# The quantized geometry of visual space: The coherent computation of depth, form, and lightness

## Stephen Grossberg

Center for Adaptive Systems, Department of Mathematics, Boston University, Boston, Mass. 02215

Abstract: A theory is presented of how global visual interactions between depth, length, lightness, and form percepts can occur. The theory suggests how quantized activity patterns which reflect these visual properties can coherently fill-in, or complete, visually ambiguous regions starting with visually informative data features. Phenomena such as the Cornsweet and Craik-O'Brien effects, phantoms and subjective contours, binocular brightness summation, the equidistance tendency, Emmert's law, allelotropia, multiple spatial frequency scaling and edge detection, figure-ground completion, coexistence of depth and binocular rivalry, reflectance rivalry, Fechner's paradox, decrease of threshold contrast with increased number of cycles in a grating pattern, hysteresis, adaptation level tuning, Weber law modulation, shift of sensitivity with background luminance, and the finite capacity of visual short term memory are discussed in terms of a small set of concepts and mechanisms. Limitations of alternative visual theories which depend upon Fourier analysis, Laplacians, zero-crossings, and cooperative depth planes are described. Relationships between monocular and binocular processing of the same visual patterns are noted, and a shift in emphasis from edge and disparity computations toward the characterization of resonant activity-scaling correlations across multiple spatial scales is recommended. This recommendation follows from the theory's distinction between the concept of a structural spatial scale, which is determined by local receptive field properties, and a functional spatial scale, which is defined by the interaction between global properties of a visual scene and the network as a whole. Functional spatial scales, but not structural spatial scales, embody the quantization of network activity that reflects a scene's global visual representation. A functional scale is generated by a filling-in resonant exchange, or FIRE, which can be ignited by an exchange of feedback signals among the binocular cells where monocular patterns are binocularly matched.

Keywords: binocular vision; brightness perception; figure-ground; feature extraction; form perception; neural network; nonlinear resonance; receptive field; short-term memory; spatial scales; visual completion

The objects of perception and the space in which they seem to lie are not abstracted by a rigid metric but a far looser one than any philosopher ever proposed or any psychologist dreamed. Jerome Lettvin (1981)

1. Introduction: The abundance of visual models. Few areas of science can boast the wealth of interesting and paradoxical phenomena readily accessible to introspection that visual perception can. The sheer variety of effects helps to explain why so many different types of theories have arisen to carve up this data landscape. Fourier analysis (Cornsweet 1970; Graham 1981; Robson 1975), projective geometry (Beck 1972; Johannson 1978; Kaufman 1974), Riemannian geometry (Blank 1978; Luneberg 1947; Watson 1978), special relativity (Caelli, Hoffman & Lindman 1978), vector analysis (Johannson 1978), analytic function theory (Schwartz 1980), potential theory (Sperling 1970), and cooperative and competitive networks (Amari & Arbib 1977; Dev 1975; Ellias & Grossberg 1975; Grossberg 1970a; 1973; 1978e; 1981; Sperling 1970; Sperling & Sondhi 1968) are just some of the formalisms which have been used to interpret and explain particular visual effects. Some of the most distinguished visual researchers believe that this diversity of formalisms is inherent in the nature of psychological phenomena. Sperling (1981, p. 282) has, for example, recently written:

In fact, as many kinds of mathematics seem to be applied to perception as there are problems in perception. I believe this multiplicity of theories without a reduction to a common core is inherent in the nature of psychology . . . , and we should not expect the situation to change. The moral, alas, is that we need many different models to deal with the many different aspects of perception.

The opinion Sperling offers is worthy of the most serious deliberation, since it predicts the type of mature science which psychology can hope to become, and thereby constrains the type of theorizing which psychologists will try to do. Is Sperling right? or Do there exist concepts and properties, heretofore not explicitly incorporated into the mainstream visual theories, which can better unify the many visual models into an integrated visual theory?

Part I of this article reviews various visual data as well as internal paradoxes and inherent limitations of some recent theories that have attempted to explain these data. Part II presents a possible approach to overcoming these paradoxes and limitations and to explaining the data in a

unified fashion. Numerical simulations that support the qualitative arguments and mathematical properties described in Part II are found in Cohen & Grossberg (1983a). Parts I and II are self-contained and can be read in either order.

## Part I

2. The quantized geometry of visual space. There is an important sense in which Sperling's assertion is surely correct, but in this sense it is also true of other sciences such as physics. Different formalisms can probe different levels of the same underlying physical reality without excluding the possibility that one formalism is more general, or physically deeper, than another. In physics, such theoretical differences can be traced to physical assumptions which approximate certain processes in order to clarify other processes. I will argue that several approaches to visual perception make approximations which do not accurately represent the physical processes which they have set out to explain. For this reason, such theories have predictive limitations which do not permit them to account, even to a first approximation, for major properties of the data. In other words, the mathematical formalism of these theories has not incorporated fundamental physical intuitions into their computational structure. Once these intuitions are translated into a suitable formalism, the theoretical diversity in visual science will, I claim, gradually become qualitatively more like that known in physics.

The comparison with physics is not an idle one. Certain of the intuitions which need to be formalized at the foundations of visual theory are well known to us all. They have not been acted upon because, despite their simplicity, they lead to conceptually radical conclusions that force a break with traditional notions of geometry. Lines and edges can no longer be thought of as a series of points; planes can no longer be built up from local surface elements or from sets of lines or points; and so on. All local entities evaporate as we build up notions of functional perceptual units which can naturally deal with the global context-dependent nature of visual percepts. The formalism in which this is achieved is a quantized dynamic geometry, and the nature of the quantization helps to explain why so many visual percepts seem to occur in a curved visual space.

When a physicist discusses quantization of curved space, he usually means joining quantum mechanics to general relativity. This goal has not yet been achieved in physics. To admit that even the simplest visual phenomena suggest such a formal step clarifies both the fragmentation of visual science into physically inadequate formalisms, and the radical nature of the conceptual leap that is needed to remedy this situation.

3. The need for theories which match the data's coherence. As background for my theoretical treatment, I will review various paradoxical data concerning interactions between the perceived depth, lightness, and form of objects in a scene. These paradoxes should not, I believe, be viewed as isolated and unimportant anomalies, but rather as informative instances of how the visual system completes a scene's global representation in response to locally ambiguous visual data. These data serve to remind us of the interdependence and context-sensitivity of visual properties; in other words, of their coherence. With these reminders fresh in our minds, I will argue in Part II that by probing important visual design principles on a deep mathematical level, one can discover, as automatic mathematical consequences, the way many visual properties are coherently caused as manifestations of these design principles.

This approach to theory construction is not in the mainstream of psychological thinking today. Instead, one often finds models capable of computing some single visual property, such as edges or cross-correlations. Even with a different model for each property, this approach does not suggest how related visual properties work together to generate a global visual representation. For example, the present penchant for modeling lateral inhibition by linear feedforward operators like a Laplacian or a Fourier transform to compute edges or cross-correlations (Marr & Hildreth 1980; Robson 1975) pays the price of omitting related nonlinear properties like reflectance processing, Weber law modulation, figure-ground fillingin, and hysteresis. To the argument that one must first understand one property at a time, I make this reply: The feedforward linear theories contain errors even in the analysis of the concepts they set out to explain. Internal problems of these theories prevent them from understanding the other phenomena that cohere in the data.

This lack of coherence, let alone correctness, will cause a heavy price to be paid in the long run, both scientifically and technologically. Unless the relationships among visual data properties are correctly represented in a distributed fashion within the system, plausible (and economic) ways to map these properties into other subsystems, whether linguistic, motor, or motivational, will be much harder to understand. Long-range progress, whether in theoretical visual science per se or in its relations to other scientific and technological disciplines, requires that the mathematical formalisms in which visual concepts are articulated be scrupulously criticized.

## 4. Some influences of perceived depth on perceived size. Interactions between an object's perceived depth, size, and lightness have been intensively studied for many years. The excellent texts by Cornsweet (1970) and by Kaufman (1974) review many of the basic phenomena.

The classical experiments of Holway and Boring (1941) show that observers can estimate the actual sizes of objects at different distances even if all the objects subtend the same visual angle on the observers' retinas. Binocular cues contribute to the invariant percept of size. For example, Emmert (1881) showed that monocular cues may be insufficient to estimate an object's length. He noted, among other properties, that a monocular afterimage seems to be located on any surface which the subject binocularly fixates while the afterimage is active. Moreover, the perceived size of the afterimage increases as the perceived distance of the surface increases. This effect is called Emmert's law. Although the use of monocular afterimages to infer properties of normal viewing is fraught with difficulties, other paradigms have also suggested an effect of perceived depth on perceived size.

For example, Gogel (1956, 1965, 1970) has reported that two objects viewed under reduction conditions (one

eye looks through a small aperture in dim light) will be more likely to be judged as equidistant from the observer as they are brought closer together in the frontal plane. In a related experiment, one object is monocularly viewed through a mirror arrangement whereas all other objects in the scene are binocularly viewed. The monocularly viewed object then seems to lie at the same distance as the edge that, among all the binocularly viewed objects, is retinally most contiguous to it. Gogel interpreted these effects as examples of an equidistance tendency in depth perception. The equidistance tendency also holds if a monocular afterimage occupies a retinal position near to that excited by a binocularly viewed object. One way to interpret these results is to assert that the perceived distance of the binocular object influences the perceived distance of the adjacent afterimage by the equidistance tendency, and thereupon influences the perceived size of the afterimage by Emmert's law.

Results such as these suggest that depth cues can influence size estimates. They also suggest that this influence can propagate between object representations whose cues excite disparate retinal points and that the patterning of all cues in the visual context of an object helps to determine its perceived length. The classical geometric notion that length can be measured by a ruler, or can be conceptualized in terms of *any* locally defined computation, thereby falls into jeopardy.

5. Some monocular constraints on size perception. Size estimates can also be modified by monocular cues, as in the corridor illusion (Richards & Miller 1971; see Figure 1a). In this illusion, two cylinders of equal size in a picture are perceived to be of different sizes because they lie in distinct positions within a rectangular grid whose spatial scale diminishes toward a fixation point on the horizon. An analogous effect occurs in the Ponzo illusion shown on the right, wherein two horizontal rods of equal pictorial length are drawn superimposed over an inverted V (Kaufman 1974; see Figure 1b). The upper rod appears longer than the lower rod. The perception of these particular figures may be influenced by learned depth perspective cues (Gregory 1966), although this hypothesis does not explain how perspective cues alter length percepts. There exist many other figures, however, in which a perspective effect on size scaling is harder to rationalize



Figure 1. (a) The corridor illusion. (b) The Ponzo illusion. (After Kaufman 1974. From Sight and Mind: An Introduction to Visual Perception. Copyright © 1974 by Oxford University Press, Inc. Reprinted by permission.)



Figure 2. Attneave's cat: Connecting points of maximum curvature with straight lines yields a recognizable caricature of a cat. (After Attneave 1954.)

(Day 1972). Several authors have therefore modeled these effects in terms of intrinsic scaling properties of the visual metric (Dodwell 1975; Eijkman, Jongsma & Vincent 1981; Restle 1971; Watson 1978).

A more dramatic version of scaling is evident when subjective contours complete the boundary of an incompletely represented figure. Then objects of equal pictorial size that lie inside and outside the completed figure may appear to be of different size (Coren 1972). The very existence of subjective contours raises the issue of how incomplete data about form can select internal representations which can span or fill-in the incomplete regions of the figure. How can we characterize those features or spatial scales in the incomplete figure which play an informative role in the completion process versus those features or scales which are irrelevant? Attneave (1954) has shown, for example, that when a drawing of a cat is replaced by a drawing in which the points of maximum curvature in the original are joined by straight lines, then the new drawing still looks like a cat (see Figure 2). Why are the points of maximum curvature such good indicators of the entire form? Is there a natural reason why certain spatial scales in a figure might have greater weight than other scales? Attneave's cat raises the question: Why does interpolation between points of maximum curvature with lines of zero curvature produce a good facsimile of the original picture? Different spatial scales somehow need to interact in our original percept for this to happen. To understand this issue, we need a correct definition of spatial scale. Such a definition should distinguish between local scaling effects, such as those which can be understood in terms of a neuron's receptive field (Robson 1975), and global scaling effects, such as those which control the filling-in of subjective contours or of phantom images across a movie screen, which subtends a visual angle much larger than that spanned by any neuron's receptive field (Smith & Over 1979; Tynan and Sekuler 1975; von Grunau 1979; Weisstein, Maguire & Berbaum (1976).

6. Multiple scales in figure and ground: simultaneous fusion and rivalry. That interactions between several spatial scales are needed for form perception is also illustrated by the following type of demonstration (Beck 1972). Represent a letter E by a series of nonintersecting straight lines of varying oblique and horizontal orientations drawn within an imaginary E contour and surrounded by a background of regular vertical lines. The E is not perceived because of the lines within the contour,



Figure 3. The Kaufman stereogram induces an impression of depth even though the darker line patterns are rivalrous. (After Kaufman 1974. From Sight and Mind: An Introduction to Visual Perception. Copyright © 1974 by Oxford University Press, Inc. Reprinted by permission.)

since the several orientations of these interior lines do not group into an E-like shape. Somehow the E is synthesized as the complement of the regular background, or, more precisely, by the statistical differences between the figure and the ground. These statistical regularities define a spatial scale — broader than the scale of the individual lines — on which the E can be perceived.

In a similar vein, construct a stereogram out of two pictures as follows (Kaufman 1974; see Figure 3). The left picture is constructed from 45°-oblique dark parallel lines bounded by an imaginary square, which is surrounded by 135°-oblique lighter parallel lines. The right picture is constructed from 135°-oblique dark parallel lines bounded by an imaginary square whose position in the picture is shifted relative to the square in the left picture. This imaginary square is surrounded by 45°-oblique lighter parallel lines. When these pictures are viewed through a stereoscope, the dark oblique lines within the square are rivalrous. Nonetheless the square as a whole is seen in depth. How does this stereogram induce rivalry on the level of the narrowly tuned scales that interact preferentially with the lines, yet simultaneously generate a coherent depth impression on the broader spatial scales that interact preferentially with the squares?

Kulikowski (1978) has also studied this phenomenon by constructing two pairs of pictures which differ in their spatial frequencies (see Figure 4). Each picture is bounded by the same frame, as well as by a pair of short vertical

reference lines attached to the outside of each frame at the same spatial locations. In one pair of pictures, spatially blurred black and white vertical bars of a fixed spatial frequency are 180° out of phase. In the other pair of pictures, sharp black and white vertical bars of the same spatial extent are also 180° out of phase. The latter pair of pictures contains high spatial frequency components (edges) as well as low spatial frequency components. During binocular viewing, subjects can fuse the two spatially blurred pictures and see them in depth with respect to the fused images of the two frames. By contrast, subjects experience binocular rivalry when they view the two pictures of sharply etched bars. Yet they still experience the rivalrous patterns in depth. This demonstration suggests that the low spatial frequencies in the bar patterns can be fused to yield a depth impression even while the higher spatial frequency components in the bars elicit an alternating rivalrous perception of the monocular patterns.

The demonstrations of Kaufman (1974) and Kulikowski (1978) raise many interesting questions. Perhaps the most pressing one is: Why are fusion and rivalry alternative binocular perceptual modes? Why are coexisting unfused monocular images so easily supplanted by rivalrous monocular images? How does fusion at one spatial scale coexist with rivalry at a different spatial scale that represents the same region of visual space?

7. Binocular matching, competitive feedback, and monocular self-matching. These facts suggest some conclusions that will be helpful in organizing my data review and will be derived on a different theoretical basis in Part II. I will indicate how rivalry suggests the existence of binocular cells that can be activated by a single monocular input and that mutually interact in a competitive feedback network. First I will indicate why these binocular cells can be monocularly activated.

The binocular cells in question are the spatial loci where monocular data from the two eyes interact to generate fusion or rivalry as the outcome. To show why at least some of these cells can be monocularly activated, I will consider implications of the following mutually exclu-



Figure 4. Demonstration of depth perception with and without fusion. (a) Sinusoidal gratings in antiphase can be fused to yield a depth impression. (b) The square wave gratings yield a depth impression even when their sharp edges become double. (c) A similar dichotomy is perceived when single sinusoidal or bars are viewed. (After Kulikowski 1978.) Reprinted by permission from Nature, volume 275, pp. 126–127. Copyright © Macmillan Journals Limited.

Suppose it does not. Then the activities of monocular cells cannot subserve perception; rather, perception is associated with activities of binocular cells or of cells more central than the binocular cells. This is because both sets of monocular cells would remain active during a rivalry percept, since the binocular interaction leading to the rivalry percept does not, by hypothesis, feed back to alter the activities of the monocular cells. Now we confront the conclusion that monocular cells do not subserve perception with the fact that the visual world can be vividly seen through a single eye. It follows that some of the binocular cells which subserve perception can be activated by input from a single eye.

Having entertained the hypothesis that the outcome of binocular matching does not feed back toward monocular cells, let us now consider the opposite hypothesis. In this case, too, I will show that a single monocular representation must be able to activate certain binocular cells. To demonstrate this fact, I will again argue by contradiction.

Suppose it does not. In other words, suppose that the outcome of binocular matching does feed back toward monocular cells but a single monocular input cannot activate binocular cells. Because the visual world can be seen through a single eye, it follows that the activities of monocular cells subserve perception in this case. Consequently, during a binocular rivalry percept, the binocular-to-monocular feedback must quickly inhibit one of the monocular representations. The signals which this monocular representation was sending to the binocular cells are thereupon also inhibited. The binocular cells then receive signals only from the other monocular representation. The hypothesis that binocular cells cannot fire in response to signals from only one monocular representation implies that the binocular cells shut off, along with all of their output signals. The suppressed monocular cells are then released from inhibition and are excited again by their monocular inputs. The cycle can now repeat itself, leading to the percept of a very fast flicker of one monocular view superimposed upon the steady percept of the other monocular view. This phenomenon does not occur during normal binocular vision. Consequently, the hypothesis that a single monocular input cannot activate binocular cells must be erroneous. Whether or not the results of binocular matching feed back toward monocular cells, certain binocular cells can be activated by a single monocular representation.

An additional conclusion can be drawn in the case wherein the results of binocular matching can feed back toward monocular cells. Here a single monocular source can activate binocular cells, which can thereupon send signals toward the monocular source. The monocular representation can thereby *self-match* at the monocular source using the binocular feedback as a matching signal. This fact implies that the monocular source cells are themselves binocular cells, because a monocular input can activate binocular cells which then send feedback signals to the monocular source cells of the other eye. In this way the monocular source cells can be activated by both eyes, albeit less symmetrically than the binocular cells at which the primary binocular matching event takes place. This conclusion can be summarized as follows: The binocular cells at which binocular matching takes place are flanked by binocular cells that satisfy the following properties: (a) they are fed by monocular signals; (b) they excite the binocular matching cells; (c) they can be excited or inhibited due to feedback from the binocular matching cells, depending upon whether fusion or rivalry occur.

It remains only to consider the possibility that the results of binocular matching do not feed back toward the monocular cells. The following argument indicates why this cannot happen. A purely feedforward interaction from monocular toward binocular cells cannot generate the main properties of rivalry, namely a sustained monocular percept followed by rapid and complete suppression of this percept when it is supplanted by the other monocular percept. This is because the very activity of the perceived representation must be the cause of its habituation and loss of competitive advantage relative to the suppressed representation. Consequently, the habituating signals from the perceived representation that inhibit the suppressed representation reach the latter at a stage at, or prior to, that representation's locus for generating signals to the perceived representation that are capable of habituating. Such an arrangement allows the signals of the perceived representation to habituate but spares the suppressed representation from habituation. By symmetry, the two representations reciprocally send signals to each other that are received at, or at a stage prior to, their own signaling cells. This arrangement of signaling pathways defines a feedback network.

