Mechanisms of stereoscopic processing: stereoaattention and surface perception in depth reconstruction

Christopher W Tyler, Leonid L Kontsevich
Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, CA 94115, USA
Based on paper presented at the Conference in Binocular Stereopsis and Optic Flow, Toronto, Canada, 22–26 June 1993

Abstract. Consideration of the range of phenomena from studies of human stereopsis suggests that a five-stage model is required to provide a complete account of the processes involved, within which any stereoaattention mechanism must operate. The information from the disparity field of the optical projections to the two eyes (stage 1) goes to a set of parallel Keplerian arrays of disparity detectors, each array selective for a different spatiotemporal property of the visual images (stage 2). Global interactions produce a cyclopean depth image that is cleaned of the spurious ghost images in the Keplerian arrays (stage 3) and that may then be processed for its (hypercyclopean) form elements (stage 4). Finally, there must be a stage of integration of the stereoscopic depth cues with monocular and kinesthetic depth cues to form the overall map of perceived distance (stage 5).

The fact that multiple cyclopean surfaces may be perceived as transparent implies that the stereoscopic system is not limited by a singular-surface constraint. However, it is unclear whether multiple surfaces can be seen simultaneously or whether only one surface is seen at a time by a selective-attention process, with the others perceived as a purely inchoate (qualitative) depth impression. New experiments on cueing of ambiguous stereocorrections by singular flat planes suggest that selective stereoaattention is a powerful mechanism. In fact, the results show that attention can be focused not just in horopteral planes but in a variety of depth configurations. Moreover, this attention focus may act as a tracking mechanism to allow perception of smooth cyclopean stereomotion, which has a frequency response up to \( \sim 5 \text{ Hz} \) (in contrast to the \( \sim 15 \text{ Hz} \) limit for detecting planar disparity shifts as jerky appearance and disappearance effects). Finally, the spatial limits of stereosurface reconstruction are explored with cyclopean targets to show some interesting asymmetries of the surface-wrapping process that may represent object-oriented constraints on depth reconstruction.

1 Overview of stages of stereoscopic processing

The current view of the visual cortex is that it consists of many separate visual representation areas, each organized along both a serial-hierarchical principle and a parallel-distributed principle (van Essen and Zeki 1978; van Essen and Maunsell 1983; Livingstone and Hubel 1988). The hierarchical principle is the organization of each neural representation into layers, with the inputs from each layer converging onto the next. The fact that this hierarchical convergence of information is happening not once but in each of a large number of visual representation areas conforms to a parallel-distributed principle on a large scale. If each visual representation has equivalent access to the attentional selectivity of the perception process, the combined network should be considered a heterarchy with distributed control properties, in which each network node can dominate the signal to form the processes perceptual output; no one representation area would be at the head of the processing stream as there would be in a signal-processing hierarchy.

There also is a medium-scale parallel organization of neural processing within each visual representation area, in that the hypercolumns process local areas of the visual field in parallel. Finally, there is a small-scale parallel organization of neurons into separate columns making up each hypercolumn, with each column processing a different attribute of the region of the visual image served by the hypercolumn.
This overlapping organization of parallel hierarchies, which may be termed a multiscale serial–parallel heterarchy, appears to be a ubiquitous principle of neural organization of sensory analysis.

Although stereopsis is generally modeled as a single unified processing system, there is significant evidence (Tyler 1983) that it has a number of independent sub-systems, each specialized for extracting depth in a different manner. Stereoscopic processing therefore also should be viewed as having a heterarchical organization, even within the domain of depth perception from binocular disparity. There are many other monocular depth cues that can be processed in parallel to generate dramatic depth impressions (such as differential-motion cues, perspective, shape-from-shading, blur, interposition, and so on) but the present analysis is focused specifically on binocular disparity cues.

Analysis of the range of stereoscopic phenomena from studies of human perception suggests that a five-stage model is required to provide a complete account of the processes involved (figure 1). The information from the disparity field of the optical projections to the two eyes (stage 1) goes to a set of parallel Keplerian arrays of disparity detectors, each array selective for a different spatiotemporal property of the visual images (stage 2). Global interactions produce a cyclopean depth image cleaned of the spurious ghost images in the Keplerian arrays (stage 3); this depth image may then be processed for its (hypercyclopean) form elements (stage 4). Finally, there must be a stage of integration of the stereoscopic depth cues with monocular and kinesthetic depth cues to form the overall perceived-distance map (stage 5). For all its complexity, this model nevertheless is based on the simplifying assumptions that there are no interactions or feedback processes between module arrays. Although this may be unrealistic, given the multiplicity of interconnections that have been mapped within the primate brain (eg van Essen and Maunsell 1983), it is important to establish the first-order structure of the network before addressing its second-order interactions.

![Diagram of stereoscopic depth perception](https://example.com/diagram.png)

**Figure 1.** A model framework for stereoscopic depth perception in the form of a multistage serial–parallel heterarchy. (a) The binocular optical transform of spatial information; (b) Keplerian arrays of neural disparity processing; (c) global cleaning of the refined depth image; (d) hypercyclopean processing of the cyclopean forms; and (e) integration with nonbinocular depth cues to form the spatial map.
1.1 Binocular optics: the first stage of stereoscopic processing

The first operation that the human visual system performs is the geometric projection that condenses the three-dimensional (3-D) world down to the planar images projected on the two retinae. This operation imposes a basic structure for which the cortex must compensate before the information can be processed into its 3-D object relations. The binocular projection geometry was first worked out by Helmholtz (1866) and is elaborated in Tyler (1991a), for example.

Figure 2a (left panel) shows the basic geometric construction of the zero-disparity horopter, consisting of two lines (dashed)—a vertical horopter line on the midline at the convergence distance and a circle (or conic section) that lies at an angle close to horizontal. Fixation at any point away from this construction results in a projection with vertical disparities on the two retinae, and is consequently off the horopter. The horopter retains this configuration even when the eyes are moved to any point within it (figure 2a, right panel). Figure 2b (left panel) shows how the regions of binocular fusion (stipple) spread around the base construction of the empirical horopter, in which measurements show that the vertical horopter line is inclined backward by an amount that varies with fixation distance and the conic section has a wider radius of

![Diagram](image_url)

**Figure 2.** (a) The geometric construction of the point horopter (dashed lines) of positions in space that are in strict binocular correspondence for fixation straight ahead. Left panel: the point horopter consists of a circle and a line; any point away from the circle and line projects with a vertical disparity and therefore is no longer in correspondence. Dotted lines through the pupils show foveal projections; full lines are construction lines to particular points in space. Right panel: the geometric point horopter remains unchanged for fixation at any point on the horopter. (b) The fusion horopter, or region of points in space that are perceived as binocularly fused around the empirical point horopter. Left panel: for fixation straight ahead; right panel: for oblique fixation up and to the right. Note the complexity that the binocular optical transform imposes on the depth-reconstruction task.
curvature than the base circle. Breitmeyer et al (1976) provided the first stereoscopic evidence for the backward inclination of the horopter (Helmholtz had relied on evidence of the difference in perceived monocular tilts). The fusion region will vary in thickness around the horopter according to the distance from the fovea. Figure 2b (right panel) shows how the horopter lines and the fusion 'sausage' are modified when fixation is moved away from the base horopter construction.

