arise from individual objects or from coincidental overlap of objects (such as the moving corners located at points of object overlap) and (2) grouping moving image features that arise from the same object.

Feature Classification

For purposes of this discussion, we can identify two types of retinal image features (Shimojo et al., 1989): those that reflect an “intrinsic” property of an environmental surface, such as the corner of a sheet of paper, and those that are “extrinsic” to any object or surface, arising from the intersecting retinal projections of distinct objects, such as the corner formed at the junction of two overlapping sheets of paper. Because they are causally linked to individual objects, the motions of intrinsic image features can provide reliable information about object motion. The motions of extrinsic features, by contrast, offer little information about object motion. Veridical interpretation of object motion is thus secondary to feature classification. In turn, intrinsic and extrinsic features in the retinal image can only be distinguished and classified using contextual cues for surface depth ordering.

Plaid patterns such as those in Figure 6 have been employed extensively to study contextual influences on motion processing (reviewed by Albright, 1993; Stoner and Albright, 1994). In one series of experiments, Stoner, Albright, and colleagues (Stoner et al., 1990; Stoner and

Albright, 1992a, 1993, 1996, 1998) noted that the pivotal image features in moving plaid patterns are “nodes” formed by the intersections of the two component gratings. Depending upon context, these nodes appear as either (1) intrinsic to the plaid, in which case their surface properties (luminance or chrominance) are seen to result from a variation in surface coloration, or (2) extrinsic, in which case their surface properties are seen as manifestations of opaque or transparent occlusion. Stoner and colleagues predicted that perceived motion would, in turn, follow directly from this feature classification, such that the plaid would be perceived as a single coherently moving pattern (“motion coherence”) when the nodes were classified as intrinsic (Figure 6A). Conversely, the plaid should appear as two gratings sliding independently across one another (“motion noncoherence”) when the nodes were classified as extrinsic (Figure 6B). Reports of human observers supported these predictions (Stoner et al., 1990; Stoner and Albright, 1996, 1998), thus demonstrating a marked influence of context on the interpretation of visual motion. In parallel neurophysiological studies, Stoner and Albright (1992a) located a cortical site in monkeys (area MT) at which these contextual influences were manifested as changes in sensitivity to visual motion. Specifically, when presented with moving plaid patterns in which contextual cues for depth ordering and feature classification were either present or absent, the directional responses of individual MT neurons were modulated in a manner that paralleled perceived motion.

Although much of the work to date on feature classification and motion processing has involved manipulations of luminance contrast, chromatic cues have also been found to make an important contribution to these phenomena. The general paradigm used is one in which the grating components of a plaid each have chromatic

contrast (Figure 7). A number of observations have been made using such stimuli (Krauskopf and Farell, 1990; Kooi et al., 1992; Dobkins et al., 1998; reviewed by Albright, 1991; Dobkins and Albright, 1998). For present purposes, the key findings are as follows: (1) if both gratings possess both chromatic and luminance contrast, and they are “symmetric” (e.g., red is brighter than green in both gratings, or vice versa), coherent motion is more likely to be perceived (Figure 7A); and (2) if both gratings possess both chromatic and luminance contrast, and they are “asymmetric” (i.e., red is brighter than green in one grating, and green is brighter than red in the other), noncoherent motion is more likely to be perceived (Figure 7B). Collectively, these observations reveal that it is not the simple presence of chromatic or luminance contrast that determines motion coherence; rather, coherence follows from the spatial relationship between the two types of contrast modulation. The significance of this finding lies in the possibility that an asymmetric luminance–chrominance relationship between the two component gratings may be a cue for transparent occlusion (Dobkins et al., 1998).

Dobkins et al. (1998) also used chromatic plaids to examine the sensitivity of MT neurons to chromatic cues for feature classification. The neurons’ directional responses were found to change in accordance with changes in the stimulus configuration, in a manner that closely mirrored the perceptual phenomenon. The results of these experiments are consistent with the proposal of Stoner and Albright (1993) that the neural substrates of motion coherence are sensitive to a wide range of contextual cues for depth ordering and feature classification.