One can now refine this conclusion by going through arguments like those above to conclude that (a) the feedback signals are received at binocular cells rather than at monocular cells, and (b) the feedback signals are not all inhibitory signals or else binocular fusion could not occur. Thus a competitive balance between excitatory and inhibitory feedback signals among binocular cells capable of monocular activation needs to be considered. Given the possibility of monocular self-matching in this framework, one also needs to ask why the process of monocular self-matching, in the absence of a competing input from the other eye, does not cause the cyclic strengthening and weakening of monocular activity that occurs when two nonfused monocular inputs are rivalrous.

One does not need a complete theory of these properties to conclude that no theory in which only a feedforward flow of visual patterns from monocular to binocular cells occurs (e.g., to compute disparity information) can explain these data. Feedback from binocular matching toward monocular computations is needed to explain rivalry data, just as such feedback is needed to explain the influence of perceived depth on perceived size or brightness. I will suggest in Part II how a suitably defined feedback scheme can give rise to all of these phenomena at once.

8. Against the Keplerian view: Scale-sensitive fusion and rivalry. The Kaufman (1974) and Kulikowski (1978) experiments also argue against the Keplerian view, which is a mainstay of modern theories of stereopsis. The Keplerian view is a realist hypothesis which suggests that the two monocular views are projected point-by-point along diagonal rays, and that their crossing-points are loci from

which the real depth of objects may be computed (Kaufman 1974). When the imaginary rays of Kepler are translated into network hardware, one is led to assume that network pathways carrying monocular visual signals merge along diagonal routes (Sperling 1970). The Keplerian view provides an elegant way to think about depth, because (other things being equal) objects which are closer should have larger disparities, and their Keplerian pathways should therefore cross at points which are further along the pathways. Moreover, all pairs of points with the same disparity cross at the same distance along their pathway, and thereby form a row of contiguous crossing-points.

This concept does not explain a result such as Kulikowski's, since all points in each figure (so the usual reasoning goes) have the same disparity with respect to the corresponding point in the other figure. Hence all points cross in the same row. In the traditional theories, this means that all points should match equally well to produce an unambiguous disparity measure. Why then do low spatial frequencies seem to match and yield a depth percept at the same disparity at which high spatial frequencies do not seem to match?

Rather than embrace the Keplerian view, I will suggest how suitably preprocessed input data of fixed disparity can be matched by certain spatial scales but not by other spatial scales. To avoid misunderstanding, I should immediately say what this hypothesis does not imply. It does not imply that a pair of high spatial frequency input patterns of large disparity cannot be matched, because only suitable statistics of the monocular input patterns will be matched, rather than the input patterns themselves. Furthermore, inferences made from linear statistics of the input patterns do not apply because the statistics in the theory need to be nonlinear averages of the input patterns to ensure basic stability properties of the feedback exchange between monocular and binocular cells. These assertions will be clarified in Part II.

Once the Keplerian view is questioned, the problem of false-images (Julesz 1971), which derives from this view and which has motivated much thinking about stereopsis, also becomes less significant. The false-images are those crossing-points in Kepler's grid that do not correspond to the objects' real disparities.

Workers like Marr and Poggio (1979) have also concluded that false images are not a serious problem if spatial scaling is taken into account. Their definition of spatial scale differs from my own in a way that highlights how a single formal definition can alter the whole character of a theory. For example, when they mixed their definition of a spatial scale with their view of the falseimage problem, Marr and Poggio (1979) were led to renounce cooperativity as well, which I view as an instance of throwing out the baby with the bathwater, since all global filling-in and figure-ground effects thereby become inexplicable in their theory. Marr and Poggio (1979) abandoned cooperativity because they did not need it to deal with false images. In a model such as theirs, the primary goal of which is to compute unambiguous disparity measures, their conclusion seems quite logical. Confronted by the greater body of phenomena that are affected by depth estimates, such a step seems unwarranted.

**9.** Local versus global spatial scales. Indeed, both the Kaufman (1974) and the Kulikowski (1978) experiments, among many others, illustrate that a figure or ground has a coherent visual existence that is more than the sum of its unambiguous feature computations. Once a given spatial scale makes a good match in these experiments, a depth percept is generated that pervades a whole region. We therefore need to distinguish the scaling property that makes good matches based on local computations from the global scaling effects that fill-in an entire region subtending an area much broader than the local scales themselves.

This distinction between local and global scaling effects is vividly demonstrated by constructing a stereogram in which the left "figure" and its "ground" are both induced by a 5% density of random dots (Julesz 1971b, p. 336) and the right "figure" of dots is shifted relative to its position in the left picture. Stereoscopically viewed, the whole figure, including the entire 95% of white background between its dots, seems to hover at the same depth. How is it that the white background of the "figure" inherits the depth quality arising from the disparities of its meagerly distributed dots, and the white background of the 'ground" inherits the depth quality of its dots? What mechanism organizes the locally ambiguous white patches that dominate 95% of the pictorial area into two distinct and internally coherent regions? Julesz (1971b, p. 256) describes another variant of the same phenomenon using a random-dot stereogram inspired by an experiment of Shipley (1965). In this stereogram, the traditional center square in depth is interrupted by a horizontal white strip that cuts both the center square and the surround in half. During binocular viewing, the white strip appears to be cut along the contours of the square and it inherits the depth of figure or ground, despite the fact that it provides no disparity or brightness cues of its own at the cut regions.

## 10. Interaction of perceived form and perceived position.

The choice of scales leading to a depth percept can also cause a shift in perceived form, notably in the relative distance between patterns in a configuration. For example, when a pattern ABC is viewed through one eye and a pattern ABC is viewed through the other eye, the letter B can be seen in depth at a position halfway between A and C (von Tschermak-Seysenegg 1952; Werner 1937). This phenomenon, called *displacement* or *allelotropia*, again suggests that the dynamic transformations in visual space are not of a local character since the location of entire letters, not to mention their points and lines, can be deformed by the spatial context in which they are placed. The nonlocal nature of visual space extends also to brightness perception, as the following section summarizes.

11. Some influences of perceived depth and form on perceived brightness. The Craik-O'Brien and Cornsweet effects (Cornsweet 1970; O'Brien 1958) show that an object's form, notably its edges or regions of rapid spatial change, can influence its apparent brightness or lightness (Figure 5). Let the luminance profile in Figure 5(a) describe a cross-section of the two-dimensional picture in Figure 5(b). Then the lightness of this picture





Figure 5. In (a), the luminance profile is depicted across a onedimensional ray through the picture in (b). Although the interiors of all the regions have equal luminance, the apparent brightness of the regions is described by (c).

appears as in Figure 5(c). The edges of the luminance profile determine the lightnesses of the adjacent regions by a filling-in process. Although the luminances of the regions are the same except near their edges, the perceived lightnesses of the regions are determined by the brightnesses of their respective edges. This remarkable property is reminiscent of Attneave's cat, since regions of maximum curvature – in the lightness domain – again help to determine how the percept is completed. In the present instance, the filling-in process overrides the visual data rather than merely completing an incomplete pattern.

Hamada (1976, 1980) has shown that this filling-in process is even more paradoxical than was previously thought. He compared the lightness of a uniform background with the lightness of the same uniform background with a less luminous Craik-O'Brien figure superimposed on it. By the usual rules of brightness contrast, the lesser brightness of the Craik-O'Brien figure should raise the lightness of the background as its own lightness is reduced. Remarkably, even the background seems darker than the uniform background of the comparison figure, although its luminance is the same.

Just as form can influence lightness, apparent depth can influence lightness. Figures which appear to lie at the same depth can influence each other's lightness in a manner analogous to that found in a monocular brightness constancy paradigm (Gilchrist 1979).

12. Some influences of perceived brightness on perceived depth. Just as depth can influence brightness estimates, brightness data can influence depth estimates. For example, Kaufman, Bacon, and Barroso (1973) studied stereograms built up from the two monocular pictures in Figure 6(a). When these pictures are viewed through a stereogram, the eyes see the lines at a different depth due to the disparity between the two monocular views. If the stereogram is changed so that the left eye sees the same picture as before, whereas the right eye sees the two pictures superimposed (Figure 6[b]), then depth is still perceived. If both eyes see the same superimposed pictures, then of course no depth is seen. However, if one eve sees the pictures superimposed with equal brightness, whereas the other eye sees the two pictures superimposed, one with less brightness and the other with more, then depth is again seen. In the latter case there is no disparity between the two figures, although there is a brightness difference. How does this brightness difference elicit a percept of depth?

The Kaufman et al. (1973) study raises an interesting possibility. If a binocular brightness difference can cause a depth percept, and if a depth percept can influence perceived length, then a binocular brightness difference should be able to cause a change in perceived length. It is also known that monocular cues can sometimes have effects on perceived length similar to those of binocular cues, as in the corridor and Ponzo illusions. When these two phenomena are combined, it is natural to ask: Under what circumstances can a monocular brightness change cause a change (albeit small) in perceived length? I will return to this question in Part II.

13. The binocular mixing of monocular brightnesses. The Kaufman et al. (1973) result illustrates the fact that brightness information from each eye somehow interacts in a binocular exchange. That this exchange is not simply



Figure 6. Combinations of the two pictures in (a), such as in (b), yield a depth percept when each picture is viewed through a separate eye. Depth can be seen even if the two pictures are combined to yield brightness differences but no disparity differences.

additive is shown by several experiments. For example, let AB on a white field be viewed with the left eye and BC on a white field be viewed with the right eye in such a way that the two Bs are superimposed. Then the B does not look significantly darker than A and C despite the fact that white is the input to the other eye corresponding to these letter positions (Helmholtz 1962). In a similar fashion, closing one eye does not make the world look half as bright despite the fact that the total luminance reaching the two eyes is halved (Levelt 1964; von Tschermak-Seysenegg 1952). This fact recalls the discussion of monocular firing of binocular cells from Section 7.

The subtlety of binocular brightness interactions is further revealed by Fechner's paradox (Hering 1964). Suppose that a scene is viewed through both eyes but that one eye sees it through a neutral filter that attenuates all wavelengths by a constant ratio. The filter does not distort the reflectances, or ratios, of light reaching its eye, but only its absolute intensity. Now let the filtered eye be entirely occluded. Then the scene looks brighter and more vivid despite the fact that less total light is reaching the two eyes, and the reflectances are still the same.

Binocular summation of brightness, in excess of probability summation, can occur when the monocular inputs are suitably matched "within some range, perhaps equivalent to Panum's area . . . stereopsis and summation may be mediated by a common neural mechanism" (Blake, Sloane & Fox 1981). I will suggest below that the coexistence of Fechner's paradox and binocular brightness summation can be explained by properties of binocular feedback exchanges among multiple spatial scales. This explanation provides a theoretical framework in which recent studies and models of interactions between binocular brightness summation and monocular flashes can be interpreted (Cogan, Silverman & Sekuler 1982).

Wallach and Adams (1954) have shown that if two figures differ only in terms of the reflectance of one region, then an effect quite the opposite of summation may be found. A rivalrous perception of brightness can be generated in which one shade, then the other, is perceived rather than a simultaneous average of the two shades. I will suggest below that this rivalry phenomenon may be related to the possibility that two monocular figures of different lightness may generate different spatial scales and thereby create a binocular mismatch.

Having reviewed some data concerning the mutual interdependence and lability of depth, form, and lightness judgments, I will now review some obvious visual facts that seem paradoxical when placed beside some of the theoretical ideas that are in vogue at this time. I will also point out that some popular and useful theoretical approaches are inherently limited in their ability to explain either these paradoxes or the visual interactions summarized above.

14. The insufficiency of disparity computations. It is a truism that the retinal images of objects at optical infinity have zero disparity, and that as an object approaches an observer, the disparities on the two retinas of corresponding object points tend to increase. This is the commonplace reason for assuming that larger disparities are an indicator of relative closeness. Julesz stereograms (Julesz 1971) have moreover provided an elegant paradigm wherein disparity computations are a sufficient

indicator of depth, since each separate Julesz random dot picture contains no monocular form cues, yet statistically reliable disparities between corresponding random dot regions yield a vivid impression of a form hovering in depth.

This stunning demonstration has encouraged a decade of ingenious neural modeling. Sperling (1970) introduced important pioneering concepts and equations in a classic paper that explains how cooperation within a disparity plane and competition between disparity planes can resolve binocular ambiguities. These ideas were developed into an effective computational procedure in Dev (1975) which led to a number of mathematical and computer studies (Amari & Arbib 1977; Marr & Poggio 1976). Due to these historical considerations, I will henceforth call models of this type Sperling–Dev models.

All Sperling–Dev models assume that corresponding to each small retinal region there exist a series of disparity detectors sensitive to distinct disparities. These disparity detectors are organized in sheets such that cooperative effects occur between detectors of like disparity within a sheet, whereas competitive interactions occur between sheets. The net effect of these interactions is to suppress spurious disparity correlations and to carve out connected regions of active disparity detectors within a given sheet. These active disparity regions are assumed to correspond to a depth plane of the underlying retinal regions. Some investigators have recently expressed their enthusiasm for this interpretation by commiting the homuncular fallacy of drawing the depth planes in impressive threedimensional figures which carry the full richness of the monocular patterns, although within the model the monocular patterns do not differentially parse themselves among the several sheets of uniformly active disparity detectors.

That something is missing from these models is indicated by the following considerations. The use of a stereogram composed of two separate pictures does not always approximate well the way two eyes view a single picture. When both eyes focus on a single point within a patterned planar surface viewed in depth, the fixation point is a point of minimal binocular disparity. Points increasingly far from the fixation point have increasingly large binocular disparities. Why does such a plane not recede toward optical infinity at the fixation point and curve toward the observer at the periphery of the visual field? Why does the plane not get distorted in a new way every time our eyes fixate on a different point within its surface? If disparities are a sufficient indicator of depth, then how do we ever see planar surfaces? Or even rigid surfaces?

This insufficiency cannot be escaped just by saying that an observer's spatial scales get bigger as retinal eccentricity increases. To see this, let a bounded planar surface have an interior which is statistically uniform with respect to an observer's spatial scales (in a sense that will be precisely defined in Part II). Then the interior disparities of the surface are ambiguous. Only its boundary disparities supply information about the position of the surface in space. Filling-in between these boundaries to create a planar impression is not just a matter of showing that the same disparity, even after an eccentricity compensation, can be locally computed at all the interior points, because an unambiguous disparity computation cannot be carried out at the interior points. The issue is not just whether the observer can estimate the depth of the planar surface, but also how the observer knows that a planar surface is being viewed.

This problem is hinted at even when Julesz stereograms are viewed. Staring at one point in the stereogram results in the gradual loss of depth (Kaufman 1974). Also, in a stereogram composed of three vertical lines to the left eye and just the two flanking lines to the right eye, the direction of depth of the middle line depends on whether the left line or the right line is fixated (Kaufman 1974). This demonstration makes the problem of perceiving planes more severe for any theory which restricts itself to disparity computations, since it shows that depth can depend on the fixation points. What is the crucial difference between the way we perceive the depths of lines and planes? Kaufman (1974) seems to have had this problem in mind when he wrote that "all theories of stereopsis are really inconsistent with the geometry of stereopsis" (p. 320).

Another problem faced by Sperling–Dev models is that they cannot explain effects of perceived depth on perceived size and lightness. The attractive property that the correct depth plane fills-in with uniform activity due to local cooperativity creates a new problem: How does the uniform pattern of activity within a disparity plane rejoin the nonuniformly patterned monocular data to influence its apparent size and lightness?

Finally, there is the problem that only a finite number of depth planes can exist in a finite neural network. Only a few such depth planes can be inferred to exist by joining data relating spatial scales to perceived depth – such as the Kaufman (1974) and Kulikowski (1978) data summarized in Section 6 – to spatial frequency data which suggest that only a few spatial scales exist (Graham 1981; Wilson & Bergen 1979). Since only one depth plane is allowed to be active at each time in any spatial position in a Sperling–Dev model, apparent depth should discretely jump a few times as an observer approaches an object. Instead, apparent depth seems to change continuously in this situation.

15. The insufficiency of Fourier models. An approach with a strong kernel of truth but a fundamental predictive limitation is the Fourier approach to spatial vision. The kernel of truth is illustrated by threshold experiments with four different types of visual patterns (Graham 1981; Graham and Nachmias 1971). Two of the patterns are gratings which vary sinusoidally across the horizontal visual field with different spatial frequencies. The other two are the sum and difference patterns of the first two. If the visual system behaved like a single channel wherein larger peak-to-trough pattern intensities were more detectable, the compound patterns would be more detectable than the sinusoidal ones. In fact, all the patterns are approximately equally detectable. A model in which the different sinusoidal spatial frequencies are independently filtered by separate spatial channels or scales fits the data much better. Recall from Section 6 some of the other data that also suggest the existence of multiple scales.

A related advantage of the multiple channel idea is that one can filter a complex pattern into its component spatial frequencies, weight each component with a factor that mirrors the sensitivity of the human observer to that channel, and then resynthesize the weighted pattern and



Figure 7. When the Cornsweet profile (a) and the rectangle (b) are filtered in such 'a way that low spatial frequencies are attenuated, both outputs look like a Cornsweet profile rather than a rectangle, as occurs during visual experience.

compare it with an observer's perceptions. This modulation transfer function approach has been used to study various effects of boundary edges on interior lightnesses (Cornsweet 1970). If the two luminance profiles in Figure 7 are filtered in this way, they both generate the same output pattern because the human visual system attenuates low spatial frequencies. Unfortunately, both output patterns look like a Cornsweet profile, whereas actually the Cornsweet profile looks like a rectangle. This is not a minor point, since the interior regions of the Cornsweet profile have the same luminance, which is false in the rectangular figure.

This application of the Fourier approach seems to me to be misplaced, since the Fourier transform is linear, whereas a reflectance computation must involve some sorts of ratios and is therefore inherently nonlinear.

The Fourier scheme is also a feedforward transformation of an input pattern into an output pattern. It cannot in principle explain how apparent depth alters apparent length and brightness, since such computations depend on a feedback exchange between monocular data to engender binocular responses. In particular, the data reviewed in Section 4 show that the very definition of a length scale can remain ambiguous until it is embedded in a binocular feedback scheme. The Fourier transform does not at all suggest why length estimates should be so labile. The multiple channel and sensitivity notions need to be explicated in a different formal framework.

16. The insufficiency of linear feedforward theories. The above criticisms of the Fourier approach to spatial vision hold for all computational theories that are based on linear and feedforward operations. For example, some recent workers in artificial intelligence (Marr & Hildreth 1980) compute a spatial scale by first linearly smoothing a pattern with respect to a Gaussian distribution and then computing an edge by setting the Laplacian (the second derivatives) of the smoothed pattern equal to zero (Figure 8). The use of the Laplacian to study edges goes back at least to the time of Mach (Ratliff 1965). The Laplacian is time-honored, but it suffers from limitations that become more severe when its zero-crossings are made the centerpiece of a theory of edges.

One of many difficulties is that zero-crossings compute only the position of an edge and not other related properties such as the brightness of the pattern near the edge. Yet the Cornsweet and Craik–O'Brien figures pointedly show that the brightnesses of edges can strongly influence the lightness of their enclosed forms. Something more than zero-crossings is therefore needed to understand



Figure 8. When a unit step in intensity (a) is smoothed by a Gaussian kernel, the result is (b). The first spatial derivative is (c), and the second spatial derivative is (d). The second derivative is zero at the location of the edge.

spatial vision. The zero-crossing computation itself does not disclose what is missing, so its advocates must guess what is needed. Marr and Hildreth (1980) guess that factors like position, orientation, contrast, length, and width should be computed at the zero-crossings. These guesses do not follow from their definition - or their computation - of an edge. Such properties lie beyond the implications of the zero-crossing computation, because this computation discards essential features of the pattern near the zero-crossing location. Even if the other properties are added to a list of data that is stored in computer memory, this list distorts - indeed entirely destroys - the intrinsic geometric structure of the pattern. The replacement of the natural internal geometrical relationships of a pattern by arbitrary numerical measures of the pattern prevents the Marr and Hildreth (1980) theory from understanding how global processes, such as filling-in, can spontaneously occur in a physical setting. Instead, the Marr and Hildreth (1980) formulation leads to an approach wherein all the intelligence of what to do next rests in the investigator rather than in the model. This restriction to local, investigator-driven computations is due not only to the present state of their model's development, but also to the philosophy of these workers, since Marr and Hildreth write (1980, p. 189): "The visual world is not constructed of ripply, wave-like primitives that extend and add together over an area." Finally, because their theory is linear, it cannot tell us how to estimate the lightnesses of objects, and because their theory is feedforward, it cannot say how apparent depth can influence the apparent size and lightness of monocular patterns.