The horopter represents the zero positions for the binocular-disparity metric, relative to which all other disparities must be scaled to generate the depth map. This structure and its variation with fixation position must be built in and redressed by the depth-reconstruction mechanism if a valid and reliable depth map is to be constructed. How the brain achieves this compensatory calibration is unknown, but it is interesting to recognize the complexity of the problem that needs to be solved.

1.2 Hierarchical disparity-encoding model

In a review of the evidence for the variety of mechanisms contributing to stereopsis, it was concluded (Tyler 1983) that it is not possible to accommodate all the diverse results within any single hierarchical model. Depth perception has been demonstrated for diverse stimulus types that are incompatible with any single process for solving the binocular-correspondence problem. The organization of the neural components of the model has recently been described in detail with supporting evidence (Tyler 1993). Only the main features of the neural model therefore will be summarized here.

A multilevel local approach has the advantage of breaking down the global nature of stereopsis (Julesz 1978) into a sequence of processes [(b)–(d) in figure 1] that are empirically distinguishable. Within this serial structure, a further parallel organization is required for analyzing specialized global features of the stereoscopic image.

1.3 Disparity-selective local mechanisms

The point of departure of the neural model is a parallel (Keplerian) array of cortical disparity 'detectors', each responding to the presence of a stimulus with a particular location in $x$, $y$, $z$ coordinates [represented by the rectangles in (b) in figure 1, with the $y$-axis omitted for clarity]. The key feature of this approach is a reliance on arrays of specialized mechanisms of binocular matching to account for a variety of attributes of the stereoscopic process. Such mechanisms are represented by the sketched rectangles extending behind the first order in (b) in figure 1. Examples of such attributes that will be discussed are direction selectivity for motion in depth and selectivity for orientation of the elements making up random-element stereograms. Such selectivities are not predicted by inhibition or facilitation between simple disparity-selective neurons. Each specialized disparity array is assumed to operate in parallel with other such arrays at the same processing level, with the overall output being determined either by a combination of the processes or by the most sensitive process under particular stimulus conditions.

Specialized disparity mechanisms may themselves exist in two classes according to whether specialization occurs at the local level (retinal-receptive-field characteristics) or global level (cortical interactions). Examples of local mechanisms that have been discussed are those specialized for the detection of stimuli with specific sizes and orientations, orientational disparities, motion in depth, or spatial frequency differences between the eyes.

As depicted in (b) in figure 1, the initial stage of binocular combination is a disparity-detection process that might be achieved by the various types of neuron with facilitatory responses to different disparities, as recorded in the visual cortex of cat and monkey (Barlow et al 1967; Poggio et al 1985). This stage may be considered as a local cross-correlation process, performed by neurons tuned to different disparities, occurring at each location in the binocular visual field. As the eyes vary their
vergence, cortical projections of the visual scene slide over one another to obtain the shift (or disparity) that produces the best match or correlation in each local region of the visual field. The images matched in this way are said to be in register. In practice there are, of course, two dimensions of field location, and the array may not be as regular as depicted here.

1.4 Mechanisms of binocular combination

The specialized mechanisms that have evolved in the human brain to accommodate the complexity of the binocular reconstruction process have been categorized (Tyler 1993) into three classes (figure 3). The first consisted of dichoptic mechanisms for reconciling the essentially disjunctive information falling at corresponding points in the two eyes. Such dichoptic mechanisms included binocular fusion of similar images in adjacent dichoptic regions, interocular suppression and rivalry of strongly discrepant, high-contrast images at corresponding locations, binocular luster of areas with different local illumination levels, and the newly discovered process of linear binocular summation of discrepant images of low contrast or high spatial frequency. Note that fusion, rivalry, luster, and summation are incompatible percepts that each require different underlying mechanisms for their operation.

The next class was static stereomechanisms for the processing of horizontal disparities, which has been proposed to consist of fine and coarse spatial mechanisms. In addition, there is evidence for specialized processing of orientational disparities between elongated receptive fields in the two eyes, which encode vertical inclinations in depth, and also for spatial frequency differences between the two eyes, which encode lateral tilts in depth. It should be noted that these latter two attributes are encoded by orthogonal properties of the arrays of binocular receptive fields with different orientation or size preferences in the two eyes, which therefore constitute independent modules for these two processes.

A final class of binocular mechanisms was that of dynamic stereomechanisms specialized for processing motion in depth. One reason for regarding these as separate mechanisms is that the receptive fields in the two eyes cannot logically be direction selective both for purely lateral motion and for purely depth motion. One requires the receptive fields to be selected for the same direction of motion in the two eyes, while the other requires opposite selectivities. There must therefore be a

![Figure 3. Types of binocular mechanism. Each row shows diagrams of a set of independent mechanisms of binocular combination in the form of its canonical receptive fields in the two eyes.](image-url)
distinct class of dynamic stereomechanisms. Evidence has shown this class to include separate mechanisms with different properties for achromatic and chromatic depth motions, again as reviewed in Tyler (1993).

1.5 Global interactions

The third cortical stage of the model is the site of the global interactions between the local disparity detectors that serve to refine the representation of the disparity image from its initial crude array of stimulated points to a coherent representation of the 3-D surfaces present in the field of view. A variety of such processes have been proposed by Julesz and others over the years, summarized in Julesz (1971, 1978) and Tyler (1983, 1991b).

Mechanisms in this class of global interactions use either conventional positional disparity information from the two retinae or specialized disparity information, but respond on the basis of some global aspect of the disparity field. Examples of such interactions are globality, cooperativity, and perceptual hysteresis.

Globality refers to the neural processes that restrict depth perception to a single perceived depth along any line of sight. Its operation is represented by the uniqueness constraint, which requires the ability to solve the binocular-correspondence problem, i.e., to identify a unified percept of a coherent depth image from the cloud of possible correspondences between all the dots projecting to one eye and all the dots projecting to the other eye.

Cooperativity is defined as any type of mutual interaction between the processing for different spatial regions of the stereo image. Such lateral cooperativity may also be involved in solving the correspondence problem effectively; it may well include such interactions as lateral inhibition between disparity detectors (figure 4), disparity-specific pooling or facilitation, the disparity-gradient limitation on the upper limit for depth reconstruction, coarse-to-fine matching processes for building up the depth image from the monocular information, and so on. These processes all may be conceived as taking place within the locus of global interactions depicted by (c) in figure 1, following the interocular-matching or disparity-detection stage but preceding the generation of a unified global depth image from the plethora of available disparity information.