These contributions of chromatic cues to feature classification and motion interpretation raise questions about the origin and nature of the chromatic signals. A first-order analysis appears to rule out signals of M pathway origin: such neurons are highly sensitive to luminance contrast (Kaplan and Shapley, 1986), and their responses can be assumed to be saturated for the high luminance contrasts in these chromatic plaid stimuli. There should thus be no additional modulation of M pathway neurons due to the chromatic contrast present in the plaids. We favor an alternative in which the P pathway conveys relevant information about the relationship between chrominance and luminance for each component of the plaid. Additional experiments are needed to evaluate this idea.

Feature Grouping

The plaid studies outlined above identified some conditions under which visual context influences the interpretation of features, such as edges and corners. We now turn to the role context plays in grouping features that belong to a particular object. As an example, consider the task of tracing the path of an acquaintance moving through a dense crowd of people. The problem facing the visual system in this case is that of pooling retinal motions according to object of origin. Efforts to obtain the motion “signals” unique to the moving acquaintance are confounded, in practice, by various types of image “noise,” caused by the presence of occluding objects, camouflaging shadows and reflections, and false leads given by people of similar size and shape. Determination

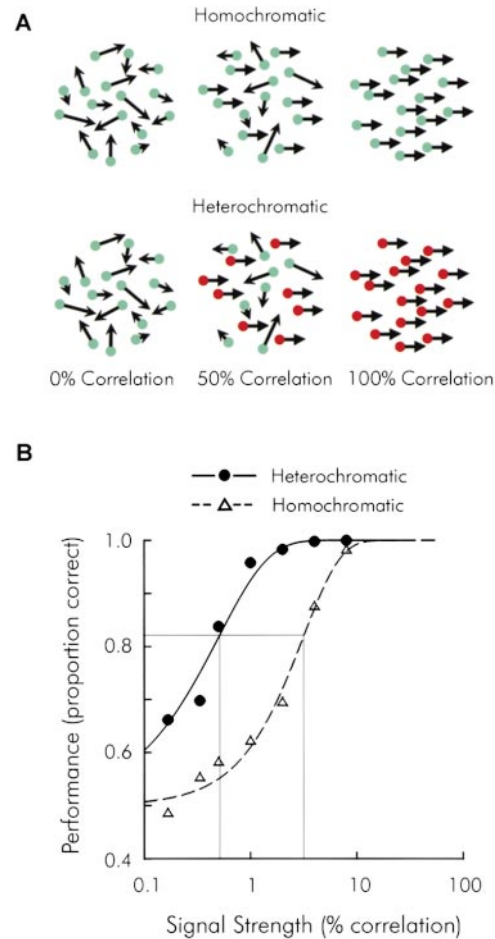


Figure 8. Discerning the Direction of a Motion Signal Is Easier if the Features Carrying Motion Signal and Noise Differ in Color

(A) Schematic diagram of the motion stimuli used by Croner and Albright (1997, 1999). Each stimulus consisted of a sequence of frames of randomly positioned bright dots appearing against a dark background on a CRT screen. Dots in each of the six circular apertures of the figure represent dots in six different stimuli. Arrows indicate velocity (direction and speed). The proportion of dots moving in the same direction at the same speed, expressed as a percentage and termed the “correlation,” describes the strength of the motion signal. At 0% correlation, all of the dots move randomly. At 50% correlation, half the dots have the same velocity. At 100% correlation, all of the dots have the same velocity. In the “homochromatic” condition, all of the dots have the same color. In the “heterochromatic” condition, the signal dots are a different color from the noise dots. Reprinted from Croner and Albright (1999) with permission from the Society for Neuroscience.

(B) Performance functions measured for a representative subject discriminating signal direction in homochromatic (open triangles) and heterochromatic (closed circles) stimuli. Also shown are curves fit to the homochromatic (dashed lines) and heterochromatic (solid lines) data. Threshold performance (0.82) is illustrated by a thin horizontal line. Where this line intersects each psychometric function, a thin vertical line is drawn to intersect the x axis at the threshold correlation for the function. Threshold correlation for the homochromatic condition was 3.11% correlation, and for the heterochromatic condition was 0.49% correlation, indicating a ~6-fold decrease in threshold when signal and noise dots were distinguished by color. Modified from Croner and Albright (1997).