17. The filling-in dilemma: To have your edge and fill-in too. Any linear and feedforward approach to spatial vision is in fact confronted with the full force of the *filling-in dilemma*: If spatial vision operates by first attenuating all but the edges in a pattern, then how do we ever arrive at a percept of rigid bodies with ample interiors, which are after all the primary objects of perception? How can we have our edges and fill-in too? How does the filling-in process span retinal areas which far exceed the spatial



Figure 9. In this luminance profile, zero-crossings provide no information about which regions are brighter than others. Auxiliary computations are needed to determine this.

bandwidths of the individual receptive fields that physically justify a Gaussian smoothing process? In particular, in the idealized luminance profile in Figure 9, after the edges are determined by a zero-crossing computation, the directions in which to fill-in are completely ambiguous without further computations tacked on. I will argue in Part II of this article that a proper definition of edges does not require auxiliary guesswork.

I should emphasize what I do not mean by a solution to the filling-in dilemma. It is not sufficient to say that edge outlines of objects constitute sufficient information for a viewer to understand a three-dimensional scene. Such a position merely says that observers can use edges to arrive at object percepts, but not how they do so. Such a view begs the question. It is also not sufficient to say that feedback expectancies, or hypotheses, can use edge information to complete an object percept. Such a view does not say how the feedback expectancies were learned, notably what substrate of completed form information was sampled by the learning process, and it also begs the question. Finally, it is inadequate to say that an abstract reconstruction process generates object representations from edges if this process would require a homunculus for its execution in real time.

Expressed in another way, the filling-in dilemma asks: If it is really so hard for us to find mechanisms which can spontaneously and unambiguously fill-in between edges, then do we not have an imperfect understanding of why the nervous system bothers to compute edges? Richards and Marr (1981) suggest that the edge computation compresses the amount of data which needs to be stored. This sort of memory load reduction is important in a computer program, but I will suggest in Part II that it is not a ratelimiting constraint on the brain design which grapples with binocular data. I will suggest, in contrast, that the edge computation sets the stage for processes which selectively amplify and fill-in among those aspects of the data which are capable of matching monocularly, binocularly, or with learned feedback expectancies, as the case might be. This conclusion will clarify both why it is that edge extraction is such an important step in the processing of visual patterns, in partial support of recent models (Marr & Hildreth 1980; Marr & Poggio 1979), and yet edge preprocessing is just one stage in the nonlinear feedback interactions that are used to achieve a coherent visual percept.

## Part II

18. Edges and fixations: The ambiguity of statistically uniform regions. The remainder of this article will outline the major concepts that are needed to build up my theory of these nonlinear interactions. I will also indicate how these concepts can be used to qualitatively interrelate data properties that often cannot be related at all by

alternative theoretical approaches. Many of these concepts are mathematical properties of the membrane equations of neurophysiology, which are the foundation of all quantitative neurophysiological experimentation. The theory provides an understanding of these equations in terms of their computational properties. When the membrane equations are used in suitably interconnected networks of cells, a number of specialized visual models are included as special cases. The theory thereby indicates how these models can be interrelated within a more general, physiologically based, computational framework. Due to the scope of this framework, the present article should be viewed as a summary of an ongoing research program, rather than as a completely tested visual theory. Although my discussion will emphasize the meaning and qualitative reasons for various data from the viewpoint of the theory, previous articles about the theory will be cited for those who wish to study mathematical proofs or numerical simulations, and Appendix A describes a system that is currently being numerically simulated to study binocular filling-in reactions.

I will motivate my theoretical constructions with two simple thought-experiments. I will use these experiments to remind us quickly of some important relationships between perceived depth and the monocular computation of spatial nonuniformities.

Suppose that an observer attempts to fixate a perceptually uniform rectangle hovering in space in front of a discriminable but perceptually uniform background. How does the observer know where to fixate the rectangle? Even if each of the observer's eyes independently fixates a different point of the rectangle's interior, both eyes will receive indentical input patterns near their fixation points due to the rectangle's uniformity. The monocular visual patterns near the fixation points *match* no matter how disparately the fixation points are chosen within the rectangle.

Several conclusions follow from this simple observation. Binocular visual matching between spatially homogeneous regions contains no information about where the eyes are pointed, since all binocular matches between homogeneous regions are equally good no matter where the eyes are pointed. The only binocular visual matches which stand out above the baseline of ambiguous homogeneous matches across the visual field are those which correlate spatially nonuniform data to the two eyes. However, the binocular correlations between these nonuniform patterns, notably their disparities, depend upon the fixation points of the two eyes. Disparity information by itself is therefore insufficient to determine the object's depth. Instead, there must exist an interaction between vergence angle and disparity information to determine where an object is in space (Foley 1980; Grossberg 1976; Marr & Poggio 1979; Sperling 1970).

This binocular constraint on resolving the ambiguity of where the two eyes are looking is one reason for the monocular extraction of the edges of a visual form and the attendant suppression of regions which are spatially homogeneous with respect to a given spatial scale. Without the ability to know where the object is in space, there would be little evolutionary advantage in perceiving its solidity or interior. In this limited sense, edge detection is more fundamental than form detection in dealing with the visual environment.

Just knowing that a feedback loop must exist between motor vergence and sensory disparities does not determine the properties of this loop. Sperling (1970) has postulated that vergence acts to minimize a global disparity measure. Such a process would tend to reduce the perception of double images (Kaufman 1974). I have suggested (Grossberg 1976b) that good binocular matches generate an amplification of network activity, or a binocular resonance. An imbalance in the total resonant output from each binocular hemifield may be an effective vergence signal leading to hemifield-symmetric resonant activity which signifies good binocular matching and stabilizes the vergence angle. The theoretical sections below will suggest how these binocular resonances also compute coherent depth, form, and lightness information.

19. Object permanence and multiple spatial scales. The second thought-experiment reviews a use for multiple spatial scales, rather than a single edge computation, corresponding to each retinal point. Again, our conclusions can be phrased in terms of the fixation process.

As a rigid object approaches an observer, the binocular disparities between its nonfixated features increase proportionally. In order to achieve a concept of object permanence, and at the very least to maintain the fixation process, mechanisms capable of maintaining a high correlation between these progressively larger disparities are needed. The largest disparities will, other things being equal, lie at the most peripheral points on the retina. The expansion of spatial scales with retinal eccentricity is easily rationalized in this way (Hubel & Wiesel 1977; Richards 1975; Schwartz 1980).

It does not suffice, however, to posit that a single scale exists at each retinal position such that scale size increases with retinal eccentricity. This is because objects of different size can approach the observer. As in the Holway and Boring (1941) experiments, objects of different size can generate the same retinal image if they lie at different distances. If these objects possess spatially uniform interiors, then the boundary disparities of their monocular retinal images carry information about their depth. Because all the objects are at different depths, these distinct disparities need to be computed with respect to that retinal position in one eye that is excited by all the objects' boundaries. Multiple spatial scales corresponding to each retinal position can carry out these multiple disparity computations. I will now discuss how the particular scales which can binocularly resonate to a given object's monocular boundary data thereupon fill-in the internal homogeneity of the object's representation with length and lightness estimates, as well as the related question of how monocular cues and learned expectancies can induce similar resonances and thus a perception of depth.

20. Cooperative versus competitive binocular interactions. One major difference between my approach to these problems and alternative approaches is the following: I suggest that a *competitive* process, not a *cooperative* process, defines a depth plane. The cooperative process that other authors have envisaged leads to sheets of network activity which are either off or maximally on. The competitive process that I posit can sustain quantized patterns of activity that reflect an object's perceived

depth, lightness, and length. In other words, the competitive patterns do not succumb to a homuncular dilemma. They are part of the representation of an object's binocular form. The cells that subserve this representative process are sensitive to binocular disparities, but they are not restricted to disparity computations. In this sense, they do not define a "depth plane" at all.

One reason that other investigators have not drawn this conclusion is because a *binary code* hypothesis is often explicit (or lurks implicitly) in their theories. The intuition that a depth plane can be perceived seems to imply cooperation, because in a binary world competition implies an *either-or* choice, which is manifestly unsuitable, whereas cooperation implies an *and* conjunction, which is at least tolerable. In actuality, a binary either-or choice does not begin to capture the properties of a competitive network. Mathematical analysis is needed to understand these properties. (I should emphasize at this point that cooperation and cooperativity are not the same notion. Both competitive and cooperative networks exhibit cooperativity, in the sense in which this word is casually used.)

A large body of mathematical results concerning competitive networks has been discovered during the past decade (Ellias & Grossberg 1975; Grossberg 1970a; 1972d; 1973; 1978a; 1978c; 1978d; 1978e; 1980a; 1980b; 1981; Grossberg & Levine 1975; Levine & Grossberg 1976). These results clarify that not all competitive networks enjoy the properties that are needed to build a visual theory. Certain competitive networks whose cells obey the membrane equations of neurophysiology do have desirable properties. Such systems are called shunting networks to describe the multiplicative relationship between membrane voltages and the conductance changes that are caused by network inputs and signals. This multiplicative relationship enables these networks to automatically retune their sensitivity in response to fluctuating background inputs. Such an automatic gain control capacity implies formal properties that are akin to reflectance processing, Weber law modulation, sensitivity shifts in response to different backgrounds, as well as other important visual effects. Most other authors have worked with additive networks, which do not possess the automatic gain control properties of shunting networks. Sperling (1970; 1981) and Sperling and Sondhi (1968) are notable among other workers in vision for understanding the need to use shunting dynamics, as opposed to mere equilibrium laws of the form  $I(A + J)^{-1}$ . However, these authors did not develop the mathematical theory far enough to have at their disposal some formal properties that I will need. A review of these and other competitive properties is found in Grossberg (1981, Sections 10–27). The sections below build up concepts leading to binocular resonances.

21. Reflectance processing, Weber law modulation, and adaptation level in feedforward shunting competitive networks. Shunting competitive networks can be derived as the solution of a processing dilemma that confronts all cellular tissues, the so-called *noise-saturation dilemma* (Grossberg 1973; 1978e). This dilemma notes that accurate processing both of low activity and high activity input patterns can be prevented by sensitivity loss due to noise (at the low activity end) and saturation(at the high activity end) of the input spectrum. Shunting competitive networks overcome this problem by enabling the cells to retune their sensitivity automatically as the overall background activity of the input pattern fluctuates through time. This result shows how cells can adapt their sensitivity to input patterns that fluctuate over a dynamical range that is much broader than the output range of the cells.

As I mentioned above, the shunting laws take the form of the familiar membrane equations of neurophysiology in neural examples. Due to the generality of the noisesaturation dilemma, formally similar laws should occur in nonneural cellular tissues. I have illustrated in Grossberg (1978b) that some principles which occur in neural tissues also regulate nonneural developmental processes for similar computational reasons.

The solution of the noise-saturation dilemma that I will review herein describes intercellular tuning mechanisms. Data describing intracellular adaptation have also been reported (Baylor and Hodgkin 1974; Baylor, Hodgkin & Lamb 1974a; 1974b) and have been quantitatively fitted by a model in which visual signals are multiplicatively gated by a slowly accumulating transmitter substance (Carpenter and Grossberg 1981). The simplest intercellular mechanism describes a competitive feedforward network in which the activity, or potential,  $x_i(t)$  of the *i*<sup>th</sup> cell (population)  $v_i$  in a field of cells  $v_1$ ,  $v_2, \ldots, v_n$  responds to a spatial pattern  $I_i(t) = \theta_i I(t)$  of inputs  $i = 1, 2, \ldots, n$ . A collection of inputs comprises a spatial pattern if each input has a fixed relative size (or reflectance)  $\theta_i$  but a possibly variable background intensity I(t) (due, say, to a fluctuating light source). The convention that  $\sum_{k=1}^{n} \theta_k = 1$  implies that I(t) is the total

input to the field; viz.  $I(t) = \sum_{k=1}^{n} I_k(t)$ . The simplest

law which solves the noise-saturation dilemma describes the net rate  $(dx_i)/(dt)$  at which sites at  $v_i$  are activated and/or inhibited through time. This law takes the form:

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i - (x_i + C)\sum_{k \neq i} I_k$$
(1)

 $i = 1, 2, \ldots, n$  where  $B > 0 \ge -C$  and  $B \ge x_i(t) \ge -C$ for all times  $t \ge 0$ . Term  $-Ax_i$  describes the spontaneous decay of activity at a constant rate -A. Term  $(B - x_i)I_i$ describes the activation due to an excitatory input  $I_i$  in the *i*<sup>th</sup> channel (Figure 10). Term  $-(x_i + C) \sum_{k \ne i} I_k$  describes the inhibition of activity by competitive inputs

 $\sum_{k\neq i} I_k \text{ from the input channels other than } v_i.$ 

In the absence of inputs (namely all  $I_i = 0$ , i = 1, 2, ..., n), the potential decays to the equilibrium potential 0 due to the decay term  $-AX_i$ . No matter how intense the chosen inputs  $I_i$ , the potential  $x_i$  remains between the values B and -C at all times because  $(B - x_i)I_i = 0$  if  $x_i = B$  and  $-(x_i + C) \sum_{k \neq i} I_k = 0$  if  $x_i = -C$ . That is why B is

called an excitatory saturation point and -C is called an inhibitory saturation point. When  $x_i > 0$ , the cell  $v_i$  is said to be depolarized. When  $x_i < 0$ , the cell  $v_i$  is hyperpolarized. The cell can be hyperpolarized only if C > 0 since  $x_i(t) \ge -C$  at all times t.



Figure 10. In the simplest feedforward competitive network, each input  $I_i$  excites its cell (population)  $v_i$  and inhibits all other populations  $v_j$ ,  $j \neq i$ . (From Grossberg 1978e.)

Before noting how system (1) solves the noise-saturation dilemma, I should clarify its role in the theory as a whole. System (1) is part of a mathematical classification theory wherein a sequence of network variations on the noise-saturation theme is analyzed. The classification theory characterizes how changes in network parameters (for example, decay rates or interaction rules) alter the transformation from input pattern  $(I_1, I_2, \ldots, I_n)$  to activity pattern  $(x_1, x_2, \ldots, x_n)$ . The classification theory thereby provides useful guidelines for designing networks to accomplish specialized processing tasks. The inverse process of inferring which network can generate prescribed data properties is also greatly facilitated. In the present case of system (1), a feedforward flow of inputs to activities occurs wherein a narrow on-center of excitatory input  $(\text{term } (B - x_i)I_i)$  is balanced against a broad offsurround of inhibitory inputs (term  $-(x_i + C) \sum_{k \neq i} I_k$ ). Deviations from these hypotheses will generate network properties that differ from those found in system (1), as I will note in subsequent examples.

To see how system (1) solves the noise-saturation dilemma, let the background input I(t) be held steady for a while. Then the activities in (1) approach equilibrium. These equilibrium values are found by setting  $dx_i/dt = 0$ in (1). They are

$$\mathbf{x}_{i} = \frac{(\mathbf{B} + \mathbf{C})\mathbf{I}}{\mathbf{A} + \mathbf{I}} \left( \boldsymbol{\theta}_{i} - \frac{\mathbf{C}}{\mathbf{B} + \mathbf{C}} \right)$$
(2)

Equation (2) exhibits four main features:

a) Factorization and automatic tuning of sensitivity.-Term  $\theta_i - C/(B+C)$  depends on the *i*<sup>th</sup> reflectance  $\theta_i$ of the input pattern. It is independent of the background intensity *I*. Formula (2) factorizes information about reflectance from information about background intensity. Due to the factorization property,  $x_i$  remains proportional to  $\theta_i - C/(B+C)$  no matter now large *I* is chosen to be. In other words,  $x_i$  does not saturate.

b) Adaptation level, featural noise suppression, and symmetry-breaking.—Output signals from cell  $v_i$  are emitted only if the potential  $x_i$  is depolarized. By (1),  $x_i$  is depolarized only if term  $\theta_i - C/(B + C)$  is positive. Because the reflectance  $\theta_i$  must exceed C/(B + C) to depolarize  $x_i$ , term C/(B + C) is called the *adaptation level*. The size of the adaptation level depends on the ratio of C to B. Typically  $B \ge C$  in vivo, which implies that  $C/(B + C) \le 1$ . Were not C/(B + C) < 1, no choice of  $\theta_i$  could depolarize the cell, since  $\theta_i$ , being a ratio, never exceeds 1.

The most perfect choice of the ratio of C to B is C/B =

1/(n-1) since then C/(B+C) = 1/n. In this case, any uniform input pattern  $I_1 = I_2 = \ldots = I_n$  is suppressed by the network because then all  $\theta_i = 1/n$ . Since also C/(B + C) = 1/n, all  $x_i = 0$  given any input intensity. This property is called *featural noise suppression*, or the suppression of zero spatial frequency patterns. Featural noise suppression guarantees that only nonuniform reflectances of the input pattern can ever generate output signals.

The inequality  $B \ge C$  is called a symmetry-breaking inequality for a reason that is best understood by considering the special case when C/B = 1/(n - 1). The ratio 1/(n - 1) is also, by (1), the ratio of the number of cells excited by input  $I_i$  divided by the number of cells inhibited by input  $I_i$ . Noise suppression is due to the fact that the asymmetry of the intercellular on-center off-surround interactions is matched by the asymmetry of the intracellular saturation points. In other words, the symmetry of the network as a whole is "broken" to achieve noise suppression. Any imbalance in this matching of intercellular to intracellular parameters will either increase or decrease the adaptation level and thereby modify the noise suppression property.

This symmetry-breaking property of shunting networks leads to a theory of how on-center off-surround anatomies develop that is different from the one implied by an additive approach, such as a Fourier or Laplacian theory, if only because additive theories do not possess excitatory and inhibitory saturation points. In Grossberg (1978e; 1982e) I suggested how the choice of intracellular saturation points in a shunting network may influence the development of intercellular on-center off-surround connections to generate the correct balance of intracellular and intercellular parameters. An incorrect balance could suppress all input patterns by causing a pathologically large adaptation level. My suggestion is that the balance of intracellular saturation points determines the balance of morphogenetic substances that are produced at the target cells to guide the growing excitatory and inhibitory pathways.

c) Weber-law modulation. – Term  $\theta_i - C/(B + C)$  is modulated by the term  $(B + C)I(A + I)^{-1}$ , which depends only on the background intensity *I*. This term takes the form of a Weber law (Cornsweet 1970). Thus (2) describes Weber law modulation of reflectance processing above an adaptation level.

d) Normalization and limited capacity.-The total activity of the network is

$$x = \sum_{k=1}^{n} x_k = \frac{[B - (n-1)C]I}{A + I}$$
(3)

By (3), x is independent of the number n of cells in the network if either C = 0 or C/(B + C) = 1/n. In every case,  $x \leq B$  no matter how intense I becomes, and B is independent of n. This tendency for total activity not to grow with n is called total activity normalization. Normalization implies that if the reflectance of one part of the input pattern increases while the total input activity remains fixed, then the cell activities corresponding to other parts of the pattern decrease.