Perceptual hysteresis is a difference in sensitivity for a change between two states according to the direction of the change. It may be termed catabolic hysteresis if the change is an order–disorder transition that occurs far more readily for increases in disorder, or entropy, in the stimulus than for the opposite transitions, decreases in entropy. An example for the neural binocular system occurs in the formation and

![Figure 4](image)

**Figure 4.** Cyclopean cooperativity in the form of the lateral-interaction field between disparity detectors derived from data on the spatial frequency tuning to cyclopean corrugation stimuli (from Tyler 1991b). For small vertical distances and somewhat larger horizontal distances there is spatial integration of disparity information (central peak). For larger vertical distances and even larger horizontal distances there is disparity inhibition (elongated moat).
dissolution of the fused depth image from the state of zero binocular correlation (figure 5). Experiments on the detection of interocular correlation show that it is easier to dissolve than to build up the perception of a fused global depth image (Julesz and Tyler 1976; Tyler and Julesz 1976, 1978).

![Graph showing correlation and decorrelation](image)

**Figure 5.** Demonstration of hysteresis in the detection of changes of binocular correlation. Detection of a binocular decorrelation event from the background of a correlated state (filled circles) is much easier than is detection of a binocular correlation from a decorrelated state (open circles), for all field sizes (from Tyler and Julesz 1978).

1.6 *Hyercyclopean perception*

The concept of hyercyclopean analysis refers to the fourth level of processing of stereoscopic images depicted by (d) in figure 1. By analogy with the cortical neurons with receptive fields selective for particular properties of the retinal image, there could be neurons at a higher level in cortex having 'receptive fields' at the level of the 'cleaned' cyclopean depth image. These receptive fields would have a cyclopean basis, in the sense of having properties specific to the disparity-selective neurons in the cyclopean retina, but would perform a hyercyclopean analysis of the spatial and temporal *form* of the depth image. Hypercyclopean receptive fields would have characteristics defined in terms of the figural properties of the cyclopean image, but independent of its specific disparity characteristics, ie which particular disparity is stimulated at any given retinal location.

The existence of such a hypercyclopean level of processing can be demonstrated by means of a stereograting adaptation paradigm in which the stereograting is moved continously across the retina, so as to avoid any stereoscopic depth afterimage. The obtained threshold elevation, which is specific to both spatial frequency and orientation of the adapting grating, therefore must be occurring at a higher level of form processing beyond that of the cyclopean processing for depth per se. Hypercyclopean specificity for adaptation to the spatial frequency content of the cyclopean image was demonstrated by Tyler (1975a) and Schumer and Ganz (1979), orientation specificity in a cyclopean tilt aftereffect by Tyler (1975a) and motion specificity in the form of a motion aftereffect to motion of the purely cyclopean depth image by Papert (1964).
1.7 The representation of visual space

Even beyond the hypercyclopean level there must be a further level of abstract spatial representation in which stereoscopic information is integrated with other spatial representations, such as those from the motion-vector field (Gibson 1950; Nakayama and Loomis 1974), the texture-gradient field (Gibson 1950), and accommodation, vergence, vestibular, and other nonstereoscopic cues. Moreover, this integrated spatial representation is a likely site for the operation of 'top–down' processes, where the spatial representation is configured to make sense in terms of object properties in the physical world. Stored knowledge of the object constraints operates to select between competing interpretations of the depth information up to this point, and to interdict aspects of the interpretations that are incompatible with object properties. This level of integration is indicated by (e) in figure 1, but is not considered in detail in the present analysis.

2 Transparency versus the 2½-D sketch

The focus in this paper is to analyze the mechanisms of scene reconstruction from the visual signal array arriving at the cortex from the two eyes, particularly the evaluation of the binocular information carried by those signals. One influential view of the process of scene reconstruction was that of Marr, summarized in his book on the subject. Marr's view was that a logical way to proceed, and perhaps a good model for the neural processing, was to use the available information to construct a surface map of the scene. Because most objects in the world are opaque, a surface map contains most of the information about the visible objects in the scene. An essential component of this map is that the configuration in the third dimension (distance from the observer) is captured as a depth token at every position on the segregated two-dimensional (2-D) surface map. Because the depth-token map is a condensed version of the full array of depth information, Marr and Nishihara (1978) somewhat jocularly called this combination of a 2-D surface map and the depth-token map the '2½-D sketch'.

Marr and colleagues viewed the 2-D surface map as being generated essentially by the local form information, which defined what would be seen at each point in the field and how it would be segregated into subregions with continuous properties divided by discontinuities in image properties. The depth-token map was considered to be derived both from stereoscopic depth cues and from monocular depth cues such as interposition, shape from shading, and so on.

The basic Marr model fails to capture two major features of scene reconstruction that may be observed in random-dot-stereogram images (Julesz 1971), which constitute valuable tools for the dissection of the neural processes of vision. The first is the fact that the features perceived from the 2-D surface map do not necessarily arise from the local form information alone. In the absence of the stereoscopic depth information, the random-dot array appears as an undifferentiated uniform surface. Unless the noise is unusually inhomogeneous, no clear features stand out. It is only after the depth is perceived that the segregation into separate regions defined by the depth borders is achievable. After this segregation, indeed, the local spatial noise may appear coarser in more distant regions and finer in closer regions owing to size scaling on the basis of the depth information. Thus, disparity information may, paradoxically, form the basis of the monocular segregation and consequently give rise to an apparent monocular depth cue of variations in perceived texture density. This sequence is in the opposite direction to the information flow postulated in the Marr model.

The second feature of random-dot stereograms is their ability to support the percept of transparent depth surfaces (Julesz 1971; Norcia and Tyler 1984). Here again, the depth tokens are assuming a primary role, for they first need to be specified at each
point in the image before the construct of a surface running through each appropriate set of points can be developed. It is as though the surface is strung across the depth tokens to segregate the relevant sets of monocular dots rather than the reverse. The visual system may be capable of supporting the simultaneous percept of up to three overlaid surfaces (Weinshall 1991) from fields of randomly intermixed dots. Such percepts make it difficult to maintain the perspective that stereopsis is derived after separation of the monocular images into locally homogeneous subregions, but instead emphasize the local depth tokens of each feature as the primary structure of visual 3-D space, with the surface superstructure erected upon their scaffolding.