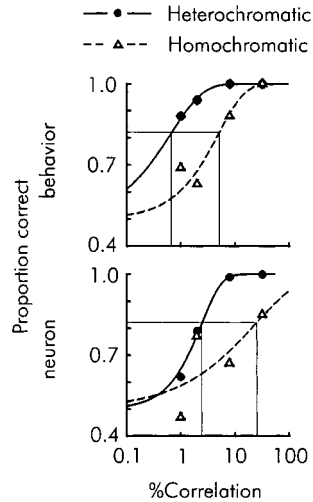


Figure 9. Simultaneously Obtained Behavioral and Neuronal Performance Functions Obtained in One Experiment Using the Random Dot Stimuli

Behavioral functions are shown at the top, and the corresponding neuronal functions obtained at the same time are shown below. Homochromatic, open triangles and dashed lines; heterochromatic, closed circles and solid lines. A large (~10-fold), statistically significant decrease in the behavioral heterochromatic threshold (thresholds: homochromatic = 5.0%, heterochromatic = 0.7%) was accompanied by a large (~10-fold), statistically significant decrease in the neuronal heterochromatic threshold (thresholds: homochromatic = 23.8%, heterochromatic = 2.4%). Reprinted from Croner and Albright (1999) with permission from the Society for Neuroscience.

of object motion under these conditions can be improved if motion signal and noise also differ along another sensory dimension, such as color. In this case, retinal motions can be grouped on the basis of color, allowing pooling of only the appropriate motions to enhance the desired motion signal. Thus, it should be easier to track our acquaintance were she to wear a distinctive red coat. Color, in this case, provides a context within which motion is more easily resolved.

Croner and Albright (1997) developed a visual stimulus that captures the essential features of this real-world example. The stimulus (Figure 8A) was based on one used widely in recent studies of motion processing (e.g., Williams and Sekuler, 1984; Newsome and Paré, 1988; Newsome et al., 1989; Britten et al., 1992) and consisted of a random array of small bright dots moving against a dark background. A varying proportion of the dots moved in the same direction and at the same speed, and thus constituted a motion signal, while the rest moved randomly and constituted motion noise. In the conventional configuration, all of the dots had the same color (Figure 8A, "homochromatic"). As shown by others (e.g., Newsome and Paré, 1988), when subjects reported the perceived direction of motion in these stimuli, their performance improved as the signal strength (the proportion of dots moving in a correlated fashion) increased. To investigate how grouping influences motion processing, Croner and Albright (1997) simply made the signal and noise dots differ along some other dimension (the contextual dimension), such as chromaticity (Figure 8A,

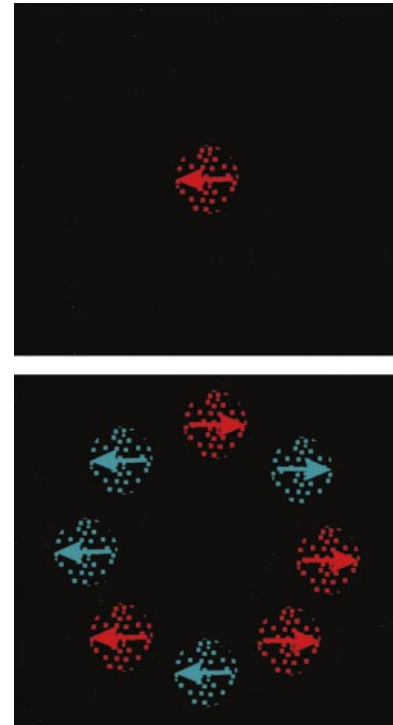


Figure 10. Schematic Diagram of Stimuli Used to Study Visual Search for Cue Conjunctions

During each trial, the subject first viewed a stimulus (top panel) that indicated the target to find in the subsequent search array (bottom panel). Arrows show motion direction in each patch of colored dots. In this example trial, the subject would search for a patch of red dots moving to the left. Based on Buračas and Albright (1999) with permission from the American Psychological Association.

"heterochromatic"). As expected, if the motion signal and noise were so distinguished, the observers' sensitivity to direction markedly improved (Figure 8B). These results demonstrated that motion signals can be selectively pooled—and their detection thus facilitated—using scene structure formed on the basis of an unrelated cue.