Weber law modulated reflectance processing helps to explain aspects of brightness constancy, whereas the normalization property helps to explain aspects of brightness contrast (Grossberg 1981). The two types of property are complementary aspects of the same dynamical process.

22. Pattern matching and multidimensional scaling without a metric. The interaction between reflectance processing and the adaptation level implies that the sum of two mismatched input patterns from two separate input sources will be inhibited by network (1). This is because the mismatched peaks and troughs of the two input patterns will add to yield an almost uniform total input pattern, which will be quenched by the noise suppression property.

By contrast, the sum of two matched input patterns is a pattern with the same reflectances  $\theta_i$  as the individual patterns. The total activity I + J of the summed pattern, however, exceeds the total activities I and J of the individual patterns. Consequently, by (2) the activities in response to the summed pattern are

$$x_i = \frac{(\mathbf{B} + \mathbf{C})(\mathbf{I} + \mathbf{J})}{\mathbf{A} + \mathbf{I} + \mathbf{J}} \left( \theta_i - \frac{\mathbf{C}}{\mathbf{B} + \mathbf{C}} \right)$$
(4)

which exceed the activities in response to the separate patterns. Network activity is thereby amplified in response to matched patterns and attenuated in response to mismatched patterns due to an interaction between reflectance processing, the adaptation level, and Weber law modulation.

The fact that the activity of each cell in a competitive network can depend on how well two input patterns match is of great importance in my theory. Pattern matching is not just a local property of input sizes at each cell. A given cell can receive two different inputs, yet these inputs may be part of perfectly matched patterns, hence the cell activity is amplified. A given cell can receive two identical inputs, yet these inputs may be part of badly mismatched patterns, hence the cell activity is suppressed.

This matching property avoids the homuncular dilemma by being an automatic consequence of the network's pattern registration process. Various models in Artificial Intelligence, by contrast, use a Euclidean distance  $\sum_{k=1}^{n} (I_k - J_k)^2$  or some other metric to compute pattern matches (Klatt 1980; Newell 1980). Such an approach requires a separate processor to compute a scalar distance between two patterns before deciding how to tack the results of this scalar computation back onto the mainstream of computational activity. A metric also misses properties of the competitive matching process which are crucial in the study of spatial vision, as well as in other pattern recognition problems wherein multiple scales are needed to represent the data unambiguously.

In the competitive matching process, a match not only encodes the matched pattern; it also amplifies it. A metric does not encode a pattern, because it is a scalar rather than a vector. A metric does not amplify the matched patterns because it is minimized rather than maximized by a pattern match. Moreover, what is meant by matching differs in a metric and in a shunting network. A metric makes local matches between corresponding input intensities, whereas a network matches reflectances, which depend upon the entire pattern. One could of course use a metric to match ratios of input intensities, but this computation requires an extra homuncular processing step and is insensitive to overall input intensity, which is not true of the network matching mechanism. When the long-range inhibitory term  $\sum_{k\neq i} I_k$  in (1) is replaced by distance-dependent inhibitory interactions, as in equation (22) of Section 24, a global match of patterns is replaced by simultaneous local matches on a spatial scale that varies monotonically with receptive field size.

Although the properties of metric matches are disappointing in comparison to properties of feedforward network matching, they are totally inadequate when compared to properties of feedback network matching. In a feedback context, there is a flexible criterion of matching called the quenching threshold (Section 28). This criterion can be tuned by attentional and other cognitive factors. Furthermore, approximately matched patterns can mutually deform one another into a fused composite pattern via positive feedback signaling (Ellias & Grossberg 1975; Grossberg 1980b). These properties endow the matching process with hysteresis properties that can maintain a match during slow deformations of the input patterns (Fender & Julesz 1967). When matching occurs betwen ambiguous bottom-up input patterns and topdown expectancies, the pattern fusion property can complete the ambiguous data leading to a cognitively mediated percept (Gregory 1966; Grossberg 1980b).

The primary use of network matching in my binocular theory is to show how those spatial scales which achieve the best binocular match of monocular data from the two eyes can resonate energetically, whereas those spatial scales which generate a mismatched binocular interpretation of the monocular data are energetically attenuated. The ease with which these multidimensional scaling effects occur is due to properties that obtain in even the simplest competitive networks. I use the term "multidimensional scaling" deliberately, since similar competitive rules often operate on a higher perceptual and cognitive level (Grossberg 1978e), where metrical concepts have also been used as explanatory tools (Osgood, Suci, and Tannenbaum 1957; Shepard 1980).

An inadequate model of how cell activity reflects matching can limit a theory's predictive range. For example, in a binocular context, I will use this relationship to suggest how several types of data can be related, including the coexistence of Fechner's paradox and binocular brightness summation (Blake et al. 1981), and the choice between binocular fusion and rivalry within a given spatial scale (Kaufman 1974; Kulikowski 1978). A reason for binocular brightness summation is already evident in equation (4). The effects of activities I and J on x, exceed those expected from noninteracting independent detectors, but are less than the sum  $I + \overline{J}$ , as a result of Weber law modulation (Cogan et al. 1982). In a feedback network, the inputs  $I_i$  and  $J_i$  are chosen to be sigmoid, or Sshaped, functions of the network activities at a prior processing stage. The sigmoid signals are needed to prevent the network as a whole from amplifying noise (Section 28). Then (4) is replaced by a nonlinear summation process that clarifies the success of power law and sigmoid summation rules in fitting data about spatial and binocular brightness interactions (Arend, Lange & Sandick 1981; Graham 1981; Grossberg 1981; Legge & Rubin 1981).

23. Weber law and shift property without logarithms. The simple equation (1) has other properties which are worthy of note. These properties describe other aspects of how the network retunes itself in response to changes in background activity.

The simplest consequence of this retuning property is the classical Weber law

$$\frac{\Delta I}{I} = \text{constant} \tag{5}$$

where  $\Delta I$  is the just noticeable increment above a background intensity *I*. The approximate validity of (5) has encouraged the belief that logarithmic processing determines visual sensitivity (Cornsweet 1970; Land 1977), since  $\Delta \log I = (\Delta I)/I$ , despite the fact that the logarithm exhibits unphysical infinities at small and large values of its argument. In fact, Cornsweet (1970) built separate theories of reflectance processing and of brightness perception by using logarithms to discuss reflectances and shunting functions like  $I(A + J)^{-1}$  to discuss brightness. By contrast, shunting equations like (2) join together reflectance processing and brightness processing into a single computational framework.

Power laws have often been used in psychophysics instead of logarithms (Stevens 1959). It is therefore of interest that equation (2) guarantees reflectance processing undistorted by saturation if the inputs  $I_i$  are power law outputs  $I_i = \lambda J_i^p$  of the activities  $J_i$  at a prior processing stage. Reflectance processing is preserved under power law transformations because the form of (2) is left invariant by such a transformation. In particular,

$$\mathbf{x}_{i} = \frac{(\mathbf{B} + \mathbf{C})\mathbf{I}}{\mathbf{A}^{*} + \mathbf{I}} \left( \boldsymbol{\theta}_{i} - \frac{\mathbf{C}}{\mathbf{B} + \mathbf{C}} \right)$$
(6)

where

$$\theta_i = \frac{\Phi_i^p}{\sum\limits_{k=1}^n \Phi_k^p} \tag{7}$$

$$\mathbf{I} = \mathbf{J}\mathbf{P} \tag{8}$$

and

$$A^* = A\lambda^{-1} \left( \sum_{k=1}^{n} \phi_k^p \right)^{-1}$$
 (9)

To show how the Weber law (5) approximately obtains in (2), choose

 $I_1 = K + \Delta I$ , and  $I_2 = I_3 = \ldots = I_n = K$  (10) Then the total input before increment  $\Delta I$  is applied to  $I_1$  is I = nK. By (2),

$$x_1 = \frac{(B+C)(I+\Delta I)}{A+I+\Delta I} \left(\frac{K+\Delta I}{nK+\Delta I} - \frac{C}{B+C}\right)$$
(11)

If  $I \ge \Delta I$  and  $n \ge 1$ , then

$$\frac{\mathbf{K} + \Delta \mathbf{I}}{\mathbf{n}\mathbf{K} + \Delta \mathbf{I}} - \frac{\mathbf{C}}{\mathbf{B} + \mathbf{C}} = \frac{\Delta \mathbf{I}(\mathbf{n} - \mathbf{I})}{\mathbf{I}} \frac{\mathbf{I}}{\mathbf{n} + \mathbf{I}} + \mathbf{D} \approx \frac{\Delta \mathbf{I}}{\mathbf{I}} + \mathbf{D} \quad (12)$$

where 
$$D = 1/n - C/(B + C)$$
. If  $I \ge A$ , then

$$\frac{(\mathbf{B} + \mathbf{C})(\mathbf{I} + \Delta \mathbf{I})}{\mathbf{A} + \mathbf{I} + \Delta \mathbf{I}} \cong \mathbf{B} + \mathbf{C}$$
(13)

Consequently

where

$$x_1 \cong (B + C) \left(\frac{\Delta I}{I} + D\right)$$
 (14)

If  $x_1$  is detectable when it exceeds a threshold  $\Gamma$ , then

$$\frac{\Delta I}{I} \cong W \tag{15}$$

(16)

$$W = \frac{\Gamma}{B+C} - D = constant$$

A more precise version of the Weber law (5) is the shift property. This property says that the region of maximal visual sensitivity shifts without compression as the background off-surround intensity is parametrically increased (Werblin 1971). The shift property obtains when the on-center input  $I_i$  is plotted in logarithmic coordinates despite the fact that (2) does not describe logarithmic processing.

The shift property is important in a multidimensional parallel processing framework wherein changes in the number and intensity of active input sources can fluctuate wildly through time. Given the shift property, one can fix the activity scale (-C, B) and the network's output threshold once and for all without distorting the network's decision rules as the inputs fluctuate through time. A fixed choice of operating range and of output thresholds is impossible in a multidimensional parallel processing theory that is built up from additive processors. If a fixed threshold is selective when m converging input channels are active, then it may not generate any outputs whatsoever when  $n \ll m$  input channels of comparable intensity are active, and may unselectively generate outputs whenever  $n \ge m$  input channels are active. Such a theory needs continually to redefine how big its thresholds should be as the input load fluctuates through time.

To derive the shift property, rewrite (2) as

х

$$_{i} = \frac{(B+C)I_{i} - CI}{A+I}$$
(17)

Also write  $I_i$  in logarithmic coordinates as  $M = \log_e I_i$ , or

$$I_i = e^M$$
, and the total off-surround input as  $L = \sum_{k \neq i} I_k$ .

Then, in logarithmic coordinates, (17) becomes

$$x_i(M,L) = \frac{Be^M - CL}{A + L + e^M}$$
(18)

The question of shift invariance is: Does there exist a shift S such that

$$\mathbf{x}_{i}(\mathbf{M} + \mathbf{S}, \mathbf{L}_{1}) \equiv \mathbf{x}_{i}(\mathbf{M}, \mathbf{L}_{2})$$
(19)

for all M, where S depends only on  $L_1$  and  $L_2$ ? The answer is yes if C = 0 (no hyperpolarization). Then

$$S = \log_{e} \left( \frac{A + L_{1}}{A + L_{2}} \right)$$
(20)

which shows that successively increasing L by linear increments  $\Delta L$  in (18) causes progressively smaller shifts S in (20). In particular, if  $L_1 = (n - 1)\Delta L$  and  $L_2 = n\Delta L$ , then S approaches zero as n approaches infinity. If C >0, then (19) implies that

$$S = \log_{e} \left[ \frac{AB + (B + C)L_{1} + AC(L_{1} - L_{2})e^{-M}}{AB + (B + C)L_{2}} \right] \quad (21)$$

By (21), S depends on M only via term  $AC(L_1 - L_2)e^{-M}$ , which rapidly decreases as M increases. Thus the shift property improves, rather than deteriorates, at the larger intensities M which might have been expected to cause saturation. Moreover, if  $B \ge C$ , as occurs physically, then (20) is approximately valid at all values of  $M \ge 0$ .

24. Edge, spatial frequency, and reflectance processing by the receptive fields of distance-dependent feedforward networks. Equation (1) is based on several assumptions which do not always occur in vivo. It is the task of the mathematical classification theory to test the consequences of modifying these assumptions. One such assumption says that the inhibitory inputs excite all offsurround channels with equal strength, as in term  $-(x_i$ 

+ C)  $\sum_{k \neq i} I_k$  of (1). Another assumption says that only

the *i*<sup>th</sup> channel is excited by the *i*<sup>th</sup> input, as in term  $(B - x_i)I_i$  of (1). In a general feedforward shunting network, both the excitatory and the inhibitory inputs can depend on the distance between cells, as in the feedforward network

$$\frac{\mathrm{d}\mathbf{x}_{i}}{\mathrm{d}\mathbf{t}} = -\mathbf{A}\mathbf{x}_{i} + (\mathbf{B} - \mathbf{x}_{i})\sum_{k=1}^{n}\mathbf{I}_{k}\mathbf{D}_{ki} - \langle \mathbf{x}_{i} + \mathbf{C} \rangle \sum_{k=1}^{n}\mathbf{I}_{k}\mathbf{E}_{ki}$$
(22)

Here the coefficients  $D_{ki}$  and  $E_{ki}$  describe the fall-off with the distance between cells  $v_k$  and  $v_i$  of the excitatory and inhibitory influences, respectively, of input  $I_k$  on cell  $v_i$ .

Equation (22) exhibits variants of all the properties enjoyed by equation (1). These properties follow from the equilibrium activities of (22), namely

$$\mathbf{x}_{i} = \frac{\mathbf{F}_{i}\mathbf{I}}{\mathbf{A} + \mathbf{G}_{i}\mathbf{I}}$$
(23)

where

$$\mathbf{F}_{i} = \sum_{k=1}^{n} \theta_{k} (BD_{ki} - CE_{ki})$$
(24)

$$\mathbf{G}_{\mathbf{i}} = \sum_{\mathbf{k}=1}^{n} \boldsymbol{\theta}_{\mathbf{k}} (\mathbf{D}_{\mathbf{k}\mathbf{i}} + \mathbf{E}_{\mathbf{k}\mathbf{i}})$$
(25)

in response to a sustained input pattern  $I_i = \theta_i I$ ,  $i = 1, 2, \ldots, n$ . See Ellias and Grossberg (1975) and Grossberg (1981) for a discussion of these properties. For present purposes, I will focus on the fact that the noise suppression property in the network (22) implies an edge detection and spatial frequency detection capability in addition to its pattern matching capability.

The noise suppression property in (23) is guaranteed by imposing the inequalities

$$B \sum_{k=1}^{n} D_{ki} \le C \sum_{k=1}^{n} E_{ki}$$
 (26)

 $i = 1, 2, \ldots, n$ . Noise suppression follows from (26) because then all  $x_i \leq 0$  in response to a uniform pattern (all  $\theta_i = 1/n$ ) by (23) and (24). The inequalities (26) say,

just as in Section 21, that there exists a matched symmetry-breaking between the spatial bandwidths of excitatory and inhibitory intercellular signaling and the choice of inhibitory and excitatory intracellular saturation points -C and B, respectively.

A distance-dependent network with the noise suppression property can detect edges and other nonuniform spatial gradients for the following reason. By (26), those cells  $v_i$  which perceive a uniform input pattern within the breadth of their excitatory and inhibitory scales are suppressed by the noise suppression property no matter how intense the pattern activity is (Figure 11). Only those cells which perceive a nonuniform pattern with respect to their scales can generate suprathreshold activity. This is also true in a suitably designed additive network (Ratliff 1965).

When the interaction coefficients  $D_{ki}$  and  $E_{ki}$  of (22) are Gaussian functions of distance, as in  $D_{ki} = D \exp \left[-\mu(k-i)^2\right]$  and  $E_{ki} = E \exp \left[-\nu(k-i)^2\right]$ , then the equilibrium activities  $x_i$  in (23) include and generalize the model of receptive field properties that is currently used to fit a variety of visual data. In particular, the term  $F_i$  in (24) that appears in the numerator of  $x_i$  depends on sums of differences of Gaussians. Difference-of-Gaussian form factors for studying receptive field responses appear in the work of various authors (Blakemore, Carpenter & Georgeson 1970; Ellias and Grossberg 1975; Enroth-Cugell & Robson 1966; Levine & Grossberg 1976; Rodieck & Stone 1965; Wilson & Bergen 1979). At least three properties of (23) can distinguish it from an additive difference-of-Gaussian form factor  $BD_{ki} - CE_{ki}$  in (24) multiplies, or weights, a reflectance  $\theta_k$ , and all the weighted reflectances are Weber-modulated by a ratio of



Figure 11. When the feedforward competitive network is exposed to the pattern in (a), it suppresses both interior and exterior regions of the pattern that look uniform to cells at these pattern locations. The result is the differential amplification of pattern regions which look nonuniform to the network, as in (b).

the background input I to itself. The difference-of-Gaussian receptive field  $BD_{ki} - CE_{ki}$  thereby becomes a weighting term in the reflectance processing of the network as a whole.

The second property is that each difference-of-Gaussian factor  $BD_{ki} - CE_{ki}$  is itself weighted by the excitatory saturation point B and the inhibitory saturation point C of the network, by contrast with a simple difference-of-Gaussian  $D_{ki} - E_{ki}$ . In networks in which zero spatial frequencies are exactly canceled by their receptive fields, the symmetry-breaking inequality  $B \ge C$  of the shunting model predicts that the ratio  $\mu\nu^{-1}$  of excitatory to inhibitory spatial bandwidths should be larger in a shunting theory than in an additive theory.

A third way to distinguish experimentally between additive and shunting receptive field models is to test whether the contrast of the patterned responses changes as a function of suprathreshold background luminance. In an additive theory, the answer is no. In a distancedependent shunting equation such as (23), the answer is yes. This breakdown is numerically and mathematically analysed in Ellias and Grossberg (1975). The ratios which determine  $x_i$  in (23) lead to changes of contrast as the background intensity I increases only because the coefficients  $D_{ki}$  and  $E_{ki}$  are distance-dependent. In a shunting network with a very narrow excitatory bandwidth and a very broad inhibitory bandwidth, the relative sizes of the x, are independent of I. The contrast changes which occur as I increases in the distance-dependent case can be viewed as a partial breakdown of reflectance processing at high I levels due to the inability of inhibitory gain control to compensate fully for saturation effects.

The edge enhancement property of a feedforward competitive network confronts us with the full force of the filling-in dilemma. If only edges can be detected by a network once it is constrained to satisfy, even approximately, such a basic property as noise suppression, then how does the visual system spontaneously fill-in among the edges to generate percepts of solid objects embedded in continuous media?

25. Statistical analysis by structural scales: Edges with scaling and reflectance properties preserved. Before facing this dilemma, I need to review other properties of the excitatory input term  $\sum_{k=1}^{n} I_k D_{ki}$  and the inhibitory input term  $\sum_{k=1}^{n} I_k E_{ki}$  in (22). Let the interaction coefficients  $D_{ki}$  and  $E_{ki}$  be distance-dependent, so that  $D_{ki} = D(|k - i|)$  and  $E_{ki} = E(|k - i|)$  where the functions D(j) and E(j) are decreasing functions of j, such as Gaussians. Then the input terms  $\sum_{k=1}^{n} I_k D_{ki}$  cross-correlate the input pattern  $(I_1, I_2, \ldots, I_n)$  with the kernel D(j). Similarly, the input terms  $\sum_{k=1}^{n} I_k E_{ki}$  cross-correlate the input pattern  $(I_1, I_2, \ldots, I_n)$  with the kernel E(j). These put pattern  $(I_1, I_2, \ldots, I_n)$  with the kernel E(j).

statistics of the input pattern, rather than the input pattern itself, are the local data to which the network reacts. I will call the kernels D(j) and E(j) structural scales of the network to distinguish them from the functional scales that will be defined below. The structural scales perform a statistical analysis of the data before the shunting dynamics further transform these data statistics. Although terms like  $\sum_{n=1}^{n} L D_{n}$  are linear furctions of the

though terms like  $\sum_{k=1}^{n} I_k D_{ki}$  are linear functions of the

inputs  $I_k$ , the inputs are themselves often nonlinear (notably S-shaped or sigmoidal) functions of outputs from prior network stages (Section 28). Thus the statistical analysis of input patterns is in general a nonlinear summation process.