![Figure 6. Depiction of the hyperglobal perception of two transparent surfaces in a random-dot stereogram.](image)

Hyperglobality is defined (Julesz 1978) as the ability to perceive multiple surfaces simultaneously in the same region of space (figure 6). This capability transcends the uniqueness constraint against such simultaneous perception of more than one surface that characterizes the simple-globality principle. The uniqueness constraint need not, however, be violated at any one spatial location; instead, the stereoscopic reconstruction process seems to be able to complete a perceived surface across neighboring elements attributable to each surface organization separately.

3 The singular constraint of stereoeattention
Before abandoning the Marr view that there is a single surface representation at any point in the field, it is important to be sure that there is no interpretation under which the single surface can remain the primary vehicle of reconstruction, even for perception of multiple transparent surfaces. One such view is that, although only a single surface may be reconstructed at any one moment in time, transparent perception may be obtained by sequential reconstruction of each of the multiple surfaces in turn. Marr and Poggio (1979) followed the approach of the cooperative Automap algorithm developed by Julesz (1963, 1971) in proposing such sequential reconstruction of depth surfaces. The idea is that surface reconstruction was achieved within a fixed Keplerian metric of disparity detectors by vergence shifts that moved the (otherwise rigid) stereoreconstructive apparatus to the appropriate regions of physical space to find the densest clusters in each disparity plane.

The hypothesis of sequential reconstruction by vergence eye movements makes two testable predictions. One is that the disparity range of the depth-reconstruction mechanism is, by postulate, limited to disparities near zero. Disparity images of flat planes near zero disparity therefore should be easier to detect than disparity images that cut through the zero-disparity plane at a steep angle. Such steep stereoidsimages should require a sequence of several vergence positions before they can be fully reconstructed. This prediction was tested by Uttal et al (1975), who generated planes
up to $80^\circ$ from frontoparallel in dynamic-noise stereograms and presented them in brief exposures too short for vergence eye movements to occur. Two-alternative forced-choice experiments (with a monocularly indistinguishable null target of random depth information) indicated that the detectability of such depth planes was almost independent of angle of tilt. This result makes it difficult to conceive how any model with a bias toward frontoparallel surface reconstruction can be operating in human vision.

A second feature of the eye-movement reconstruction approach is that it does not include a mechanism of attentional enhancement of surfaces projecting within the Keplerian array; the only local focusing mechanism is presumed to be that of vergence tracking through the 3-D optical image. Since attention is being shown to have a ubiquitous role throughout many domains of sensory analysis, it is natural to ask whether attentional enhancement operates within the mechanism of stereoscopic depth reconstruction? There are previous suggestions that attention may be subject to endogenous control through the third dimension (Downing and Pinker 1985; Gawriszewski et al 1987) and that it may be subject to exogenous control in search tasks involving separation of targets in the third dimension (Nakayama and Silverman 1986; Holliday and Braddick 1991).

Julesz and Chang (1976) approached the question of stereoscopic enhancement by a preceding disparity stimulus from the point of view of disparity pooling. They found that a plane presented briefly at one depth would bias the subsequent perception of a pair of ambiguous planes in the direction of the disparity of the initial plane. Although there may have been an attentional component to this biasing, Julesz and Chang explained the result as a disparity-pooling effect in the cooperative interactions in the global stereoscopic computation. On this account, the neural trace of the initial plane lingered to pool with the signal of the nearer of the two test planes, rendering it stronger than the signal from the other test plane. Competitive inhibition between the two test planes might then have operated to eliminate perception of the weaker signal.

The Julesz and Chang results also are consistent with an explanation in terms of rapid enhancement by the exogeneous attentional signal of the seed plane. The rapid time course of the seeding, while incompatible with the slower time course of endogenous or voluntary attention, is similar to that of the facilitation by exogenous or automatic attentional shifts (Nakayama and Mackeben 1989). However, the seeding does not provide strong evidence for a stereoscopic-attention mechanism because a cueing effect would be expected with an additive disparity-pooling mechanism of the non-cooperative type proposed by Richards (1971).

Our approach to establishing the role of stereoscopic attention was to present a modulated-disparity test signal on one of the two simultaneous test planes that would not be enhanced by any kind of disparity pooling. This modulation signal was a sinusoidal disparity corrugation that could appear on either of the two ambiguous planes in a random-dot stereogram on each trial, the other remaining flat. The corrugation itself could be in one of two phases (sine or $-\text{sine}$ relative to the fixation point that the observer was instructed to fixate) to form the forced-choice discrimination task. The attentional cue was a flat stereoplane presented for 105 ms prior to the 45 ms test stimulus (see figure 7a). This sequence amounted to a total time of 150 ms, which is less than the latency for any vergence eye movement (Rashbass and Westheimer 1961). Any disparity averaging or pooling between the flat priming plane and the test corrugation would tend to reduce rather than facilitate the perceived depth corrugation of the test, making discrimination of its phase more difficult. Only if there were attentional enhancement in the region of the test plane could performance improve.
Because we are attempting to drive attention through the third dimension with an external priming cue rather than by some voluntary behavior on the part of the observer, the experiment is designed to explore the properties of exogenous rather than endogenous attention. We therefore would expect the time course of any facilitation to the priming cue to be similar to that found in two-dimensional paradigms (e.g. Nakayama and Makepen 1989).

Figure 7. Demonstration of the operation of stereoscopic attention controlled by an exogenous cue plane, for two observers, LL and AC. (a) The task is detection of the phase of depth corrugations presented ambiguously for 45 ms in one of a pair of transparent planes. The cue plane was presented for 105 ms at one of five disparities spanning the test-plane disparities. (b) Without the cue (filled squares) the corrugations are indiscriminable. The two graphs show narrow disparity tuning of attentional enhancement by the cue plane when it is proximal in depth to the test corrugation.
In fact, when the corrugated test surface was away from the horopter, the experimental results (upper panel, figure 7b) showed dramatic improvement in the corrugation-phase discriminability when the priming plane was near to the disparity of the test surface. The test durations were set so that the corrugation-phase discriminability was close to chance in the absence of the priming plane. When the priming plane fell within about 10 arc min of the disparity of the corrugations, its phase became readily discriminable. Beyond that range, there was little enhancement above the baseline performance level. Because the priming plane contained no corrugations, it added no information to the discrimination task. Its effect, therefore, must have been due to a non-feature-specific enhancement of the information-processing capability in a limited disparity range, which may be attributed to the operation of disparity-specific attention.

When both the corrugated test surface and the cue plane were at the horopter (zero disparity; lower panel, figure 7b), no enhancement was observed from the moderate phase discriminability already present, but discriminability fell to lower values as the cue plane was shifted away from the horopter. This result is interpreted to imply that the attentional focus normally resided at or near the horopter locus, so no enhancement was obtained by attracting it to that position. Drawing the attentional focus to disparities away from the horopter, however, degraded performance back to chance levels, with a similar tuning to that obtained for off-horopteral test disparities.