To examine the neural basis of this kind of cue interaction, Croner and Albright (1999) trained monkeys to report the perceived direction of motion of stimuli such as those diagramed in Figure 8A. While the monkeys performed the direction discrimination task, the activity of individual directionally selective MT neurons was recorded. Approximately 20% of the MT neurons studied showed a significant improvement in their ability to discriminate direction when the signal and noise dots were different colors. An example is shown in Figure 9. Thus, a subpopulation of MT neurons is strongly influenced by chromatic scene structure, just as are observers' perceptual judgments.

What kind of input is required to mediate this chromatic contribution to MT responses? As for the chromatic plaid studies described above, the dots in the stimuli diagramed in Figure 8A were of high luminance contrast relative to background. Since the responses of M pathway neurons saturate under these high-contrast conditions, it is unlikely that the relevant input comes

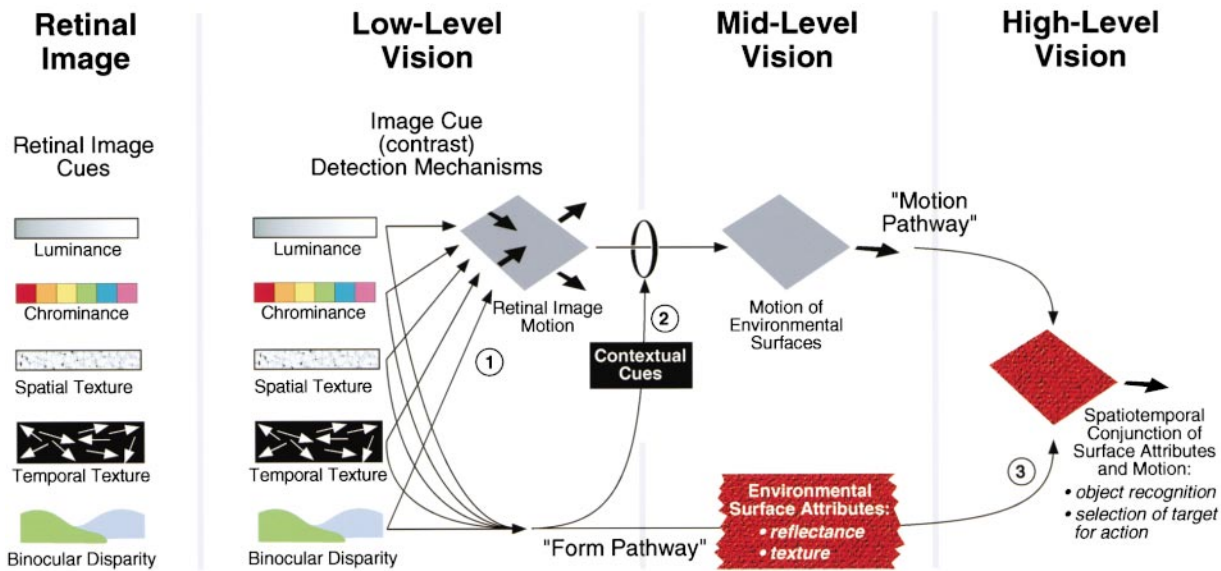


Figure 11. Schematic Illustration of Proposed Interactions between Different Sources of Image Information

The retinal image (left) contains a variety of visual cues. These cues are detected by specialized mechanisms at an early stage of visual processing (low-level vision), ultimately yielding a representation of retinal image features. Because motion detection is secondary to the detection of spatial contrast in the image, primary motion detectors receive convergent input from neurons representing spatial cues, such as luminance and chrominance. This convergence is the implementation of "principle 1" proposed herein. The goal of mid-level vision is to construct a neuronal representation of environmental features, such as the surface properties (e.g., reflectance) and motions of objects. Achievement of that goal requires that image features be categorized and grouped on the basis of contextual information in the image, which is the implementation of "principle 2" proposed herein. Note that properties that are physically independent in the environment (e.g., motion and surface reflectance) are independently represented, despite the necessity for cross-talk between the representations of different image cues. Finally, the goal of high-level vision is to link the independent representations of environmental features, such as the path of motion and the surface reflectance of an object, to enable object recognition and the execution of movements (e.g., reaching) directed at specific targets. This is the solution to the classic binding problem and the implementation of "principle 3" proposed herein.

from residual chromatic sensitivity of neurons in this pathway. Rather, information such as that carried by the P pathway must be involved. Based on their findings, Crone and Albright (1997) proposed that contextual cues for feature grouping lead to adjustments of the relative gain of the sensory signals reaching or passing through area MT. These gain adjustments, accordingly, enable motion signals bearing a common property (e.g., the color red) to be processed selectively with minimal disruption by dynamic noise.