These concepts are elementary, as well as insufficient for our purposes. It is, however, instructive to review how statistical preprocessing of an input pattern influences the network's reaction to patterns more complex than a rectangle, say, a periodic pattern of high spatial frequency bars superimposed on a periodic pattern of low spatial frequency bars (Figure 12[a]). Suppose for definiteness that the excitatory scale D(i) is narrower than the inhibitory scale E(j) to prevent the occurrence of spurious peak splits and multiple edge effects that can occur even in a feedforward network's response to spots and bars of input (Ellias & Grossberg 1975). Then the excitatory structural bandwidth determines a unit length over which input data is statistically pooled, whereas the inhibitory structural bandwidth determines a unit length over which the pooled data of nearby populations are evaluated for their uniformity.



Figure 12. Transitions in the response of a network to a pattern (a) with multiple spatial frequencies progressively alters from (b) through (d) as the structural scales of the network expand.

It is easily seen that a feedforward network in which featural noise suppression holds and whose excitatory bandwidth approximates  $\alpha$  can react to the input pattern with a periodic series of smoothed bumps (Figure 12[b]). By contrast, a network whose excitatory bandwidth equals period  $2\alpha$  but is less than the entire pattern width reacts only to the smoothed edges of the input pattern (Figure 12[c]). The interior of the input pattern is statistically uniform with respect to the larger structural scale, and therefore its interior is inhibited by noise suppression. As the excitatory bandwidth increases further, the smoothed edges are lumped together until the pattern generates a single centered hump, or spot, of network activity (Figure 12[d]). This example illustrates how the interaction of a broad structural scale with the noise suppression mechanism can inhibit all but the smoothed edges of a finely and regularly textured input pattern. After inhibition takes place, the spatial breadth of the surviving edge responses depends on both the input texture and the structural scale; the edges have not lost their scaling properties. The peak height of these edge responses compute a measure of the pattern's reflectances near its boundary, since ratios of input intensities across the network determine the steady-state potentials  $x_i$  in (23). Rather than discard these monocular scaling and lightness properties, as in a zero-crossing computation, I will use them in an essential way below as the data with which to build up binocular resonances.

26. Correlation of monocular scaling with binocular fusion. The sequence of activity patterns in Figure 12(b), 12(c), and 12(d) is reversed when an observer steadily approaches the picture in Figure 12(a). Then the spot in Figure 12(d) bifurcates into two boundary responses, which in turn bifurcate into a regular pattern of smoothed bumps, which finally bifurcate once again to reveal the high frequency components within each bump. If the picture starts out sufficiently far away from the observer, then the first response in each of the observer's spatial scales is a spot, and the bifurcations in the spot will occur in the same order. However, the distance at which a given bifurcation occurs depends on the spatial scale in question. Other things being equal, a prescribed bifurcation will occur at a greater distance if the excitatory bandwidth of the spatial scale is narrower (high spatial frequency). Furthermore, the registration of multiple spatial frequencies (or even of multiple spots) in the picture will not occur in a spatial scale whose excitatory bandwidth is too broad (low spatial frequency).

The same sequence of bifurcations can occur within the multiple spatial scales corresponding to each eye. If the picture is simultaneously viewed by both eyes, the question naturally arises: How do the two activity patterns within each monocular scale binocularly interact at each distance? Let us assume for the moment, as in the Kaufman (1974) and Kulikowski (1978) experiments, that as the disparity of two monocular patterns increases, it becomes harder for the high spatial frequency scales to fuse them. Since disparity decreases with increasing distance, all scales can binocularly fuse their respective patterns (assuming they are detectable at all) when the distance is great enough, but the lower spatial frequency scales can maintain fusion over a broader range of decreasing distances than can the higher spatial frequency scales. Other things being equal, the scales which can most easily binocularly fuse their two monocular representations of a picture at a given distance are the scales which average away the finer features in the picture. It therefore seems natural to ask: Does the broad spatial smoothing within low spatial frequency scales enhance their ability to binocularly fuse disparate monocular activity patterns?

Having arrived at this issue, we now need to study those properties of *feedback* competitive shunting networks that will be needed to design scale-sensitive binocular resonances in which the fusion event is only one of a constellation of interrelated depth, length, and lightness properties.

27. Noise suppression in feedback competitive networks. The noise-saturation dilemma confronts all cellular tissues which process input patterns, whether the cells exist in a feedforward or in a feedback anatomy. As part of the mathematical classification theory, I will therefore consider shunting interactions in a feedback network wherein excitatory signals are balanced by inhibitory ones. Together, these feedback signals are capable of retuning network sensitivity in response to fluctuating background activity levels.

The feedback analog of the distance-dependent feedforward network (22) is

$$\frac{\mathrm{d}\mathbf{x}_{i}}{\mathrm{d}\mathbf{t}} = -\mathbf{A}\mathbf{x}_{i} + (\mathbf{B} + \mathbf{x}_{i}) \left[ \mathbf{J}_{i} + \sum_{k=1}^{n} \mathbf{f}(\mathbf{x}_{k}) \mathbf{D}_{ki} \right]$$
$$- (\mathbf{x}_{i} + \mathbf{C}) \left[ \mathbf{K}_{i} + \sum_{k=1}^{n} \mathbf{g}(\mathbf{x}_{k}) \mathbf{E}_{ki} \right]$$
(27)

 $i = 1, 2, \ldots, n$ . As in (22), term  $-Ax_i$  describes the spontaneous decay of activity at rate -A. Term  $(B - x_i)J_i$  describes the excitatory effect of the feedforward excitatory input  $J_i$ , which was chosen equal to  $\sum_{k=1}^{n} I_k D_{ki}$  in

(22). Term  $-(x_i + C)K_i$  is also a feedforward term due to inhibition of activity by the feedforward inhibitory input  $K_i$ , which was chosen equal to  $\sum_{k=1}^{n} I_k E_{ki}$  in (22). The

 $X_i$ , which was chosen equal to  $\sum_{k=1}^{n} T_k E_{ki}$  in (22). The

new excitatory feedback term  $\sum_{k=1}^{n} f(x_k)D_{ki}$  describes

the total effect of all the excitatory feedback signals  $f(x_k)D_{ki}$ from the cells  $v_k$  to  $v_i$ . The function  $f(x_i)$  transmutes the activity, or potential, of  $x_i$  into a feedback signal  $f(x_i)$ , which can be interpreted either as a density of spikes per unit time interval or as an electrotonic influence, depending on the situation. The inhibitory feedback term

 $\sum_{k=1}^{n} g(x_k) E_{ki}$  determines the total effect of all the inhibi-

tory feedback signals  $g(x_i)E_{ki}$  from the cells  $v_k$  to  $v_i$ . As in (22), the interaction coefficients  $D_{ki}$  and  $E_{ki}$  are often defined by kernels D(j) and E(j), such that E(j) decreases more slowly than D(j) as a function of increasing values of j.

The problem of noise suppression is just as basic in feedback networks as in feedforward networks. Suppose, for example, that the feedforward inputs and the feedback signals both use the same interneurons and the same statistics of feedback signaling  $(f(x_i) = g(x_i))$  to distribute their values across the network. Then (27) becomes

$$\frac{dx_{i}}{dt} = -Ax_{i} + (B - x_{i})\sum_{k=1}^{n} [I_{k} + f(x_{k})]D_{ki}$$
$$-(x_{i} + C)\sum_{k=1}^{n} [I_{k} + f(x_{k})]E_{ki}$$
(28)

 $i = 1, 2, \ldots, n$ . In such a network, the same criterion of uniformity is applied both to feedforward and to feedback signals. Both processes share the same structural scales. Correspondingly, in (28) as in (22) the single inequality

$$B\sum_{k=1}^{n} D_{ki} \leq C\sum_{k=1}^{n} E_{ki}$$
(26)

suffices to suppress both uniform feedforward patterns and uniform feedback patterns.

28. Sigmoid feedback signals and tuning. Another type of noise suppression, called signal noise suppression, is also needed for a feedback network to function properly. This is true because certain positive feedback functions f(w) can amplify even very small activities w into large activities. Noise amplification due to positive feedback signaling can flood the network with internally generated noise capable of massively distorting the processing of feedforward inputs. Pathologies of feedback signaling have been suggested to cause certain seizures and hallucinations (Ellias & Grossberg 1975; Grossberg 1973; Kaczmarek & Babloyantz 1977).

In Grossberg (1973), I proved as part of the mathematical classification theory that the simplest physically plausible feedback signal which is capable of attenuating, rather than amplifying, small activities is a sigmoid, or Sshaped, signal function (Figure 13). Several remarks should be made about this result.

The comment is sometimes made that you only need a signal *threshold* to prevent noise amplification (Figure 13). This is true, but insufficient, because a threshold signal function does not perform the same pattern transformation as a sigmoid signal function. For example, in a



### ACTIVITY

Figure 13. A sigmoid signal f(w) of cell activity w can suppress noise, contrast enhance suprathreshold activities, normalize total activity, and store the contrast enhanced and normalized pattern in short term memory within a suitably designed feedback competitive network.



Figure 14. In Figures 10(a) and 10(b), the same input pattern is differently transformed and stored in short term memory due to different settings of the network quenching threshold.

shunting network with a narrow on-center and a broad off-surround, a threshold signal chooses the population that receive the largest input for activity storage and suppresses the activities of all other populations. By contrast, a sigmoid signal implies the existence of a quenching threshold (QT). This means that the activities of populations whose initial activation is less than the QT are suppressed, whereas the activity pattern of populations whose initial activities exceed the QT is contrast enhanced before being stored. I identify this storage process with storage in short term memory (STM). In a network that possesses a QT, any operation which alters the QT can sensitize or desensitize the network's ability to store input data (Figure 14). This tuning property is trivialized in a network that chooses the population which receives the largest input for STM storage. In either case, a nonlinear signal function is needed to prevent noise amplification in a feedback network. This fact presents a serious challenge to all linear feedforward models, such as Fourier and Gaussian models.

A proper choice of signal function can be made by mathematically classifying how different signal functions transduce input patterns before they are stored in STM. Consider, for example, the following special case of (28):

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{x}_{i} = -\mathbf{A}\mathbf{x}_{i} + (\mathbf{B} - \mathbf{x}_{i})[\mathbf{I}_{i} + \mathbf{f}(\mathbf{x}_{i})] - \mathbf{x}_{i}\left[\mathbf{J}_{i} + \sum_{\mathbf{k}\neq i} \mathbf{f}(\mathbf{x}_{k})\right], \quad (29)$$

i = 1, 2, ..., n. In (29), the competitive feedback term  $\sum_{i=1}^{n} f(r_i)$  describes long range lateral inhibition just like

 $\sum_{k \neq i} f(x_k) \text{ describes long-range lateral inhibition, just like term } \sum_{k \neq i} I_k \text{ in the feedforward network (1). Network (29)}$ 

strips away all extraneous factors to focus on the following issue. After an input pattern  $(I_1, I_2, \ldots, I_n, J_1, J_2, \ldots, J_n)$  delivered before time t = 0 establishes an initial pattern  $(x_1(0), x_2(0), \ldots, x_n(0))$  in the network's activities, how does feedback signaling within the network transform the initial pattern before it is stored in STM? This problem was solved in Grossberg (1973).

Chart 1 summarizes the main features of the solution. The function  $g(w) = w^{-1}f(w)$  is graphed in Chart 1



Chart 1. Influence of signal function f(w) on input pattern transformation and short term memory storage.

because the property that determines the pattern transformation is whether g(w) is an increasing, constant, or decreasing function at prescribed activities w. For example, a linear f(w) = aw determines a constant g(w) = a; a slower-than-linear  $f(w) = aw(b + w)^{-1}$  determines a decreasing  $g(w) = a(b + w)^{-1}$ ; a faster-than-linear f(w) = $aw^n$ , n > 1, determines an increasing  $g(w) = aw^{n-1}$ ; and a sigmoid signal function  $f(w) = aw^2 (b + w^2)^{-1}$  determines a concave  $g(w) = aw(b + w^2)^{-1}$ . Both linear and slower-than-linear signal functions amplify noise, and are therefore unsatisfactory. Faster-than-linear signal functions, such as power laws with powers greater than one, or threshold rules, suppress noise so vigorously that they make a choice. Sigmoid signal functions determine a QT by mixing together properties of the other types of signal functions.

Another important point is that the QT does not equal the turning point, or manifest threshold, of the sigmoid signal function. The QT depends on all of the parameters of the network. This fact must be understood to argue effectively that the breakdown of any of several mechanisms can induce pathological network properties, such as seizures or hallucinations, by causing the QT to assume abnormally small values. Similarly, an understanding of the factors that control the QT is needed to analyze possible attentional and cognitive mechanisms that can modulate how precise a binocular or bottom-up and topdown match has to be in order to generate fusion and resonance.

A formula for the QT of (29) has been computed when this network is in its short term memory mode (set all inputs  $I_i = J_i = 0$ ). Let the feedback signal function f(w)satisfy

$$f(w) = Cwg(w) \tag{30}$$

where  $C \ge 0$ , g(w) is increasing if  $0 \le w \le x^{(1)}$ , and g(w) = 1 if  $x^{(1)} \le w \le B$ . Thus f(w) grows faster-than-linearly if 0

 $\leq w \leq x^{(1)}$ , linearly if  $x^{(1)} \leq w \leq B$ , and attains a maximum value of C at w = B within the activity interval from 0 to B. The values of f(w) at activities  $w \geq B$  do not affect network dynamics because each  $x_i \leq B$  in (29). It was proved in Grossberg (1973, pp. 355–359) that the QT of (29) is

$$QT = \frac{x^{(1)}}{B - AC^{-1}}$$
(31)

By (31), the QT is not the manifest threshold of f(w), which occurs where g(w) is increasing. Rather, the QT depends on the transition activity where f(w) changes from faster-than-linear to linear, upon the maximum size C of the signal function in the physiological range, upon the number B of excitable sites in each population, and upon the decay rate A.

By (31), an increase in C causes a decrease in the QT. Increasing a shunting signal C that nonspecifically gates all the network's feedback signals can thereby facilitate STM storage. Such a decrease in the QT can facilitate binocular matching by weakening the criterion of how well matched two input patterns need to be in order for some network nodes to supraliminally reverberate in STM. It cannot be overemphasized that this and other desirable tuning properties of competitive feedback networks depend upon the existence of a nonlinear signal function f(w). For example, if f(w) is linear, then  $x^{(1)} = 0$  in (30) and the QT = 0 by (31). Then all positive network activities, no matter how small, can be amplified and stored in STM, including activities due to internal cellular noise.

29. The interdependence of contrast enhancement and tuning. The existence of a QT suggests that the contrast enhancement of input patterns that is ubiquitous in the nervous system is not an end in itself (Ratliff 1965). In feedback competitive shunting networks, contrast enhancement is a mathematical consequence of the signal noise suppression property. This fact is emphasized by the observation that linear feedback signals can perfectly store an input pattern's reflectances – in particular, they do not enhance the pattern – but only at the price of amplifying network noise (Chart 1). Contrast enhancement by a feedback network in its *suprathreshold* activity range follows from noise suppression by the network in its *subthreshold* activity range. Contrast enhancement can intuitively be understood if a feedback competitive network possesses a normalization property like that of a feedforward competitive network (Section 21). If small activities are attenuated by noise suppression and total activity is approximately conserved due to normalization, then large activities will be enhanced.

The simplest example of total activity normalization in a feedback competitive network follows. Consider network (29) in its short term memory mode (all inputs  $I_i = J_i = 0$ ). Let  $x = \sum_{i=1}^{n} x_i$  be the total STM activity and let  $F = \sum_{i=1}^{n} f(x_i)$  be the total feedback signal. Sum over the index *i* in (29) to find that

$$dx/dt = -Ax + (B - x)F$$
(32)

To solve for the possible equilibrium activities of x(t), let dx/dt = 0 in (32). Then

$$\frac{Ax}{B-x} = F$$
(33)

By Chart 1, a network with a faster-than-linear signal function chooses just one activity, say  $x_i$ , for storage in STM. Hence only one summand in F remains positive as time goes on, and its  $x_i(t)$  value approaches that of x(t). Thus (33) can be rewritten as

$$\frac{Ax}{B-x} = f(x) \tag{34}$$

or equivalently

$$\frac{A}{B-x} = g(x) \tag{35}$$

Equation (35) is independent of the number of active cells. Hence the total stored STM activity is independent of the number of active cells. The limiting equation (33) is analyzed for other choices of signal function in Grossberg (1973).

30. Normalization and multistability in a feedback competitive network: A limited capacity short term memory system. Thus suitably designed feedback competitive networks do possess a normalization property. Recall from Section 21 that in a feedforward competitive network, the total activity can increase with the total input intensity but is independent of the number of active cells. This is true only if the inhibitory feedforward interaction  $\sum_{k\neq i} I_k$  in (1) is of long range across the network cells. If the strengths of the inhibitory pathways are weakened or fall off rapidly with distance, then the normalization property is weakened also, and saturation can set in at high input intensities. The same property tends to hold for the feedforward terms  $(B - x_i)J_i$  and  $-(x_i + C)K_i$  of (27).

The normalization property of a feedback competitive network is more subtle (Grossberg 1973; 1981). If such a network is excited to suprathreshold activities and if the exciting inputs are then terminated, then the total activity of the network can approach one of perhaps several positive equilibrium values, all of which tend to be independent of the number of active cells. Thus if the activity of one cell is for some reason increased, then the activities of other cells will decrease to satisfy the normalization constraint unless the system as a whole is attracted to a different equilibrium value. This limited capacity constraint on short term memory is an automatic property in our setting. It is postulated without a mechanistic explanation in various other accounts of short term memory processing (Raaijmakers & Shiffrin 1981, p. 126).

The existence of multistable equilibria in a competitive feedback network is illustrated by equation (35). When f(w) is a faster-than-linear signal function, both  $A(B - x)^{-1}$  and g(x) in (35) are increasing functions of  $x, 0 \le x \le B$ , and g(x) may be chosen so that these functions intersect at arbitrarily many values  $E_1, E_2, \ldots$  of x. Every other value in such a sequence is a possible stable equilibrium point of x, and the remaining values are unstable equilibrium point of w, as when f(w) is a sigmoid signal function, a tendency exists for the suprathreshold equilibria of x to be unique or closely clustered together. These assertions are mathematically characterized in Grossberg (1973).

31. Propagation of normalized disinhibitory cues. Just as in feedforward networks, the feedback normalization property is weakened if the inhibitory path strengths are chosen to decrease more rapidly with distance. Then the normalization property tends to hold among subsets of cells that lie within one bandwidth of the network's inhibitory structural scale. In particular, if some cell activities are enhanced by a given amount, then their neighbors will tend to be suppressed by a comparable amount. The neighbors of these neighbors will then be enhanced by a similar amount, and so on. In this way, a disinhibitory wave can propagate across a network in such a way that each crest of the wave inherits, or "remembers," the activity of the previous crest. This implication of the normalization property in a feedback network with finite structural scales will be important in my account of filling-in. Normalization within a structural scale also endows the network's activity patterns with constancy and contrast patterns, as in the case of feedforward competitive networks (Section 24). In a feedback context, however, constancy and contrast properties can propagate far beyond the confines of a single structural scale because of normalized disinhibitory properties such as those Figure 15 depicts.