There are two competing interpretations of the attentional enhancement shown in figure 7. One possibility is that it is compatible with the operation of fixed lateral facilitation between disparity detectors of the types proposed by either Julesz (1971), Tyler (1975b), or Marr and Poggio (1979). Although each type is different, they share in common the fixed spatial orientation of the facilitation, which is hard-wired to be maximal between disparity detectors sensitive to the same binocular disparities, i.e., parallel to the horopter plane. The alternative view is that the attentional enhancement is stimulus driven; specifically, the facilitation shown in figure 7 just happens to be revealed in the frontoparallel plane because that is where we tested for it. On this hypothesis, the enhancement would be equally evident whatever the 3-D spatial distribution of the priming stimulus. This concept of a flexible 3-D-attentional mechanism will be evaluated in the next section.

4 Wrapping the surface with the attentional shroud
The surfaces of the natural world typically do not form flat planes; even when they do, the planes are unlikely to fall conveniently in the singular position of the observer's frontoparallel plane. A stereoscopic-attentional process therefore would be much more valuable if it could be wrapped around the form of any spatial object, rather than being restricted to frontoparallel planes. The wrapping process could take one of two forms—it could be an exogenous attentive process driven by the local stereoscopic information immediately preceding the test stimulus at each point in the visual field, or it could be a cooperative endogenous process in the form of a flexible planar coupling within the Keplerian array. This latter process corresponds to what Julesz (1971) called "the search for dense surfaces".

A more vivid representation of this process is to think of it as an attentional shroud, wrapping the dense locus of activated disparity detectors as a cloth wraps a structured object. A depiction of such a wrapping has been provided by René Magritte (figure 8). This depiction shows how the cloth may envelop an object to capture the broad features of its shape, although some detail may be lost. Such a loss of detail is characteristic of the stereoscopic process. The graph in figure 9, reproduced from Tyler (1974), shows how stereoscopic depth reconstruction of corrugated surfaces is limited to a spatial bandwidth of only about 4 cycles deg⁻¹, as much as one log unit
less than the bandwidth for resolution of luminance information. Thus, stereoscopic reproduction is capable of rendering depth variations only to a coarse scale of representation, as though the connectivity of depth reconstruction were by a flexible material that was too stiff to match local discontinuities in the depth information.

To begin to explore the nature of the attentional mechanism, we conducted an experiment to test for the existence of attentional enhancement of depth detection in a situation where the priming stimulus was not a frontoparallel plane. For our experiment, the stereoscopic stimulus consisted of a plane tilted around the vertical axis from which two squares protruded in depth, 3 deg on each side of the fixation point. Each square could be with equal probability either in front of or behind the tilted plane by 2 arc min of disparity. The observer had to make a forced-choice judgment of whether the two squares had the same or opposite directions of disparity offset, which required simultaneous depth processing in both locations. If attention was focused on only one location, to the left or right of fixation, performance could be only 50% correct since there was random correlation between the disparity directions of the two cyclopean squares. Thus, attention was required in two different ranges of depth relative to the horopter in order to perform the task.

The attentional-priming stimulus was a random-dot plane at various tilts relative to the test plane. It was presented for 105 ms, and was immediately followed by the 45 ms test configuration, just as in the previous experiment. The question was whether presenting the priming plane at the same tilt as the test plane facilitated disparity discrimination relative to the absence of the priming stimulus, and whether there was a depth-orientation tuning to such facilitation. Again, any such priming effect would have to be operating at a minimum of two depth locations in order to enhance performance on the task of whether the disparity directions of the squares was the same or opposite. This task therefore tested for the first-order (tilt) component of stereoscopic-attentional control, as opposed to the first experiment that tested the zero-order or absolute-disparity selectivity of the mechanism.

The results for two values of test tilt, expressed in disparity-gradient ratio (Burt and Julesz 1980) are shown in figure 10b; each of them shows both strong attentional facilitation by the priming stimulus and selectivity for the primed depth orientation compared with for other nearby orientations. These results support the idea that

![Figure 8. Illustration of a shroud wrapping a face, from “L’histoire centrale” by René Magritte.](image)
stereoattention is not simply limited to frontoparallel planes, as would be expected if it were a hard-wired connectivity of lateral facilitation (Julesz and Johnson 1968; Tyler 1975; Marr and Poggio 1979). Instead, the task requires that stereoattention must be flexible enough to facilitate at least two locations at different disparities simultaneously. While the fully elaborated concept of an attentional shroud still needs broader evaluation, it has passed the first test of its flexibility in allowing facilitation of first-order depth tilts as opposed to frontoparallel (zero-order) depth planes.

Figure 9. (a) Two-panel stereogram depicting a sinusoidal depth corrugation increasing in amplitude to the left and in frequency upwards. Binocular fusion of this stereogram allows one to see the range of the stereospatial depth resolution. Note that the corrugation is invisible in the upper left sector of the stereogram, showing that the stereospatial resolution is limited to about 3 cycles deg\(^{-1}\). (b) Disparity threshold as a function of spatial frequency, with a sensitivity peak at about 0.4 cycles deg\(^{-1}\).
Figure 10. (a) Paradigm for demonstrating that stereotypes can wrap onto the first-order tilt of a plane in depth. At the top, a tilted test plane is shown with squares protruding in opposite directions on either side of fixation. Below this is a test plane with a different tilt from which the squares protrude in the same direction. On each trial, the observer had to estimate whether the disparities of the squares were the same or opposite in direction, to ensure that both were attended simultaneously. Three alternative tilts for the cue plane are also shown. At the bottom the cue and test sequence over time for this figure and figure 11, are shown. (b) Tuning of stereodynamics to depth tilt, expressed in units of disparity-gradient ratio (disparity gradient of cue plane × 16). Cue time was 105 ms. Clear enhancement was obtained for cue tilts near those of the test targets, relative to the lack of discriminability without a cue plane (filled squares).

5 Dynamics of stereodynamics

Having established some properties of the spatial control of 3-D-attentional facilitation by exogenous disparity priming, we now turn to the temporal dimension of this mechanism. In previous work, Norcia and Tyler (1985) measured the temporal limits for the perception of apparent motion between two planes alternating in a dynamic-noise stereodisplay. The maximum rate at which smooth motion could be seen was 5–6 Hz, although alternation between the planes was detectable up to about 14 Hz. It is therefore clear that the limit to smooth stereodynamics is set not by a temporal limitation on the processing of stereoscopic depth per se, but by some mechanism required for integrating or tracking the smooth motion path in depth. We would like to propose that this depth-tracking mechanism is not a separate system specialized for motion tracking, but the same attention mechanism that is demonstrated in the previous figures. This approach provides a parsimonious interpretation of the diverse data according to a single mechanism.