Summary

These examples of interactions between different visual cues in the service of motion interpretation offer a second type of qualification to the hypothesis that elemental visual dimensions are each processed independently by specialized cortical areas. In contrast with Principle 1 above, in which nonmotion spatial contrast cues are utilized for the simple detection of retinal image motion, Principle 2 asserts that such cues are used to segment the image into meaningful scene-based motions.

Next, we consider our third principle governing cue interactions—that involving the binding of visual cues to form representations of objects.

Principle 3: Object Recognition and Visual Target Selection—Identity Lies in the Conjunction of Visual Cues

The classic form of visual cue interaction, which has long been a focus of lively debate in circles ranging from

philosophy to cognitive psychology and more recently to neurobiology, is the so-called "binding problem." As commonly presented, the problem is that of linking together the independent neuronal representations of image features that arise from a single object. The conjunction of features, of course, provides identity to objects and is the key to their recognition and the selection of targets for action. That this binding problem is solved by the visual system is a given, which attests to additional interactions between the representations of visual cues. Important questions concern the means by which the solution is achieved. How is it, for example, that my visual system is able to link independent neuronal events corresponding to the color and the motion of a red balloon moving leftward, such that I recognize and direct appropriate motor output to that balloon?

A behavioral task known as "visual search" has been used for many years to probe this binding process. Subjects are presented with an array of complex visual stimuli and required to locate a remembered target in the array—a task that captures the essence of commonplace real-world situations, such as recognizing and reaching for the salt shaker on a densely arrayed dining table. In one simple version of this task (Figure 10), the target is defined by a unique (i.e., solitary) conjunction of color and motion cues—in Figure 10, a patch of red dots moving to the left—and "distractors" are comprised of other color/motion conjunctions (red-rightward, green-rightward, green-leftward). The time that it takes a

human or monkey subject to detect the target increases linearly with the number of distractors (Nakayama and Silverman, 1986; Buračas and Albright, 1999). Since target detection is a behavioral index of cue binding, this paradigm can be used to study perceptual and neurophysiological events surrounding the binding of different cues.

Manifestations of the target detection process, assessed behaviorally and neurophysiologically, are highly revealing of the interactions between cues. To begin with, the very fact that a unique conjunction of color and direction of motion can be detected by observers argues that the neuronal representation of motion is somehow linked with that for color. It has been suggested that such linking may be achieved by temporal synchrony of neuronal responses (von der Malsburg, 1981; Gray et al., 1989), but there currently exists little empirical support for a general form of that hypothesis. Buračas and Albright (1995, *Soc. Neurosci.*, abstract) found, alternatively, that neuronal responses to motion in area MT were modulated in amplitude upon behavioral detection of a target defined by a conjunction of color and motion. The timing of this response enhancement just prior to the behavioral report suggests that it may reflect the binding of color and motion cues via selection by focal attention, which leads in turn to target awareness, recognition, and the initiation of an action directed at the target.

Another process, known as "guided search," further refines our understanding of cue interactions that give rise to object recognition and target-directed actions (Egeth et al., 1984; Wolfe et al., 1989). The guided search hypothesis argues that search for a conjunction of cues proceeds selectively amongst those objects that have a cue in common with the desired target. For example, this idea predicts that an observer searching for the red object moving leftward in Figure 10 will begin by evaluating the motions of red objects only (i.e., to the exclusion of green objects). Behavioral data from humans (Egeth et al., 1984; Wolfe et al., 1989) and nonhuman primates (Motter and Belky, 1998; Buračas and Albright, 1999) support this idea, suggesting that one cue (color in the previous example) provides a context for selective modulation of neurons representing another cue (e.g., motion), even prior to object recognition. Preliminary results indicate that the responses of MT neurons are consistent with this view (Buračas and Albright, submitted).

Little is known of the mechanisms responsible for these effects, nor of the source of the contextual signals for feature binding and object recognition. Their existence nonetheless documents a third form of functional interaction between the neuronal representations of different visual cues, which is likely to be a major focus of future research in this area.