**32. Structural versus functional scales.** The propagation process depicted in Figure 15 needs to be understood in greater detail because it will be fundamental in all that follows. A good way to approach this understanding is to compare the reactions of competitive feedforward networks with those of competitive feedback networks to the same input patterns.

Let us start with the simplest case. Choose C = 0 in (22) and (27). This prevents the noise suppression inequalities (26) from holding. Although feedforward and feedback inhibition are still operative, activities cannot be inhibited below zero in this case. Consequently, a uniform input pattern can be attenuated but not entirely sup-



Figure 15. Reaction of a feedforward competitive network (b) and a feedback competitive network (c) to the same input pattern (a). Only the feedback network can activate the interior of the region which receives the input pattern with unattenuated activity.

pressed. Choose a sigmoidal feedback signal function to prevent noise amplification, and thus to contrast-enhance the pattern of suprathreshold activities. These hypotheses enable us to study the main effects of feedback signalling unconfounded by the effect of noise suppression.

What happens when we present a rectangular input pattern (Figure 15[a]) to both networks? Due to the feedforward inhibition in (22), the feedforward network enhances the edges of the rectangle and attenuates its interior (Figure 15[b]). By contrast, the feedback network elicits a regularly spaced series of excitatory peaks across the cells that receive the rectangular input (Figure 15[c]). This type of reaction occurs even if the input pattern is not contrast-enhanced by a feedforward inhibitory stage, as in Figure 15(b), before feedback inhibition can act on the contrast-enhanced pattern. The pattern of Figure 15(c) is elicited even if the feedback acts directly on the rectangular input pattern. Parametric numerical studies of this type of disinhibitory feedback reaction are found in Ellias and Grossberg (1975).

The spatial bandwidth between successive peaks in Figure 15(c) is called the *functional scale* of the feedback network. My first robust points are that a functional scale can exist in a feedback network but not in a feedforward network, and that, although the functional scale is related to the structural scale of a feedback network, the two scales are not identical. I will discuss the functional scale given C = 0 before reinstating the noise suppression inequalities (26) because the interaction between contrast enhancement and noise suppression in a feedback network is a much more subtle issue. 33. Disinhibitory propagation of functional scaling from boundaries to interiors. To see how a functional scale develops, let us consider the network's response to the rectangular input pattern on a moment-to-moment basis. All the populations  $v_m$  that are excited by the rectangle initially receive equal inputs. All the activities  $x_m$  of these populations therefore start to grow at the same rate. This growth process continues until the feedback signals  $f(x_m)D_{mi}$  and  $g(x_m)E_{mi}$  can be registered by the other populations  $v_i$ . Populations  $v_i$  which are near the rectangle's boundary receive smaller total inhibitory signals

 $\sum_{m=1}^{n} g(x_m) E_{mi}$  than populations which lie nearer to the rectangle's center, even when all the rectangle-excited activities  $x_m$  are equal. This is because the interaction strengths  $E_{mi} = E(|m-i|)$  are distance-dependent, and the boundary populations receive no inhibition from contiguous populations that lie outside the rectangle.

As a result of this inhibitory asymmetry, the activities  $x_i$ near the boundary start to grow faster than contiguous activities  $x_j$  nearer to the center. The inhibitory feedback signal  $g(x_i)E_{ij}$  from  $v_i$  to  $v_j$  begins to exceed the inhibitory feedback signal  $g(x_j)E_{ji}$  from  $v_j$  to  $v_i$ , because  $x_i > x_j$  and  $E_{ij}$  $= E_{ji}$ . Thus although all *individual* feedback signals among rectangle-excited populations start out equal, they are soon differentiated due to a second-order effect whereby the boundary bias in the spatial distribution of the *total* inhibitory feedback signals is mediated by the activities of *individual* populations.

As the interior activities  $x_i$  get differentially inhibited, their inhibitory signals  $g(x_j)E_{jk}$  to populations  $v_k$  which lie even deeper within the rectangle's interior become smaller. Now the *total* pattern of inputs plus feedback signals is no longer uniform across the populations  $v_j$  and  $v_k$ . The populations  $v_k$  are favored. Contrast enhancement bootstraps their activities  $x_k$  to larger values. Now these populations can more strongly inhibit neighboring populations that lie even deeper into the rectangle's interior, and the process continues in this fashion.

The boundary asymmetry in the total inhibitory feedback signals hereby propagates ever deeper into the rectangle's interior by a process of distance-dependent disinhibition and contrast enhancement until all the rectangle-excited populations are filled-in by a series of regularly spaced activity peaks as in Figure 11(c).

34. Quantization of functional scales: Hysteresis and uncertainty. As I mentioned in Section 32, two distinct types of spatial scales can be distinguished in a feedback network. The structural scales D(j) and E(j) describe how rapidly the network's feedback interaction coefficients decrease as a function of distance. The functional scale describes the spatial wavelength of the disinhibitory peaks that arise in response to prescribed input patterns. Although these two types of scale are related, they differ in fundamental ways.

They are related because an increase in a network's structural scales can cause an increase in the functional scale with which it fills-in a given input pattern, as in the numerical studies of Ellias and Grossberg (1975). This is due to two effects acting together. A slower decrease of D(j) with increasing distance j can increase the number of contiguous populations that pool excitatory feedback.

This effect can broaden the peaks in the activity pattern. A slower decrease of E(j) with increasing distance j can increase the number of contiguous populations which can be inhibited by an activity peak. This effect can broaden the troughs in the activity pattern. This relationship between structural and functional scales partially supports the intuition that visual processing includes a spatial frequency analysis of visual data (Graham 1981; Robson 1975), because if several feedback networks with distinct structural scales received the same input pattern, then they would each generate distinct functional scales such that smaller structural scales tended to generate smaller functional scales. However, the functional scale does not equal the structural scale, and its properties represent a radical departure from feedforward linear ideas.

The most important of these differences can be summarized as follows. The functional scale is a *quantized* property of the *interaction* between the network and *global* features of an input pattern, such as its length. Unlike a structural scale, a functional scale is not just a property of the network. Nor is it just a property of the input pattern. The interaction between pattern and network literally creates the functional scale. The quantized nature of this interaction is easy to state because it is so fundamental. (The reader who knows some quantum theory, notably Bohr's original model of the hydrogen atom, might find it instructive to compare the two types of quantization.)

The length L of a rectangular input pattern might equal a nonintegral multiple of a network's structural scales, but obviously there can only exist an integral number of disinhibitory peaks in the activity pattern induced by the rectangle. The feedback network therefore quantizes its activity in a way that depends on the global structure of the input pattern. The functional scales must change to satisfy the quantum property as distinct patterns perturb the network, even though the network's structural scales remain fixed.

For example, rectangular inputs of length L,  $L + \Delta L$ , L +  $2\Delta L$ , . . . ,  $L + \omega \Delta L$  might all induce  $M_L$  peaks in the network's activity pattern. Not until a rectangle of length  $L + (\omega + 1)\Delta L$  is presented might the network respond with  $M_L + 1$  peaks. This length quantization property suggests a new reason why a network, and perception, can exhibit hysteresis as an input pattern is slowly deformed through time. This hysteresis property can contribute to, but is not identical with, the hysteresis that is due to persistent binocular matching as a result of positive feedback signaling when two monocular patterns are slowly deformed after first being binocularly matched (Fender & Julesz 1967; Grossberg 1980b). Another consequence of the quantization property is that the network cannot distinguish certain differences between input patterns. Quantization implies a certain degree of perceptual uncertainty.

**35. Phantoms.** The reader might by now have entertained the following objection to these ideas. If percepts really involve spatially regular patterned responses even to uniform input regions, then why don't we easily see these patterns? I suggest that we sometimes do, as when spatially periodic visual phantoms can be seen superimposed upon otherwise uniform, and surprisingly large,

regions (Smith & Over 1979; Tynan & Sekuler 1975; Weisstein, Maguire & Berbaum 1976). The disinhibitory filling-in process clarifies how these phantoms can cover regions which excite a retinal area much larger than a single structural scale. I suggest that we do not see phantoms more often for three related reasons.

During day-to-day visual experience, several functional scales are often simultaneously active. The peaks of higher spatial frequency functional scales can overlay the spaces between lower spatial frequency functional scales. Retinal tremor and other eve movements can randomize the spatial phases of, and thereby spatially smooth, the higher frequency scales across the lower frequency scales through time. Even within a single structural scale, if the boundary of an input pattern curves in two dimensions, then the disinhibitory wavelets can cause interference patterns as they propagate into the interior of the activity pattern along rays perpendicular to each boundary element. These interference patterns can also obscure the visibility of a functional scale. Such considerations clarify why experiments in which visual phantoms are easily seen usually use patterns that selectively resonate with a low spatial frequency structural scale that varies in only one spatial dimension. This suggestion that filling-in by functional scales may subserve phantoms does not imply that the perceived wavelength of a phantom is commensurate with any structural scale of the underlying network. Rather I suggest that once a pattern of functional wavelets is established by a boundary figure, it can quickly propagate by a resonant filling-in reaction into the interior of the figure if the shape of the interior does not define functional barriers to filling-in (Section 40).

An important issue concerning the perception of phantoms is whether they are, of necessity, perceivable only if moving displays are used, or whether the primary effect of moving a properly chosen spatial frequency at a properly chosen velocity is to selectively suppress all but the perceived spatial wavelength via noise suppression. The latter interpretation is compatible with an explanation of spatial frequency adaptation using properties of shunting feedback networks (Grossberg 1980b, Section 12).

A possible experimental approach to seeing functional scales using a stationary display takes the form of a twostage experiment. First adapt out the high spatial frequencies using a spatial frequency adaptation paradigm. Then fixate a bounded display which is large enough and is shaped properly to strongly activate a low spatial frequency scale in one dimension, and which possesses a uniform interior that can energize periodic network activity.

36. Functional length and Emmert's law. Two more important properties of functional scales are related to length and lightness estimates. The functional wavelength defines a length scale. To understand what I mean by this, let a rectangular input pattern of fixed length L excite networks with different structural scales. I hypothesize that the apparent length of the rectangle in each network will depend on the functional scale generated therein. Since a broader structural scale induces a broader functional scale, the activity pattern in such a network will contain fewer active functional wavelengths. I suggest that this property is associated with an

impression of a shorter object, despite the fact that L is fixed.

The reader might object that this property implies too much. Why can a monocularly viewed object have ambiguous length if it can excite a functional scale? I suggest that under certain, but not all, monocular viewing conditions, an object may excite all the structural scales of the observer. When this happens, the object's length may seem ambiguous. I will also suggest in Section 39 how binocular viewing of a nearby object can selectively excite structural scales which subserve large functional scales, thereby making the object look shorter. By contrast, binocular viewing of a far-away object can selectively excite structural scales which subserve small functional scales, thereby making the object look longer. Thus the combination of binocular selection of structural scales that vary inversely with an object's distance, along with the inverse variation of length estimates with functional scales, may contribute to an explanation of Emmert's law.

This view of the correlation between perceived length and perceived distance does not imply that the relationship should be veridical – and indeed sometimes it is not (Hagen & Teghtsoonian 1981) – for the following reasons. The functional scale is a quantized collective property of a nonlinear feedback network rather than a linear ruler. The selection of which structural scales will resonate to a given object and of which functional scales will be generated within these structural scales depends on the interaction with the object in different ways; for one, the choice of structural scale does not depend on a filling-in reaction.

These remarks indicate a sense in which functional scales define an "intrinsic metric," which is independent of cognitive influences but on whose shoulders correlations with motor maps, adaptive chunking and learned feedback expectancy computations can build (Grossberg 1978e; 1980b). This intrinsic metric helps to explain how monocular scaling effects, such as those described in Section 5, can occur. Once the relevance of the functional scale concept to metrical estimates is broached, one can begin to appreciate how a dynamic "tension" or "force field" or "curved metric" can be generated whereby objects which excite one part of the visual field can influence the perception of objects at distant visual positions (Koffka 1935; Watson 1978). I believe that the functional scale concept explicates a notion of dynamic field interactions that escapes the difficulties faced by the Gestaltists in their pioneering efforts to explain global visual interactions.

**37. Functional lightness and the Cornsweet effect.** The functional scale concept clarifies how object boundaries can determine the lightness of object interiors, as in the Cornsweet effect. Other things being equal, a more intense pattern edge will cause larger inhibitory troughs around itself. The inhibitory trough which is interior to the pattern will thereby create a larger disinhibitory peak due to pattern normalization within the structural scale. This disinhibitory process continues to penetrate the pattern in such a way that all the interior peak heights are influenced by the boundary peak height because each inhibitory trough "remembers" the previous peak height. The sensitivity of filled-in interior peak size to

boundary peak size helps to explain the Cornsweet effect (Section 11).

Crucial to this type of explanation is the idea that the disinhibitory filling-in process feeds off the input intensity within the object interior. The reader can now better appreciate why I set C = 0 to start off my exposition. Suppose that a feedforward inhibitory stage acts on an input pattern before the feedback network responds to the transformed pattern. Let the feedforward stage use its noise suppression property to convert a rectangular input pattern into an edge reaction that suppresses the rectangle's interior (Figure 15[b]). Then let the feedback network transform the edge-enhanced pattern. Where does the feedback network get the input energy to fill-in off the edge reactions into the pattern's interior if the interior activities have already been suppressed? How does the feedback network know that the original input pattern had an interior at all? This is the technical version of the "To Have Your Edge and Fill-In Too" dilemma that I raised in Section 17. We are now much closer to an answer.

38. The monocular length-luminance effect. Before suggesting a resolution of this dilemma, I will note a property of functional scales which seems to be reflected in various data, such as the Wallach and Adams (1954) experiment, but seems not to have been studied directly. This property concerns changes in functional scaling that are due to changes in the luminance of an input pattern. To illustrate the phenomenon in its simplest form. I will consider qualitatively the response of a competitive feedback network such as (27) to a rectangular input pattern of increasing luminance. In Figure 16(a) the rectangle intensity is too low to elicit any suprathreshold reaction. In Figure 16(b) a higher rectangle intensity fills-in the region with a single interior peak and two boundary peaks. At the still higher intensity of Figure 16(c), two interior peaks emerge. At successively higher intensities, more peaks emerge until the intensity gets so high that a smaller number of peaks again occurs (Figure 16[d]). This progressive increase followed by a progressive decrease in the number of interior peaks has been found in many computer runs (Cohen & Grossberg 1983a; Ellias & Grossberg 1975). It reflects the network's increasing sensitivity at higher input intensities until such high intensities are reached that the network starts to saturate and is gradually desensitized. The quantitative change in the relative number of peaks is not so dramatic as Figure 16 suggests.

If we assume that the total area under an activity pattern within a unit spatial region estimates the lightness of the pattern, then it is tempting to interpret the above result as a perceived lightness change when the luminance of an object, but not of its background, is parametrically increased. This interpretation cannot be made without extreme caution, however, because the functional scaling change within one monocular representation may alter the ability of this representation to match the other monocular representation within a given structural scale. In other words, by replacing spatially homogeneous regions in a figure by spatially patterned functional scales, we can think about whether these patterns match or mismatch under prescribed conditions. A



Figure 16. Response of a feedback competitive network to a rectangle of increasing luminance on a black background.

change in the scales which are capable of binocular matching implies change in the scales which can energetically resonate. A complex change in perceived brightness, depth, and length may hereby be caused.

Even during conditions of monocular viewing, the phenomenon depicted by Figure 16 has challenging implications. Consider an input pattern which is a figure against a ground with nonzero reflectance. Let the entire pattern be illuminated at successively higher luminances. Within the energy region of brightness constancy, the balance between the functional scales of figure and ground can be maintained. At extreme luminances, however, the sensitivity changes illustrated in Figure 16 can take effect and may cause a coordinated change in both perceived brightness and perceived length. If the functional wavelength, as opposed to a more global estimate of the total activated region within a structural scale, influences length judgments, then a small length reduction may be detectable at both low and high luminances. This effect should at the present time be thought of as an intriguing possibility rather than as a necessary prediction of the theory because, in realistic binocular networks, interactive effects between monocular and binocular cells and between multiple structural scales may alter the properties of Figure 16.

39. Spreading FIRE: Pooled binocular edges, false matches, allelotropia, binocular brightness summation, and binocular length scaling. Now that the concept of a functional scale in a competitive feedback network is clearly in view, I can reintroduce the noise suppression inequalities (26) to show how the joint action of noise suppression and functional scaling can generate a fillingin resonant exchange (FIRE) that is sensitive to binocular properties such as disparity. Within the framework I have built up, starting a FIRE capable of global effects on perceived depth, form, and lightness is intuitively simple. I will nonetheless describe the main ideas in mechanistic terms, since if certain constraints are not obeyed, the FIRE will not ignite (Cohen & Grossberg 1983a). I will also restrict my attention to the simplest, or minimal, network which exhibits the properties that I seek. It will be apparent that the same types of properties can be obtained in a wide variety of related network designs. The equations that have been used to simulate such a FIRE numerically are described in the Appendix.

First I will restrict attention to the case of a single structural scale, which is defined by excitatory and inhibitory kernels D(j) and E(j), respectively. Three main intuitions go into the construction.

Proposition I.-Only input pattern data which are spatially nonuniform with respect to a structural scale are informative (Section 18).

Proposition II.-The ease with which two monocular input patterns of fixed disparity can be binocularly fused depends on the spatial frequencies in the patterns (Section 6 and 8). This dependence is not, however, a direct one. It is mediated by statistical preprocessing of the input patterns using nonlinear cross-correlations, as in Section 25. Henceforth when I discuss an "edge," I will mean a statistical edge rather than an edge within the input pattern itself.

**Proposition III.**-Filling-in a functional scale can only be achieved if there exists an input source on which the FIRE can feed (Section 33).

To fix ideas, let a rectangular input pattern idealize a preprocessed segment of a scene. The interior of the rectangle idealizes an ambiguous region and the boundaries of the rectangle idealize informative regions of the scene with respect to the structural scale in question. A copy of the rectangular input pattern is processed by each monocular representation. Since the scene is viewed from a distance, the two rectangular inputs will excite disparate positions within their respective monocular representations (Figure 17[a]). In general, the more peripheral boundary with respect to the foveal fixation point will correspond to a larger disparity.

Proposition I suggests that the rectangles are passed through a feedforward competitive network capable of noise suppression to extract their statistical edges (Figure 17[b]). Keep in mind that these edges are not zerocrossings. Rather, their breadth is commensurate with the bandwidth of the excitatory kernel D(j) (Section 25). This property is used to realize Proposition II as follows.

Suppose that the edge-enhanced monocular patterns are matched at binocular cells, where I mean matching in the sense of Sections 22 and 24. Because these networks possess distance-dependent structural scales, the suppressive effects of mismatch are restricted to the spatial



Figure 17. After the two monocular patterns (a) are passed through a feedforward competitive network to extract their nonuniform data with respect to the network's structural scales (b), the filtered patterns are topographically matched to allow pooled binocular edges to form (c) if the relationship between disparity and monocular functional scaling is favorable.

wavelength of an inhibitory scale, E(j), rather than involving the entire network. Because the edges are statistically defined, the concepts of match and mismatch refer to the degree of coherence between monocular statistics rather than to comparisons of individual edges. Three possible cases can occur.

The case of primary interest is the one in which the two monocular edge reactions overlap enough to fall within each other's excitatory on-center D(j). This will happen, for example, if the disparity between the edge centers does not exceed half the width of the excitatory on-center. Marr and Poggio (1979) have pointed out that, within this range, the probability of false matches is very small, in fact less than 5%. Within the zero-crossing formalism of Marr and Poggio (1979), however, the decision to restrict matches to this distance is not part of their definition of an edge. In a theory in which the edge computation retains its spatial scale at a topographically organized binocular matching interface, this restriction is automatic.