More generally, attentional tracking could form the basis of all manifestations of the well-known long-range motion system (Braddick 1974). The hypothesis is that the attentional mechanism tends to be attracted to any region of energy change in the image, regardless of its form, Fourier or non-Fourier nature, direction in chromaticity space, etc., whereas the short-range motion systems require the existence of specific channels tuned to the spatiotemporal properties of motion in some local region of the visual field. The sense of motion in a long-range motion stimulus would then be attributed to the motion of the attentional focus under exogenous control. This hypothesis does not need to be specific about the mechanism by which the motion of the attentional focus is sensed; it may be a bottom-up or a top-down process. An example of a bottom-up process would be the spread of the neural excitation trace through the cortical neural network representing the local properties of the visual field. The continuity of positions of the local excitation peak would provide the
moving focus for the motion reconstruction. An example of a top-down process would be the control of an attentional focus in the visual cortex by corollary discharge from an object-reconstruction network in, for example, the parietal cortex. The top-down process would be subject to the learned structural constraints of the properties of objects in the world, whereas the characteristics of the bottom-up process would be inherent in the properties of neural interactions within the cortex. Note that both types of influence are conceived as aspects of the exogenous attentional mechanism. These are questions that remain to be decided by other means. The point for the present purpose is to propose that the mechanism for apparent motion between two physically stationary stimuli is identical with the mechanism for the motion of the attentional focus. The assumption is that the stimulus change for the apparent-motion paradigm forms an exogenous cue to drive the attentional focus, which may be subject to habituation, and may also be under some degree of endogenous (voluntary) control.

Under the attentional hypothesis for apparent motion, the temporal limits of apparent motion would provide a measure of the temporal limits of the attentional mechanism. For lateral motion, these limits were measured by Zeeman and Roelofs (1953) and Tyler (1973) for repetitive or alternating apparent motion of squares or line stimuli. The maximum frequency for which apparent motion was observable at any line separation in central view was ~10 Hz for motion of the bright bars (and ~6 Hz for the alternative percept of motion of the dark space across the bars). Beyond this frequency, the alternating stimuli were perceived as flickering in a fixed position, with no smooth motion between them. For the case of apparent motion in the third dimension, Norcia and Tyler (1984) measured the limits of depth motion to be stable at 5–6 Hz for disparities from 15 to 45 arc min, suggesting that the temporal limit for attentional tracking is similar in all directions in visual space.

To verify that the temporal properties of the 3-D-attentional mechanism operating in the attentional tasks illustrated in figures 7 and 10 match those derived from apparent-motion paradigms, we measured the optimal delay between the onset of the priming plane and the test stimuli. The results (figure 11a) show that performance is a nonmonotonic function of onset delay, peaking at about 100 ms. This behavior mirrors that reported for attentional priming of luminance stimuli by Nakayama and

![Graphs showing discriminability and discriminability ratio over time and frequency](image)

**Figure 11.** (a) Tuning of stereocognition to cue exposure time in the depth-tilt paradigm of figure 10a. A valid cue (filled circles), at the same tilt as the test, was maximally effective when presented for 100 ms prior to the stimulus, but almost ineffective by a duration of 200 ms. An invalid cue (open circles) produced negligible facilitation at any duration. (b) Fourier transform of valid-cue data.
Mackeben (1989). In both cases the priming effect is not instantaneous, taking a significant time to build up, but it then decays rapidly even though the prime is still present for the full duration of the prestimulus interval. In terms of the apparent-motion hypothesis, the time course of this cueing effect provides reasonable confirmation of the expected low-temporal-frequency limit, since the Fourier transform of this function shows a frequency cutoff of about 6 Hz (figure 11b).

6 Object-oriented constraints on surface reconstruction

The concept of the attentional shroud is intended to capture the idea of a mechanism that acts like a soap film in minimizing the curvature of the perceived depth surface consistent with the available disparity information. There is evidence from Julesz (1971) that the perceived surface projects across the space between discrete disparity signs to form the minimal smooth surface. He treated this process as an extension of the ‘search for dense surfaces’ within the disparity array and attributed it to a specialized aspect of the stereoscopic system. Perhaps this process is better viewed as a property of the attentional mechanism, and that this mechanism is not just a local spotlight but a flexible surface that can wrap subsets of the Keplerian disparity field to provide a sense of the object array in the field of view. What, then, are the properties of the proposed surface-wrapping operation? Does it extend uniformly over the full disparity field or are there differentiated edge-specific properties as in luminance processing?

One corollary of this surface-reconstruction approach is a postulate that the object array is represented strictly in terms of its surfaces, as proposed by Nakayama and Shimojo (1990a). As with the majority of objects in the physical world, the reconstructed surface is assumed to be inherently opaque, as seen from the point of view of the observer. This may seem like a tautology, but it is possible, in principle, to imagine that the observer could perceive the surface reconstruction as viewed from any angle, including from a position behind the surface with respect to the observer. However, just as Nakayama and Shimojo have postulated that the borders of ambiguous figure/ground objects ‘belong’ to the shape perceived as ‘figure’ at any instant, so the reconstructed surface is proposed to belong to the outside of the object that it is wrapping (ie the side nearest the observer). The observer is constrained, on this hypothesis, always to be outside convex objects in the scene. Concave objects, on the other hand, are viewed as though from the inside (especially, perhaps, a concave ‘cave’ filling the whole visual scene) but never as solid objects, since we could not then see through them to their surface. However, the observers can never adopt the perceptual point of view of looking back through the scene at their face, presumably because of lack of familiarity with this perspective. Even when looking at one’s own image in the mirror, it seems to be impossible to adopt the perspective of the image looking at oneself.

The point of this discussion is that there is an inherent asymmetry in the viewpoint of the reconstructed surface, which is viewed always from the ‘front’ rather than the ‘back’, with respect to the viewer. There is a corresponding asymmetry in the 3-D reconstruction of arrangements of occluding surfaces. All 3-D arrangements are possible in principle, so there are no absolute constraints in 3-D image reconstruction. However, the observer-centered viewpoint, together with the principle of minimum complexity in the reconstructed image, suggests a set of constraints that may be operating to limit performance in human vision.

One effect of the observer-centered viewpoint that was noted by Julesz (1971) applies to the regions of the scene that are visible to only one eye. These are the regions where the horizontal disparity gradient exceeds a slope of 2 (Burt and Julesz 1980), which is the condition where a surface falls exactly along the line sight of one eye.
For steeper surfaces, the nearer edge of the surface occludes the remainder from that eye, so that it is only visible to the other eye. These monocular 'no-man's-lands' in the depth array carry no stereoscopic depth information, so they could, in principle, be perceived as anywhere behind the boundary of visibility in which the image is visible only monocularly. In practice, these regions usually are perceived as continuous with the far surface running into the monocular region rather than any other more exotic examples from the infinite set of possible percepts (see figure 12). Note that the perception of stereoscopic surfaces requires a cyclopean view in which the two eyes' views are integrated into a unified percept. This percept must be perceived at a particular location in space as viewed from a unitary egocenter known as the cyclopean eye (Hering 1868). It is only the monocular segment visible in the projection to the cyclopean eye that requires to be assigned a depth in the stereoreconstruction process.