A Reconciliation of Modular and Interactive Processes?

Is it possible to reconcile the reigning segregationist view of cortical visual processing with these and other data demonstrating functional interactions between different sensory cues? We believe that the solution lies in a refinement of our understanding of which processes are segregated.

Clearly, the purpose of vision is to guide behavior. To that end, there is considerable benefit afforded by independent internal representations of object properties that enjoy physical independence from one another in the real world. The motions and surface colors of an object (which are truly independent object properties) are thus prime candidates for modularity. Confusion results, however, when one fails to distinguish between object properties and retinal image properties. The key insight is that in order to construct modular representations of object color and motion, the visual system must generate context-dependent interpretations of retinal image features. By definition, each image feature's context consists of other image features, defined by any visual cue, nearby in time and space. Along the cortical path to an independent representation of object motions, it should thus not surprise us to find the footprints of other retinal image properties, such as chrominance.

Resolution can therefore be achieved simply by distinguishing between the physical properties of objects and the retinal image features cast by those objects. At the early stages of visual processing, retinal image features may be processed in relative isolation from each other, depending upon the cues that define them. At later processing stages, these initially segregated signals must be integrated (by and for the various means and purposes outlined above) in order to achieve representations of the object features that gave rise to them (Figure 11). Thus, the neuronal representation of object motion, which we believe occurs at least partially in MT, reflects the integration of signals initially carried separately by the M and P pathways.

We have proposed that MT neurons represent the motions of environmental features—objects—and constitute, in that sense, a motion module. Indeed, none of the data summarized herein counter the prevailing view that area MT represents motion selectively, with the key qualification that the motion represented is that of objects in the visual scene. But, as we have described, it is fundamentally impossible to achieve a representation of object motion without convergence of different kinds of cues appearing in the retinal image. With this perspective, we can more easily understand the discovery that neuronal responses in area MT reflect a convergence of information thought to be carried by the earlier M and P pathways.

Acknowledgments

We thank G. T. Buračas, G. R. Stoner, and A. Thiele for comments. Much of the work from our laboratory cited herein was supported by grants from the National Eye Institute, the National Institutes of Mental Health, the McKnight Foundation for Neuroscience, the Fight for Sight research division of Prevent Blindness America, the Human Frontiers Science Foundation, the Alfred P. Sloan Foundation, and the McDonnell-Pew Center for Cognitive Neuroscience at San Diego. T. D. A. is an Investigator of the Howard Hughes Medical Institute.

References

- Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol.* 52, 1106–1130.
- Albright, T.D. (1991). Color and the integration of motion signals. *Trends Neurosci.* 14, 266–269.