If this matching constraint is satisfied, then a *pooled* binocular edge is formed that is centered between the loci of the monocular edges (Figure 17[c]). See Ellias and Grossberg (1975, Figure 25) for an example of this shift phenomenon. The shift in position of a pooled binocular edge also has no analog in the Marr and Poggio (1979) theory. I suggest that this binocularly-driven shift is the basis for allelotropia (Section 10).

If the two distal edges fall outside their respective oncenters, but within their off-surrounds, then they will annihilate each other if they enjoy identical parameters, or one will suppress the other by contrast enhancement if it has a sufficient energetic advantage. This unstable competition will be used to suggest an explanation of binocular rivalry in Section 44.

Finally, the two edges might fall entirely outside each other's receptive fields. Then each can be registered at the binocular cells, albeit with less intensity than a pooled binocular edge, due to equations (2) and (4). A double image can then occur. I consider the dependence of intensity on matching to be the basis for binocular brightness summation (Section 13).

The net effect of the above operations is to generate two amplified pooled binocular edges at the boundaries of an ambiguous region if the spatial scale of the network can match the boundary disparities of the region. Networks which cannot make this match are energetically attenuated. Having used disparity (and thus depth) information to select suitable scales and to amplify the informative data within these scales, we must face the filling-in dilemma posed by Proposition III. How do the binocular cells know how to fill-in between the pooled binocular edges to recover a binocular representation of the entire pattern? Where do these cells get the input energy to spread the FIRE? In other words, having used noise suppression to achieve selective binocular matching, how do we bypass noise suppression to recover the form of the object?

If we restrict ourselves to the minimal solution of this problem, then one answer is strongly suggested. Signals from the pooled binocular edges are topographically fed back to the processing stage at which the rectangular input is registered. This is the stage just before the feedforward competitive step that extracts the monocular edges (Figure 18). Several important conclusions follow immediately from this suggestion:



Figure 18. Monocular processing of patterns through feedforward competitive networks is followed by binocular matching of the two transformed monocular patterns. The pooled binocular edges are then fed back to both monocular representations at a processing stage where they can feed off monocular activity to start a FIRE.

1) The network becomes a feedback competitive network in which binocular matching modulates the patterning of monocular representations.

2) If filling-in can occur, a functional scale is defined within this feedback competitive network. A larger disparity between monocular patterns resonates best with a larger structural scale, which generates a larger functional scale. Thus perceived length depends on perceived depth.

3) The activity pattern across the functional scale is constrained by the network's normalization property. Thus perceived depth influences perceived brightness, notably the lightnesses of objects which seem to lie at the same depth.

In short, if we can overcome the filling-in dilemma at all within feedback competitive shunting networks, then known dependencies between perceived depth, length, form, and lightness begin to emerge as natural consequences. I know of no other theoretical approach in which this is true.

It remains to indicate how the FIRE can spread despite the action of the noise suppression inequalities (26). The main problem to avoid is summarized in Figure 19. Figure 19(a) depicts a pooled binocular edge. When this edge adds onto the rectangular pattern, we find Figure 19(b). Here there is a hump on the rectangle. If this pattern is then fed through the feedforward competitive network, a pattern such as that in Figure 19(c) is produced. In other words, the FIRE is quenched. This is because the noise suppression property of feedforward competition drives all activities outside the hump to subthreshold values before the positive feedback loops in the total network can enhance any of these activities.

I have exposed the reader to this difficulty to emphasize a crucial property of pooled binocular edges. If C > 0 in (27), then an inhibitory trough surrounds the edge (Figure 19[d]). (If C is too small to yield a significant trough, then the pooled edge must be passed through another



Figure 19. The FIRE is quenched in (a)-(c) because there exists no nonuniform region off the pooled binocular edge which can be amplified by the feedback exchange. In (d)-(f), the inhibitory troughs of the edges enables the FIRE to propagate.

stage of feedforward competition.) When the edge in Figure 19(d) is added to the rectangular input by a competitive interaction, the pattern in Figure 19(e) is generated. The region off the hump is no longer uniform. The uniform region is separated from the hump by a trough whose width is commensurate with the inhibitory scale E(j). When this pattern is passed through the feedforward competition, Figure 19(f) is generated. The nonuniform region has been contrast-enhanced into a second hump, whereas the remaining uniform region has been annihilated by noise suppression. Now the pattern is fed back to the rectangular pattern stage and the cycle repeats itself. A third hump is thereby generated, and the FIRE rapidly spreads, or "develops," across the entire rectangular region at a rate commensurate with the time it takes to feed a signal through the feedback loop. Since the cells which are excited by the rectangle are already processing the input pattern when the FIRE begins, it can now spread very quickly.

Some further remarks need to be made to clarify how the edge in Figure 19(d) adds to the rectangular input pattern. The inhibited regions in the edge can generate signals only if they excite off-cells whose signals have a net inhibitory effect on the rectangle. This option is not acceptable because mismatched patterns at the binocular matching cells would then elicit FIRES via off-cell signaling. Rather, the edge activities in Figure 19(d) are rectified when they generate output signals. These signals are distributed by a competitive (on-center off-surround) anatomy whose net effect is to add a signal pattern of the shape in Figure 19(d) to the rectangular input pattern. In other words, if all signaling stages of Figure 18 are chosen to be competitive to overcome the noise-saturation dilemma (Section 21), then the desired pattern transformations are achieved. This hypothesis does not necessarily imply that the pathways between the processing stages are both excitatory and inhibitory. Purely excitatory pathways can activate each level's internal on-center offsurround interneurons to achieve the desired effect. From this perspective, one can see that the two monocular edge-extraction stages and the binocular matching stage at the top of Figure 18 can all be lumped into a single binocular edge matching stage. If this is done, then the mechanism for generating FIREs seems elementary indeed. If competitive signaling is used to binocularly match monocular representation and to feed the results back to the monocular representations, then a filling-in reaction will spontaneously occur within the matched scales.

40. Figure-ground separation by filling-in barriers. Now that we have seen how a FIRE can spread, it remains to say how it can be prevented from inappropriately covering the entire visual field. A case in point is the Julesz (1971) 5% solution of dots on a white background in the stereo-gram of Section 9. How do the different binocular disparities of the dots in the "figure" and "ground" regions impart distinct depths to the white backgrounds of these two regions? This is an issue because the same ambiguous white background fills both regions.

I suggest that the boundary disparities of the "figure" dots can form pooled binocular edges in a spatial scale different from the one that best pools binocular edges in the "ground" scale. At the binocular cells of the "ground"

scale, mismatch of the monocular edges of the "figure" can produce an inhibitory trough whose breadth is commensurate with two inhibitory structural wavelengths. The spreading FIRE cannot cross a filling-in barrier (FIB) any more than a forest fire can across a sufficiently broad trench.

Thus, within a scale whose pooled binocular edges can feed off the ambiguous background activity, FIREs can spread in all directions until they run into FIBS. This mechanism does not imply that a FIRE can rush through all spaces between adjacent FIBS, because the functional scale is a coherent dynamic entity that will collapse if the spaces between FIBS, relative to the functional scale, are sufficiently small. Thus a random placement of dots may. other things being equal, form better FIBS than a deterministic placement which permits a coherent flow of FIRE to run between rows of FIBS. A rigorous study of the interaction between (passive) texture statistics and (coherent) functional scaling may shed further light on the discriminability of figure-ground separation. The important pioneering studies of Julesz (1978) and his colleagues on texture statistics have thus far been restricted to conclusions which can be drawn from (passive) correlational estimates.

41. The principle of scale equivalence and the curvature of activity-scale correlations: Fechner's paradox, equidistance tendency, and depth without disparity. My description of how a FIRE can be spread and blocked sheds light on several types of data from a unified perspective. Suppose that, as in Section 36, an ambiguous monocular view of an object excites all structural scales due to selfmatching of the monocular data at each scale's binocular cells. Suppose that a binocular view of an object can selectively excite some structural scales more intensely than others due to the relationship between matching and activity amplification (Section 22). These assumptions are compatible with data concerning the simultaneous activation of several spatial scales at each position in the visual field during binocular viewing (Graham, Robson & Nachmias 1978; Robson & Graham 1981), with data on binocular brightness summation (Blake, Sloane & Fox 1981; Cogan, Silverman & Sekuler 1982), and with data concerning the simultaneous visibility of rivalrous patterns and a depth percept (Kaufman 1974; Kulikowski 1978). The suggestion that a depth percept can be generated by a selective amplification of activity in some scales above others also allows us to understand: (1) why a monocular view does not lose its filling-in capability or other resonant properties (since it can excite some structural scales via self-matches); (2) why a monocular view need not have greater visual sensitivity than a binocular view, despite the possibility of activating several scales due to self-matches (since a binocular view may excite its scales more selectively and with greater intensity due to binocular brightness summation); (3) why a monocular view may look brighter than a binocular view (Fechner's paradox) (since although the matched scales during a binocular view are amplified, so that activity lost by binocular mismatch in some scales is partially gained by binocular summation in other scales, the monocular view may excite more scales by self-matches); and (4) yet why a monocular view may have a more ambiguous depth than a binocular view (since a given scene may fail to selectively

THE BEHAVIORAL AND BRAIN SCIENCES (1983) 4 652

amplify some scales more than others due to its lack of spatial gradients [Gibson, 1950]).

The selective amplification that enhances a depth percept is sometimes due to the selectivity of disparity matches, but it need not be. The experiment of Kaufman, Bacon, and Barroso (1973) shows that depth can be altered, even when no absolute disparities exist, by varying the relative brightnesses of monocular pattern features. The present framework interprets this result as an external manipulation of the energies that cause selective amplification of certain scales above others, and as one that does so in such a way that the preferred scales are altered as the experimental inputs are varied.

The same ideas indicate how a combination of monocular motion cues and/or motion-dependent input energy changes can enhance a depth percept. Motions that selectively enhance delayed self-matches in certain scales above others can contribute to a depth percept. All of these remarks need quantitative implementation via a major program of computer simulations. The simulations that have already been completed do, however, support the mathematical, numerical, and qualitative results on which the theory is founded (Cohen & Grossberg 1983a). Although this program is not yet complete, the qualitative concepts indicate how to proceed and how various data may be explained in a unified fashion that are not discussed in a unified way by competing theories.

The idea that depth can be controlled by the energy balance across several active scales overcomes a problem in Sperling-Dev models. Due to the competition between depth planes in these models, only one depth plane at a time can be active in each spatial location. However, there can exist only finitely many depth planes, both on general grounds due to the finite dimension of neural networks, and on specific grounds due to inferences from spatial frequency data wherein only a few scales are needed to interpret the data (Graham 1981; Wilson & Bergen 1979). Why, then, do we not perceive just three or four different depths, one depth corresponding to activity in each depth plane? Why does the depth not seem to jump discretely from scale to scale as an object approaches us? Depth seems to change continuously as an object approaches us despite the existence of only a few structural scales. The idea that the energy balance across functional scales changes continuously as the object approaches, and thereby continuously alters the depth percept, provides an intuitively appealing answer. This idea also mechanistically explicates the popular thesis that the workings of spatial scales may be analogous to the workings of color vision, wherein the pattern of activity across a few cone receptor types forms the substrate for color percepts.

The present framework suggests an explanation of Gogel's equidistance tendency (Section 4). Suppose that a monocularly viewed object of ambiguous depth is viewed which excites most, or all, of its structural scales through self-matches. Let a nearby binocularly viewed object selectively amplify the scales with which it forms the best pooled binocular edges. Let a FIRE spread with the greatest vigor through these amplified scales. When the FIRE reaches the monocular self-matches within its scale, it can amplify the activity of these self-matches, much as occurs during binocular brightness summation. This shift in the energy balance across the scales which represent the monocularly viewed object impart it with depthfulness. This conclusion follows – and this is the crucial point – even though no new disparity information is produced within the self-matches by the FIRE. Only an energy shift occurs. Thus, although disparities may be sufficient to produce a depth percept, they may not be necessary to produce one.

I suggest instead that suitable correlations between activity and scaling across the network loci that represent different spatial positions produce a depth percept. Depth is perceived whenever the resonant activity distribution is "curved" among several structural scales as representational space is traversed, no matter how – monocularly or binocularly – the activity distribution achieves its curvature. This conclusion may be restated as a deceptively simple proposition: An object in the outside world is perceived to be curved if it induces a curvature in the abstract representational space of activity-scale correlations.

Such a conclusion seems to smack of naive realism, but it is saved from the perils of naive realism by the highly nonlinear and nonlocal nature of the shunting network representation of input patterns. The conclusion does, however, provide a scientific rationale for the temptations of naive realism, and points the way to a form of neorealism if one entertains the quantum-mechanical proposition that the curvature of an object in the outside world is also due to curved activity-scaling correlations in an abstract representational space. Such considerations lead beyond the scope of this article.

The view that all external operations that cause equivalent activity-scaling correlations generate equivalent depth percepts liberates our thinking from the current addiction to disparity computations and suggests how monocular gradients, monocular motion cues, and learned cognitive feedback signals can all contribute to a depth percept. Because of the importance of this conception to my theory, I give it a name: the *principle of scale equivalence*.

42. Reflectance rivalry and spatial frequency detection. The same ideas suggest an explanation of the Wallach and Adams (1954) data on rivalry between two central figures of different lightness (Section 13). Suppose that each monocular pattern generates a different functional scale when it is viewed monocularly (Section 38). Suppose, moreover, that the monocular input intensities are chosen so that the functional scales are spatially out of phase with each other. Then when a different input pattern is presented to each eye, the feedback exchange between monocular and binocular cells, being out of phase, can become rivalrous.

This explanation leads to a fascinating experimental possibility: Given an input figure of fixed size, test a series of lightness differences to the two eyes. Can one find ranges of lightness where the functional scales are rivalrous followed by ranges of lightness in which the functional scales can match? If this is possible, then it is probably due to the fact that only certain peaks in the two scales match binocularly. The extra peaks self-match. Should this happen, it may be possible to detect small spatial periodicities in lightness such that binocular matches are brighter than self-matches. I am not certain that these differences will be visible, because the fillingin process from the locations of amplified binocular matches across the regions of monocular self-matches may totally obscure the lightness differences of the two types of matches. Such a filling-in process may be interpreted as a type of brightness summation.

Another summation phenomenon which may reflect the activation of a functional scale is the decrease in threshold contrast needed to detect an extended grating pattern as the number of cycles in the pattern is increased. Robson and Graham (1981) explain this phenomenon quantitatively "by assuming that an extended grating pattern will be detected if any of the independently perturbed detectors on whose receptive field the stimulus falls signals its presence" (p. 409). What is perplexing about this phenomenon is that "some kind of summation process takes place over at least something approaching 64 cycles of our patterns . . . it is stretching credulity rather far to suppose that the visual system contains detectors with receptive fields having as many as 64 pairs of excitatory and inhibitory regions" (p. 413). This phenomenon seems less paradoxical if we suppose that a single suprathreshold peak within a structural scale can drive contiguous subthreshold peaks within that scale to suprathreshold values via a disinhibitory action. Suppose, moreover, that increasing the number of cycles increases the expected number of suprathreshold peaks that will occur at a fixed contrast. Then a summation effect across 64 structural wavelengths is not paradoxical if it is viewed as a filling-in reaction from suprathreshold peaks to subthreshold peaks, much like the filling-in reaction that may occur between binocular matches and selfmatches in the Wallach and Adams (1954) paradigm.

Due to the large number of phenomena which become intuitively more plausible using this type of filling-in idea, I believe that quantitative studies of how to vary input brightnesses to change the functional scales generated by complex visual stimuli deserve more experimental and theoretical study. One challenge is to find new ways to selectively increase or decrease the activity within one structural scale without inadvertently increasing or decreasing the activities within other active scales as well. In meeting this challenge, possible effects of brightness changes on perceived length are no less interesting than their effects on perceived depth. For example, suppose that an increase in input contrast decreases the functional scale within a prescribed structural scale. Even if the individual peaks in the several functional scales retain approximately the same height, a lightness difference may occur due to the increased density of peaks within a unit cellular region. This lightness difference will alter length scaling in the limited sense that it can alter the ease with which matching can occur between monocular signals at their binocular interface, as I have just argued. It remains quite obscure, however, how such a functional length change in a network's perceptual representation is related to the genesis of motor actions, or whether motor commands are synthesized from more global properties of the regions in which activity is concentrated across all scales. To the extent that motor consequences help to shape the synthesis of perceptual invariants, no more than a qualitative appreciation of how functional length changes can influence effects like Emmert's law may be possible until quantitative sensory-motor models are defined and simulated.

43. Resonance in a feedback dipole field: Binocular development and figure-ground completion. My discussions of how a FIRE spreads (Section 39) and of figure-ground completion (Section 40) tacitly used properties that require another design principle to be realized. This design suggests how visual networks are organized into *dipole fields* consisting of subfields of on-cells and subfields of off-cells with the on-cells joined together and the off-cells joined together by competitive interactions. Because this concept has been extensively discussed elsewhere (Grossberg 1980b; 1982c; 1982d), I will only sketch the properties which I need here.

I will start with a disclaimer to emphasize that I have a very specific concept in mind. My dipoles are not the classical dipoles which Julesz (1971b) used to build an analog model of stereopsis. My dipoles are on-cell off-cell pairs such that a sudden offset of a previously sustained input to the on-cell can elicit a transient antagonistic rebound, or off-reaction, in the activity of the off-cell. Similarly, a sudden and equal arousal increment to both the on-cell and the off-cell can elicit a transient antagonistic rebound in off-cell activity if the arousal increment occurs while the on-cell is active (Figure 20). Thus my notion of dipole describes how STM can be rapidly reset, either by temporal fluctuations in specific visual cues or by unexpected events, not necessarily visual at all, which are capable of triggering an arousal increment at visually responsive cells. In my theory, such an unexpected event



Figure 20. An antagonistic rebound, or off-reaction, in a gated dipole can be caused either by rapid offset of a phasic input or rapid onset of a nonspecific arousal input. As in Figure 21, function J(t) represents a phasic input, function I(t) represents a nonspecific arousal input, function  $x_5(t)$  represents the potential, or activity, of the on-channel's final stage, and function  $x_6(t)$  represents the potential, or activity, of the off-channel's final stage. (From Grossberg 1982c.)



Figure 21. In the simplest example of a gated dipole, phasic input J and arousal input I add in the on-channel to activate the potential  $x_1$ . The arousal input alone activates  $x_2$ . Signals  $S_1 = f(x_1)$  and  $S_2 = f(x_2)$  such that  $S_1 > S_2$  are thereby generated. In the square synapses, transmitters  $x_1$  and  $x_2$  slowly accumulate to a target level. Transmitter is also released at a rate proportional to  $S_1 z_1$  in the on-channel and  $S_2 z_2$  in the off-channel. This is the transmitter gating step. These signals perturb the potentials  $x_3$ and  $x_4$ , which thereupon compete to elicit the net on-reaction  $x_5$ and off-reaction  $x_6$ . See Grossberg (1980b; 1982d) for a mathematical analysis of gated dipole properties. (From Grossberg 1982c.)

is hypothesized to elicit the mismatch negativity component of the N200 evoked potential, and such an antagonistic rebound, or STM reset, is hypothesized to elicit the P300 evoked potential. These reactions to specific and nonspecific inputs are suggested to be mediated by slowly varying transmitter substances – notably catecholamines like norepinephrine – which multiplicatively gate, and thereby habituate to, input signals on their way to the oncells and the off-cells. The outputs of these cells thereupon compete before eliciting net on-reactions and offreactions, respectively, from the dipole (Figure 21).