As shown by Julesz (1971), the principle of completion of far surfaces operates for simple surfaces such as planes, whatever the shape of the aperture through which the far surface is viewed. What has not been tested is the degree of generality of this reconstruction process. Will it just extend the local disparity horizontally across the region being reconstructed in a simple zero-order extrapolation, or will the brain employ top-down processes to optimize the reconstructed segment according to a variety of real-world constraints?

Observation of such cyclopean steps in the two-panel stereograms of figures 13–16 reveals that some constraints are observed while others are not. Figure 13 shows a step between zero-order planes in which the edges of the correlated planes are indicated by the thin black lines extending from the top and bottom of the figure. The upper and lower pairs show opposite directions of steps. The pair of lines in the upper left and lower right panels are provided to define the 'no-man's-land' of monocular dots with no corresponding mates in the opposite panel. All other dots occur in matched pairs between the two images. The question is, what depth do these uncorrelated dots appear to take when the fields are viewed stereoscopically?

Figure 12. Diagram of the binocular geometry and cyclopean view of a stereoscopic edge. Note that there is a monocular zone visible only to the right eye (for an edge with the left side closer). Dotted lines, optical lines of sight; thin lines, cyclopean projections; thick lines, stereoscopic image; thick dashed lines, cyclopean reconstructions of the monocular zone; CE, functional spatial location of the cyclopean eye. The question is, what depth does the brain assign to this monocular zone? Is it influenced by the depth of the adjacent binocularly visible regions? Three possible solutions are depicted for the monocular zone; A, continuation of the far plane; B, interpolation between the binocularly defined edges; C, continuation of the near plane.
Observation of figure 13 reveals that, with fixation on the front plane close to the step, the monocular zone appears to be continuous with the rear plane; this case corresponds to solution A as depicted in figure 12. (Both edge directions are included because this effect may be more pronounced in one direction than the other.) With fixation on the rear plane, many observers report fingers extending from the front plane and tending to fill in across the monocular zone in front of the fixation plane, corresponding to solution C in figure 12. Interestingly, solution B never seems to be observed, although it would make a perfectly logical interpretation of the ambiguous information. This selectivity implies that object-oriented constraints are brought to bear on the stereoreconstruction process in selecting surface continuity and right angles over other possible interpretations.

**Figure 13.** Two-panel random-dot stereograms for evaluating the stereoreconstruction of the depth in the monocular region between zero-order planes across a horizontal disparity step. Vertical lines delimit the width of the monocular region. Upper and lower pairs present opposite directions of the disparity step.
The dependence of the interpretation of the monocular zone on the disparities of the surrounding planes is contrary to the predictions of the occlusion hypothesis of Nakayama and Shimojo (1990b). Occlusion would predict the monocular zone to appear behind the edge when the front edge was fixated, just as is observed, but that it also should be behind the fixated plane in the case of rear fixation. The observation that the front plane may be perceived as continuing across the monocular zone implies that higher-order reconstruction processes are at work beyond the simple precepts of da Vinci stereopsis (Nakayama and Shimojo 1990b).

To characterize more of the properties of the depth-reconstruction process, we introduced higher-order surfaces on either side of the edge. Figure 14 shows a horizontal disparity ramp in the rear surface. With fixation in the front plane, the monocular zone seems to take a depth close to that of the closest edge of the rear plane and project across horizontally to the edge, with perhaps some curvature up to

Figure 14. Two-panel random-dot stereograms for depth reconstruction in the monocular region across a horizontal disparity step with a first-order gradient in the rear plane. Vertical lines delimit the width of the monocular region. Upper and lower pairs present opposite directions of the disparity step.
the front plane close to its edge. With fixation on the rear plane at the edge on the monocular zone, the depth extends out from the front plane, as before. In neither fixation position is there any tendency to complete with the slope of the ramp. The same results are obtained when the rear plane contains a horizontal disparity modulation (figure 15). Note that a sinusoidal modulation contains derivatives of all orders, so it provides a test of the completion effect to all orders of surface curvature. It therefore seems that the completion is dominated by the disparity of the edge adjacent to the monocular zone and is uninfluenced by the horizontal derivatives of the disparity modulation.

The remaining question in this context is whether there is any effect of the vertical component of disparity modulation. If the stereoreconstruction process can be induced to complete across from the depth of the nonfixated edge (figure 12), can this process adapt locally to variations in the depth of the inducing edge? This question is addressed

Figure 15. Two-panel random-dot stereograms for depth reconstruction in the monocular region across a disparity step with a horizontal sinusoidal disparity-correlation profile in the rear plane. Vertical lines delimit the width of the monocular region. Upper and lower pairs present opposite directions of the disparity step.
with the stereograms shown in figure 16, where the rear plane is given a vertical sinusoidal disparity corrugation (appearing as horizontal depth bars). Most observers see a clear continuation of these corrugations across the monocular 'no-man's-land' up to the resolution limit of the corrugations (cf figure 9). It therefore appears that the spatial resolution for the monocular region is the same as that for the binocularly-defined surface, suggesting that the stereoreconstruction process operates under similar constraints in both cases.

Beyond these properties of two intersecting surfaces are the issues of the spatial configuration of multiple intersections. For example, the same principle of monocular no-man's-lands depicted in figures 13–16 implies that narrow channels should be seen as wider than narrow ridges when at any orientation other than horizontal (Julesz 1971). The reason for this asymmetry is that the monocular regions tend to be associated with the far plane, which is the base plane in the case of the ridge but

Figure 16. Two-panel random-dot stereograms for depth reconstruction in the monocular region across a disparity step with a vertical sinusoidal disparity-corrugation profile in the rear plane. Vertical lines delimit the width of the monocular region. Upper and lower pairs present opposite directions of the disparity step.
the bottom of the channel when the same set of disparity relations are viewed as uncrossed rather than crossed (as depicted in figure 17).