- Albright, T.D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science* 255, 1141–1143.
- Albright, T.D. (1993). Cortical processing of visual motion. *Rev. Oculomot. Res.* 5, 177–201.
- Albright, T.D., and Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Exp. Brain Res.* 65, 582–592.
- Albright, T.D., Desimone, R., and Gross, C.G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* 51, 16–31.
- Anstis, S.M. (1980). The perception of apparent movement. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 153–168.
- Baker, J.F., Petersen, S.E., Newsome, W.T., and Allman, J.M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*): a quantitative comparison of medial, dorsomedial, dorsolateral and middle temporal areas. *J. Neurophysiol.* 45, 387–405.
- Beck, J., Prazdny, K., and Ivry, R. (1984). The perception of transparency with achromatic colors. *Percept. Psychophys.* 35, 407–422.
- Boring, E.G. (1942). *Sensation and Perception in the History of Experimental Psychology* (New York: D. Appleton-Century Company).
- Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765.
- Buračas, G.T., and Albright, T.D. (1999). Covert visual search: a comparison of performance by humans and macaques. *Behav. Neurosci.* 113, 451–464.
- Cavanagh, P., and Anstis, S. (1991). The contribution of color to motion in normal and color-deficient observers. *Vision Res.* 31, 2109–2148.
- Cavanagh, P., Tyler, C.W., and Favareau, O.E. (1984). Perceived velocity of moving chromatic gratings. *J. Opt. Soc. Am. (A)* 1, 893–899.
- Chubb, C., and Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. (A)* 5, 1986–2007.
- Croner, L.J., and Albright, T.D. (1997). Image segmentation enhances discrimination of motion in visual noise. *Vision Res.* 37, 1415–1427.
- Croner, L.J., and Albright, T.D. (1999). Segmentation by color influences responses of motion-sensitive neurons in the cortical middle temporal visual area. *J. Neurosci.* 19, 3935–3951.
- Dobkins, K.R., and Albright, T.D. (1993). Color, luminance, and the detection of visual motion. *Curr. Dir. Psychol. Sci.* 2, 189–193.
- Dobkins, K.R., and Albright, T.D. (1994). What happens if it changes color when it moves? The nature of chromatic input to macaque visual area MT. *J. Neurosci.* 14, 4854–4870.
- Dobkins, K.R., and Albright, T.D. (1998). The influence of chromatic information on visual motion processing in the primate visual system. In *High-Level Motion Processing—Computational, Neurobiological, and Psychophysical Perspectives*, T. Watanabe, ed. (Cambridge, MA: MIT Press), pp. 53–94.
- Dobkins, K.R., Stoner, G.R., and Albright, T.D. (1998). Perceptual, oculomotor, and neural responses to moving color plaids. *Perception* 27, 681–709.
- Egeth, H.E., Virzi, R.A., and Garbart, H. (1984). Searching for conjunctively defined targets. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 32–39.
- Gattass R., and Gross, C.G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 621–638.
- Gegenfurtner, K.R., Kiper, D.C., Beusmans, J.M., Carandini, M., Zaidi, Q., and Movshon, J.A. (1994). Chromatic properties of neurons in macaque MT. *Vis. Neurosci.* 11, 455–466.
- Gray, C.M., Konig, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337.
- Hawken, M.J., Gegenfurtner, K.R., and Tang, C. (1994). Contrast dependence of colour and luminance motion mechanisms in human vision. *Nature* 367, 268–270.
- Julesz, B., and Payne, R.A. (1968). Differences between monocular and binocular stroboscopic movement perception. *Vision Res.* 8, 433–444.
- Kaplan, E., and Shapley, R.M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc. Natl. Acad. Sci. USA* 83, 2755–2757.
- Koffka, K. (1935). *Principles of Gestalt Psychology* (New York: Harcourt, Brace, and Company).
- Kooi, F.L., De Valois, K.K., Switkes, E., and Grosop, D.H. (1992). Higher-order factors influencing the perception of sliding and coherence of a plaid. *Perception* 21, 583–598.
- Krauskopf, J., and Farell, B. (1990). Influence of colour on the perception of coherent motion. *Nature* 348, 328–331.
- Lee, B.B., Martin, P.R., and Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J. Physiol.* 404, 323–347.
- Lee, B.B., Martin, P.R., and Valberg, A. (1989a). Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *J. Physiol.* 414, 223–243.
- Lee, B.B., Martin, P.R., and Valberg, A. (1989b). Amplitude and phase of response of macaque retinal ganglion cells to flickering stimuli. *J. Physiol.* 414, 245–263.
- Lee, B.B., Martin, P.R., and Valberg, A. (1989c). Nonlinear summation of M- and L-cone inputs to phasic retinal ganglion cells of the macaque. *J. Neurosci.* 9, 1433–1442.
- Lelkens, A.M., and Koenderink, J.J. (1984). Illusory motion in visual displays. *Vision Res.* 24, 1083–1090.
- Livingstone, M.S., and Hubel, D.H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J. Neurosci.* 7, 3416–3468.
- Livingstone, M.S., and Hubel, D.H. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749.
- Logothetis, N.K., Schiller, P.H., Charles, E.R., and Hurlbert, A.C. (1990). Perceptual deficits and the activity of the color-opponent and broad-band pathways at isoluminance. *Science* 247, 214–217.
- Lu, Z.L., and Sperling, G. (1995). Attention-generated apparent motion. *Nature* 377, 237–239.
- Maunsell, J.H., and Van Essen, D.C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* 49, 1127–1147.
- Merigan, W.H., and Maunsell, J.H. (1993). How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* 16, 369–402.
- Metelli, F. (1974). The perception of transparency. *Sci. Am.* 230, 90–98.
- Milner, A.D., and Goodale, M.A. (1995). *The Visual Brain in Action* (New York: Oxford University Press)
- Motter, B.C., and Belky, E.J. (1998). The guidance of eye movements during active visual search. *Vision Res.* 38, 1805–1815.
- Nakayama, K., and Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature* 320, 264–265.
- Newsome, W.T., and Pare, E.B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* 8, 2201–2211.
- Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.
- O’Keefe, L.P., and Movshon, J.A. (1998). Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Vis. Neurosci.* 15, 305–317.
- Palmer, J., Mobley, L.A., and Teller, D.Y. (1993). Motion at isoluminance: discrimination/detection ratios and the summation of luminance and chromatic signals. *J. Opt. Soc. Am. (A)* 10, 1353–1362.
- Petersik, J.T., Hicks, K.I., and Pantle, A.J. (1978). Apparent movement of successively generated subjective figures. *Perception* 7, 371–383.
- Ramachandran, V.S., Rao, V.M., and Vidyasagar, T.R. (1973). Apparent movement with subjective contours. *Vision Res.* 13, 1399–1401.
- Saito, H., Tanaka, K., Isono, H., Yasuda, M., and Mikami, A. (1989).

- Directionally selective response of cells in the middle temporal area (MT) of the macaque monkey to the movement of equiluminous opponent color stimuli. *Exp. Brain Res.* 75, 1–14.
- Salzman, C.D., Britten, K.H., and Newsome, W.T. (1990). Cortical microstimulation influences perceptual judgments of motion direction. *Nature* 346, 174–177.
- Salzman, C.D., Murasugi, C.M., Britten, K.H., and Newsome, W.T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* 12, 2331–2355.
- Schiller, P.H., and Colby, C.L. (1983). The responses of single cells in the lateral geniculate nucleus of the rhesus monkey to color and luminance contrast. *Vision Res.* 23, 1631–1641.
- Schiller, P.H., and Logothetis, N.K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends Neurosci.* 13, 392–398.
- Shimojo, S., Silverman, G.H., and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Res.* 29, 619–626.
- Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behav. Res. Methods Instr.* 8, 144–151.
- Stoner, G.R., and Albright, T.D. (1992a). Neural correlates of perceptual motion coherence. *Nature* 358, 412–414.
- Stoner, G.R., and Albright, T.D. (1992b). Motion coherency rules are form-cue invariant. *Vision Res.* 32, 465–475.
- Stoner, G.R., and Albright, T.D. (1993). Image segmentation cues in motion processing: implications for modularity in vision. *J. Cogn. Neurosci.* 5, 129–149.
- Stoner, G.R., and Albright, T.D. (1994). Visual motion integration: a neurophysiological and psychophysical perspective. In *Visual Detection of Motion*, A.T. Smith and R.J. Snowden, eds. (London: Academic Press), pp. 253–290.
- Stoner, G.R., and Albright, T.D. (1996). The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res.* 36, 1291–1310.
- Stoner, G.R., and Albright, T.D. (1998). Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. *Vision Res.* 38, 387–401.
- Stoner, G.R., Albright, T.D., and Ramachandran, V.S. (1990). Transparency and coherence in human motion perception. *Nature* 344, 153–155.
- Stromeyer III, C.F., Kronauer, R.E., Ryu, A., Chaparro, A., and Eskew, R.T., Jr. (1995). Contributions of human long-wave and middle-wave cones to motion detection. *J. Physiol. (Lond.)* 485, 221–243.
- Thiele, A., Dobkins, K.R., and Albright, T.D. (1999). The contribution of color to motion processing in macaque area MT. *J. Neurosci.*, in press.
- Ungerleider, L.G., and Haxby, J.V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165.
- Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In *Analysis of Visual Behavior*, D.J. Ingle, M.S. Goodale, and R.J.W. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549–586.
- Van Essen, D.C., Maunsell, J.H.R., and Bixby, J.L. (1981). The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic connections. *J. Comp. Neurol.* 199, 293–326.
- von der Malsburg, C. (1981). The correlation theory of brain function. Internal report, Max-Planck-Institute of Biophysical Chemistry, Göttingen, West Germany.
- Williams, D.W., and Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Res.* 24, 55–62.
- Wolfe, J.M., Cave, K.R., and Franzel, S.L. (1989). Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433.
- Zeki, S.M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol.* 236, 549–573.
- Zihl, J., Von Cramon, D., and Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain* 106, 313–340.