This figure should make the asymmetry between the two constructions clear. In reconstructing the channel (figure 17a), its width is extended across the entire monocular zone, extending from the depth of the narrow binocular portion of the channel. If there is no binocular portion, this region is rivalrous and is perceived according to the principles of rivaldepth (O'Shea and Blake 1987; Tyler 1991b). Reconstruction of the ridge (figure 17b) is slightly different because there may be a portion of the base plane that is invisible to both eyes. This region may be apprehended as continuing behind the ridge by the process of amodal completion, but it is not perceived in as direct a fashion as the monocular regions on either side. As may be seen in the auto-stereogram of intersecting channels and ridges (figure 18), the monocularly visible regions are completed according to the depth of the binocularly visible regions beyond, so that the vertical channels appear substantially wider than the ridges and no rivalry is perceived in the monocular strips. This is the case when fixation is maintained on the base plane, so there is no question of the asymmetry being due to convergence micropia for the nearer versus the further bar. The horizontal channels and ridges appear the same width because no monocular regions are created by the vertical disparity changes across the horizontal edges [since these changes are orthogonal to the disparities themselves (Julesz 1971; Tyler and Raibert 1975)].

A further property of stereoreconstruction that is illustrated in figure 18 is the facilitation of the depth solution when it conforms with readily recognizable object properties. Four kinds of intersections between horizontal and vertical bars at different depths are depicted. Most observers find it much easier to see the near bars intersecting, which can be parsed as two elevated roadways crossing at different heights, than to see the far bars intersecting, which does not match the profile required for two channels intersecting at different depths. This difference occurs

![Diagram of binocular geometry and cyclopean view of stereoscopic bars](image)

**Figure 17.** Diagram of the binocular geometry and cyclopean view of stereoscopic bars. Dotted lines, optical lines of sight; thin lines, cyclopean projection; thick lines, stereoscopic image; thick dashed lines, cyclopean reconstructions of the monocular zone; CE, functional spatial location of the cyclopean eye. (a) The bar has an uncrossed disparity to form a channel. (b) The bar has a crossed disparity to form a ridge. Note that in both cases there are zones visible only to one eye on either side of the bar. Most observers see these monocular zones as continuous with the far plane, so that the channel appears broader than the ridge when fixation is on the base plane (corresponding to solution a in figure 12).
despite that fact that the local structure and 3-D spatial frequency content is identical for the two types. Familiarity with the kinds of structure that are common in the real world therefore seems to play a role in the perception of these structures.

The principles illustrated in figures 12-18 provide insights into the operation of the human stereoreconstruction process when it is faced with ambiguous stimuli in the monocularly visible regions at disparity edges. They suggest that the constraints employed by the human visual system are object oriented even at this relatively low level of the reconstruction process.

Figure 18. Autostereogram of intersecting channels and ridges.

7 Concluding remarks
We have presented evidence for a rapid, surface-based attentional mechanism operating under exogenous control in the process of 3-D image reconstruction. This raises the question of how the attentional mechanism obtains the information about where to focus in the depth field, since the implication is that within 100 ms it focuses on the salient depth level at any position in the image. In a sense, this pushes the issue of depth reconstruction back a level from that of the full process to the question of depth-reconstruction process that drives the attentional mechanism.

First, we should emphasize that the 3-D-attentional mechanism that we are exploring is an exogenous one, driven by the information in the stimuli alone. The experimental paradigms were designed to randomize across the observers' expectations, so that no endogenous control of the depth or the 3-D form of the attentional enhancement according to the observers' expectations was possible.
This leads to an analysis of what one means by a stimulus-driven attentional enhancement as opposed to a direct response to the stimulus. In current terminology, attentional enhancement may be viewed as a second-order process, one insensitive to the sign or local features of the information that is enhanced, and must have a temporal memory (albeit a short one). Thus, in the absence of an attentional prime, there would be some defined response to a particular stimulus. If an attentional prime preceded the stimulus by a short time, the response would be facilitated regardless of its specific characteristics.

One appealing process by which such a second-order facilitation could occur is by limiting the observer's response to control by a small subset of all the neurons that could possibly be activated. Let us say that the observer is initially set to respond to a signal in any of the neural elements in the Keplerian array (ignoring all other neural signals on the basis of endogenous attention, perhaps). The noise against which a test signal is detected or discriminated would then be composed of the combined noise of all these elements. The exogenous attention might operate to limit the processing to those elements that were previously activated by the attentional prime, thus eliminating the noise (or uncertainty) associated with all the unactivated elements. Such selectivity could form the basis of the attentional enhancement of whatever features were presented to the activated elements after the prime. This noise-reduction hypothesis could be tested in future experiments by measuring the equivalent noise threshold for external noise added to the enhanced and nonenhanced regions, respectively.

It is a further property of any second-order process that it forms a coarser representation of the input image than the first-order process on which it is based. In this sense, the operation of an exogenous attentional process represents a coarse-to-fine analysis that is somewhat analogous to that proposed by Marr and Poggio (1979). In the context of a 3-D-attentional process, however, the coarse and fine representations refer to resolution in the disparity domain (i.e., coarse and fine determination of value of the depth token) rather than in the spatial frequency domain.

To have significant selectivity within the Keplerian array when stimulated by random-dot stereogram forms, the attentional mechanism has to solve the correspondence problem to some degree, otherwise it would be swamped by all the disparities of the spurious correspondence matches. From the viewpoint of the cooperative algorithms of Julesz (1971) and Marr and Poggio (1979), it might seem unlikely that such cooperative process could operate within the 100 ms time constant of the attentional enhancement shown in figure 11. The selectivity is, however, compatible with the 'local cleaning' operations, such as disparity inhibition across the Keplerian array, that were proposed by Tyler (1975a, 1983, 1991b, 1993; see c in figure 1). There are direct-interaction processes that could be hard wired to operate within the empirical constraints. Such local processes would be sufficient to generate a coarse depth representation by which to drive the attentional shroud, to allow selective enhancement of a complex depth profile for subsequent processing of its detailed depth features. It is hoped that this coarse picture of the operation of the exogenous 3-D-attentional mechanism will serve to focus studies on its predictions that will subsequently refine the details of our understanding.

Acknowledgement. This work was supported by NIH grant EY7890.

References
Breitmeyer B, Julesz B, Kropfl W, 1976 "Dynamic random-dot stereograms reveal up-down anisotropy and left-right isotropy between cortical hemifields" Science 187 269–270
Hering E, 1868 *Die Lehre von Binocularen Sehen* (Leipzig: Engelmann)
Helmholtz H von, 1866 *Handbuch der Physiologischen Optik* (Hamburg: Vos)
Papert S, 1964 “Stereoscopic synthesis as a technique for locating visual mechanisms” *MI Quarterly Progress Report* **73** 239–243
Tyler C W, 1975a “Spatial organization of binocular disparity sensitivity” Vision Research 15 583–590
Tyler C W, 1975b “Stereoscopic tilt and size aftereffects” Perception 4 187–192
Tyler C W, Julesz B, 1976 “The neural transfer characteristic (neurometry) for binocular stochastic stimulation” Biological Cybernetics 23 33–37
Tyler C W, Raibert M, 1975 “Generation of random-dot stereogratings” Behavioral Research Methods and Instrumentation 7 37–41