In a dipole field, the on-cells are hypothesized to interact via a shunting on-center off-surround network. The off-cells are also hypothesized to interact via a shunting on-center off-surround network. These shunting networks normalize and tune the STM activity within the onsubfield and the off-subfield of the total dipole field network. The dipole interactions between on-cells and off-cells enable an on-cell onset to cause a complementary off-cell suppression, and an on-cell offset to cause a complementary off-cell enhancement. This duality of reactions makes sense of structural neural arrangements such as on-center off-surround networks juxtaposed against off-center on-surround networks and uses this unified processing framework to qualitatively explain visual phenomena such as positive and negative aftereffects, the McCollough effect, spatial frequency adaptation, monocular rivalry, and Gestalt switching between ambiguous figures (Grossberg 1980b).

The new features that justify mentioning dipole fields here are that the on-fields and off-fields can interact to generate functional scales, and that the signals which

regulate the balance of activity between on-cells and offcells can habituate as the transmitter substances that gate these signals are progressively depleted. These facts will now be used to clarify how figure-ground completion and binocular rivalry might occur. I wish to emphasize, however, that dipole fields were not invented to explain such visual effects. Rather, they were invented to explain how internal representations which self-organize (e.g., develop, learn) as a result of experience can be stabilized against the erosive effects of later environmental fluctuations. My adaptive resonance theory suggests how learning can occur in response to resonant activity patterns, yet is prevented from occurring when rapid STM reset and memory search routines are triggered by unexpected events. In the present instance, if LTM traces are placed in the feedforward and feedback pathways that subserve binocular resonances, then the theory suggests that binocular development will occur only in response to resonant data patterns, notably to objects to which attention is paid (Grossberg 1976b; 1978e; 1980b; Singer 1982). Because the mechanistic substrates needed for the stable self-organization of perceptual and cognitive codes are not peculiar to visual data, one can immediately understand why so many visual effects have analogs in other modalities.

An instructive instance of figure-ground completion is Beck's phantom letter E (Section 6). To fully explain this percept, one needs a good model of competition between orientation sensitive dipole fields; in particular, a good physiological model of cortical hypercolumn organization (Hubel & Wiesel 1977). Some observations can be made about the relevance of dipole field organization in the absence of a complete model.

Suppose that the regularly spaced vertical dark lines of the "ground" are sufficiently dense to create a statistically smoothed pattern when they are preprocessed by the nonlinear cross-correlators of some structural scales (Glass & Switkes 1976). When such a smoothed pattern undergoes noise suppression within a structural scale, it generates statistical edges at the boundary of the "ground" region due to the sudden change in input statistics at this boundary. These edges of the (black) offfield generate complementary edges of the (white) onfield due to dipole inhibition within this structural scale. These complementary edges can use the ambiguous (preprocessed) white as an energy source to generate a FIRE that fills in the interior of the "ground." This FIRE *defines* the ground as a coherent entity. The "ground" does not penetrate the "figure" because FIBs are generated by the competition which exists between orientation detectors of sufficiently different orientation.

A "figure" percept can arise in this situation as the complement of the coherently filled-in "ground," which creates a large shift in activity-scale correlations at the representational loci corresponding to the "ground" region. In order for the "figure" to achieve a unitary existence other than as the complement of the "ground," a mechanism must operate on a broader structural scale than that of the variously oriented lines that fill the figure. For example, suppose that, due to the greater spatial extent of vertical ground lines than nonvertical figure lines, the smoothed vertical edges can almost completely inhibit *all* smoothed nonvertical edges near the figureground boundary. Then the "figure" can be completed as a disinhibitory filling-in reaction among *all* the smoothed nonvertical orientations of this structural scale. Thus, according to this view, "figure" and "ground" fill-in due to disinhibitory reactions among different subsets of cells. A lightness difference may be produced between such a "figure" and a "ground" (Dodwell 1975).

A similar argument sharpens the description of how figure-ground completion occurs during viewing of the Julesz 5% stereogram (Section 40). In this situation, black dots that can be fused by one structural scale may nonetheless form FIBs in other structural scales. A FIRE is triggered in the structural scales with fused black dots by the disinhibitory edges which flank the dots in the scale's white off-field. This FIRE propagates until it reaches FIBs that are generated by the nonfused dots corresponding to an input region of different disparity. The same thing happens in all structural scales which can fuse some of the dots. The figure-ground percept is a statistical property of all the FIREs that occur across scales.

44. Binocular rivalry. Binocular rivalry can occur in a feedback dipole field. The dynamics of a dipole field also explain why sustained monocular viewing of a scene does not routinely cause a perceived waxing and waning of the scene at the frequency of binocular rivalry, but may nonetheless cause monocular rivalry in response to suitably constructed pictures at a rate that depends on the juxtaposition of features in the picture (Grossberg 1980b, Section 12). I will here focus on how the slowly habituating transmitter gates in the dipole field could cause binocular rivalry without necessarily causing monocular waxing and waning.

Let a pair of smoothed monocular edges mismatch at the binocular matching cells. Also suppose that one edge momentarily enjoys a sufficient energetic advantage over the other to be amplified by contrast enhancement as the other is completely suppressed. This suppression can be mediated by the competition between the off-cells that correspond to the rivalrous edges. In particular, the oncells of the enhanced edge inhibit their off-cells via dipole competition. Due to the tonic activation of off-cells, the off-cells of the other edge are disinhibited via the shunting competition that normalizes and tunes the off-field. The on-cells of these disinhibited off-cells are thereupon inhibited via dipole competition.

As this is going on, the winning edge at the binocular matching cells elicits the feedback signals that ignite whatever FIRES can be supported by the monocular data. This resonant activity gradually depletes the transmitters which gate the resonating pathways. As the habituation of transmitter progresses, the net sizes of the gated signals decrease.

The inhibited monocular representation does not suffer this disadvantage because its signals, having been suppressed, do not habituate the transmitter gates in their pathways. Finally, a time may be reached when the winning monocular representation loses its competitive advantage due to progressive habituation of its transmitter gates. As soon as the binocular competition favors the other monocular representation, contrast enhancement bootstraps it into a winning position and a rivalrous cycle is initiated.

A monocularly viewed scene would not inevitably wax and wane, for the following reason. Other things being

equal, its transmitter gates habituate to a steady level such that the habituated gated signals are an increasing function of their input sizes (Grossberg 1968; 1981; 1982e). Rivalry occurs only when competitive feedback signaling, by rapidly suppressing some populations but not others, sets the stage for the competitive balance to slowly reverse as the active pathways that sustain the suppression habituate faster than the inactive pathways. The same mechanism can cause a percept of monocular rivalry to occur when the monocular input pattern contains a suitable spatial juxtaposition of mutually competitive features (Rauschecker, Campbell & Atkinson 1973).

45. Concluding remarks about filling-in and quantization. The quantized dynamic geometry of FIRE provides a mechanistic framework in which the experimental interdependence of many visual properties may be discussed in a unified fashion. Of course, a great deal of theoretical work remains to be done (even assuming all the concepts are correct), not only in working out the physiological designs in which these dynamic transactions take place but also in subjecting the numerical and mathematical properties of these designs to a confrontation with quantitative data. Also, the discussion of disinhibitory fillingin needs to be complemented by a discussion of how hierarchical feedback interactions between the feedforward adaptive filters (features) and feedback adaptive templates (expectancies) that define and stabilize a developing code can generate pattern completion effects, which are another form of filling-in (Dodwell 1975; Grossberg 1978e, Sections 21-22; 1980b, Section 17; Lanze, Weisstein & Harris 1982). Despite the incompleteness of this program, the very existence of such a quantization scheme suggests an answer to some fundamental questions.

Many scientists have, for example, realized that since the brain is a universal measurement device acting on the quantum level, its dynamics should in some sense be quantized. This article suggests a new sense in which this is true by explicating some quantized properties of binocular resonances. One can press this question further by asking why binocular resonances are nonlinear phenomena that do not take the form of classical linear quantum theory. I have elsewhere argued that this is because of the crucial role which resonance plays in stabilizing the brain's self-organization (Grossberg 1976; 1978e; 1980b). The traditional quantum theory is not derived from principles of self-organization, despite the fact that the evolution of physical matter is as much a fundamental problem of self-organization on the quantum level as are the problems of brain development, perception, and learning. It will be interesting to see, as the years go by, whether traditional quantum theory looks more like an adaptive resonance theory as it too incorporates self-organizing principles into its computational structure.

## APPENDIX

The following system of equations defines a binocular interaction capable of supporting a filling-in resonant exchange (Cohen & Grossberg 1983a). Monocular representations

$$\frac{d}{dt} x_{iL} = -Ax_{iL} + (B - x_{iL}) \sum_{k=1}^{n} I_{kL} [J_{kL} + z_k] + C_{ki}$$
$$-(x_{iL} + D) \sum_{k=1}^{n} I_{kL} [J_{kL} + z_k] + E_{ki} \quad (A1)$$

$$\frac{d}{dt} x_{iR} = -Ax_{iR} + (B - x_{iR}) \sum_{k=1}^{n} I_{kR} [J_{kR} + z_k] + C_{ki}$$
$$-(x_{iR} + D) \sum_{k=1}^{n} I_{kR} [J_{kR} + z_k] + E_{ki} \quad (A2)$$

where  $[\xi]^+ = \max(\xi, 0)$ .

Binocular matching

$$y_{i} = \frac{\sum_{k=1}^{n} \tilde{F}_{ki}[f(x_{kL}) + f(x_{kR})]}{\tilde{A} + \sum_{k=1}^{n} \tilde{C}_{ki}[f(x_{RL}) + f(x_{kR})]}$$
(A3)

$$\tilde{\mathbf{F}}_{\mathbf{k}i} = \tilde{\mathbf{B}}\tilde{\mathbf{C}}_{\mathbf{k}i} - \tilde{\mathbf{D}}\tilde{\mathbf{E}}_{\mathbf{k}i} \tag{A4}$$

and

$$\tilde{\mathbf{G}}_{\mathbf{k}i} = \tilde{\mathbf{C}}_{\mathbf{k}i} + \tilde{\mathbf{E}}_{\mathbf{k}i} \tag{A5}$$

Binocular-to-monocular feedback

$$z_{i} = \frac{\sum_{k=1}^{n} F_{ki}^{*} g(y_{k})}{A^{*} + \sum_{k=1}^{n} G_{ki}^{*} g(y_{k})}$$
(A6)

$$F_{ki}^* = B^*C_{ki}^* - D^*E_{ki}^*$$
 (A7)

and

$$G_{ki}^* = C_{ki}^* + E_{ki}^*$$
 (A8)

Equation (A1) describes the response of the activities  $x_{iL}$ , i = 1, 2, ..., n, in the left monocular representation. Each  $x_{iL}$  obeys a shunting equation in which both the excitatory interaction coefficients  $C_{ki}$  and the inhibitory interaction coefficients  $E_{ki}$  are Gaussian functions of the distance between  $v_k$  and  $v_i$ . Two types of simulations have been studied:

Additive inputs.-All  $I_{kL}$  are chosen equal. The terms  $J_{kL}$  register the input pattern and summate with the binocular-tomonocular feedback functions  $z_k$ .

monocular feedback functions  $z_k$ . Shunting inputs.-All  $J_{kL}$  are chosen equal. The terms  $I_{kL}$  register the input pattern. The binocular-to-monocular feedback functions  $z_k$  modulate the system's sensitivity to the inputs  $I_{kL}$  in the form of gain control signals.

Equation (A2) for the activities  $x_{iR}$ ,  $i = 1, 2, \ldots, n$ , in the right monocular representation has a similar interpretation. Note that the same binocular-to-monocular feedback functions  $z_k$  are fed back to the left and right monocular representations.

The binocular matching stage (A3) obeys an algebraic equation rather than a differential equation due to the simplifying assumption that the differential equation for the matching activities  $y_i$  reacts quickly to the monocular signals  $f(x_{kL})$  and  $f(x_{kR})$ . Consequently,  $y_i$  is always in an approximate equilibrium with respect to its input signals. This equilibrium equation says that the monocular inputs  $f(x_{kL})$  and  $f(x_{kR})$  are added before being matched by the shunting interaction. The signal functions f(w) are chosen to be sigmoid functions of activity w. The excitatory interaction coefficients  $\tilde{C}_{ki}$  and inhibitory interaction coefficients  $\tilde{E}_{ki}$  are chosen to be Gaussian functions of distance. The spatial decay rates of  $C_{ki}$ ,  $\tilde{C}_{ki}$ , and  $C^*_{ki}$  are chosen equal. The spatial decay rates of  $E_{ki}$ ,  $\tilde{E}_{ki}$ , and  $E^*_{ki}$  are chosen equal. The oncenter is chosen narrower than the off-surround.

After the monocular signal patterns  $(f(x_{1L}), f(x_{2L}), \ldots, f(x_{nL}))$ and  $(f(x_{1R}), f(x_{2R}), \ldots, f(x_{nR}))$  are matched at the binocular matching stage, the binocular activities  $y_k$  are rectified by the output signal function  $g(y_k)$ , which is typically chosen to be a sigmoid function of  $y_k$ . Then these rectified output signals are distributed back to the monocular representations via competitive signals (A6) with the same spatial bandwidths as are used throughout the computation.

Numerical studies have been undertaken with the following types of results (Cohen & Grossberg 1983a). An "edgeless blob," or Gaussianly smoothed rectangular input, does not supraliminally excite the network at any input intensity. By contrast, when a rectangle is added to the blob input, the network generates a FIRE that globally fills-in the "figure" defined by the rectangle and uses the rectangle's edges to generate a globally structured "ground" (Figure 22). Despite



Figure-ground filling-in due to a rectangle on an Figure 22. "edgeless blob": By itself, the blob elicits no suprathreshold reaction in the binocular matching field at any input intensity. By itself, in a network without feedback from the matching field, the rectangle elicits only a pair of boundary edges at any input intensity. Given a fixed ratio of rectangle to blob intensity in the full network, as the background input intensity is parametrically increased, the network first elicits subthreshold reactions to the edges of the rectangle. Once the quenching threshold is exceeded, a full blown global resonance is triggered. Then the rectangle fills-in an intensity estimate between its edges (the "figure"), and structures the blob so that it fills-in an intensity estimate across the entire blob (the "ground"). The two intensity estimates reflect the ratio of rectangle-to-blob input intensities. (From Cohen & Grossberg 1982.)

the fact that the network is totally insensitive to the blob's intensity in the absence of the rectangle, the rectangle's presence in the blob sensitizes the network to the ratio of rectangleplus-blob to blob intensities, and globally fills-in these figure and ground lightness estimates. Parametric input series have been done with rectangles on rectangles, rectangles on blobs, triangles on rectangles, and so forth to study how the network estimates and globally fills-in lightness estimates that are sensitive to the figure-to-ground intensity ratio.

Monocular patterns that are mismatched relative to a prescribed structural scale do not activate a FIRE at input intensities that are suprathreshold for matched monocular patterns. Thus, different structural scales selectively resonate to the patterns that they can match. Different structural scales also generate different functional scales, other things being equal.

Matched monocular patterns such as those described above have been shown to elicit only subliminal feedforward edge reactions until their intensities exceed the network's quenching threshold, whereupon a full-blown global resonance is initiated which reflects disparity, length, and lightness data in the manner previously described.

## ACKNOWLEDGMENTS

This work was supported in part by the Air Force Office of Scientific Research (AFOSR 82-0148), the National Science Foundation (NSF IST-80-00257), and the Office of Naval Research (N00014-83-K0337).

## **Open Peer Commentary**

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

## "Filling-in" between edges

### Lawrence E. Arend, Jr.

Eye Research Institute of Retina Foundation, Boston, Mass. 02114

I would like to mention several "filling-in" phenomena, some of which seem to me to be particularly problematic for the resonance explanations of Grossberg's model, and some problematic for other filling-in models as well. Since I have attempted no simulations with this model I present them as potential difficulties rather than contradictions.

The first issue concerns the smoothing of filled-in areas. In Grossberg's target article little attention is devoted to the fact that no periodicity of brightness is perceived in homogeneous image regions. Such a fundamental observation should not be minimized, even in the developmental stages of a model with other, more interesting properties. If the smoothness is to be attributed to summation among the responses at different functional scales, there must evidently be many (and even more structural scales since pattern/structure interaction limits the number of functional scales). Summation of outputs of such a dense population of structural scales would tend to diminish the perceptual impact of the quantum properties of the model.

The problem of limiting the spread of effects from an edge is more complicated than is evident from the simple Craik-O'Brien-Cornsweet edge alone. The observation that the scallop luminance distribution and true luminance edge both look like the edge rather than the scallop takes on additional importance in light of the global nature of the filling-in. In fact, the term "filling-in" is itself something of a misnomer when applied to the Craik-O'Brien-Cornsweet illusion. Extensions of the original observations show that something resembling indefinite integration over differential information from multiple edges is involved (Arend 1973; Arend, Buehler & Lockhead 1971). If several Cornsweet scallops of the same handedness are present (Figure 1[b]) the effects of one edge are not terminated by a second edge, but rather extend undiminished to the areas beyond, summing with the effects of succeeding edges. Analogous effects in arrays of colored papers under nonuniform illumination (Land & McCann 1971) extend over very broad regions. In these cases filling-in due to resonant-exchange would have to occur over a number of functional scales and over long distances, without creating visible periodic nonuniformities within subregions of the pattern.

A further problem is a challenge not only to the resonantexchange model, but to all current filling-in models: The domain affected by an edge is limited. The resonant-exchange model uses binocular mechanisms to limit interactions to areas in the same depth planes. There are, however, similar limits to the domain of interactions even within the same depth plane. When the disk containing the scallops of Figure 1(b) is placed beside a second disk on a black background, the scallops do not alter the appearance of grays on the second disk, even though it lies in the same depth plane physically and perceptually. The outer ring of the scalloped disk appears darker than a physically separated patch of the same luminance. If the integration extended over the intervening disk borders, the two equal-luminance regions should have the same brightness.

A similar domain-of-integration effect occurs in Land's color invariance demonstration (Land 1977). Two identical colored paper arrays were illuminated by separate light sources, the source for each consisting of three narrowband lights. Land demonstrated that within each array the papers' hues depend on the relationships among all the chromaticities in the array rather than local chromaticities. By separately adjusting the two illuminants' three components, Land matched the chromaticity of one array's "green" paper to that of a "blue" paper in the second. In spite of physical identity of the light reaching the eye from the



Figure 1 (Arend). Spatial luminance distributions (left) and corresponding brightness distributions (right) along the radius of a spinning disk. Each luminance distribution begins at 10 mL at the extreme left. (From Arend, Buehler & Lockhead 1971.)

two papers, they retained their white-illuminant hues within the tolerances of color names, i.e., there was a normalization of color within the array. In Land's model the normalization occurs in the process of integrating edge information along lines through the array. As in the case of the two disks just described, the integral must not extend over the edges intervening between the two arrays if the "green" and "blue" patches are to look different. If the integrals over edges along a line between the "green" and "blue" patches include the differences at the edges of the two arrays, the two patches should have the same hue rather than the reported "green" and "blue."

I know of no model (including Land's and my own) capable in its present form of explaining this limitation to subdomains of integration. In the model presented by Grossberg the spread process is stopped by well-defined processes. The model should therefore be capable of generating psychophysically testable predictions of stimulus conditions defining boundaries of perceptual domains of integration.

## Isomorphism is where you find it

### Bruce Bridgeman

Psychology Board of Studies, University of California, Santa Cruz, Calif. 95064

One of the values of mathematical modeling is that it makes explicit what we know and what we don't know about sensory processes. Usually the latter is more dramatic than the former. Grossberg's work of the last decade or so has made contributions to both sides.

Several of the mathematical properties of the shunting inhibition networks which Grossberg analyzes are shared with related networks invoking lateral interaction among elements arranged in parallel. When we translate Grossberg's nomenclature into physiological terms (for instance, quantum equals multistable; competitive equals lateral inhibitory), we see that some of the psychologically useful properties of these networks are not new. The automatic tuning of sensitivity (Section 21A), for instance, is a property of the subtractive lateral inhibition networks based on the Hartline-Ratliff model from *Limulus*. Grossberg's basic model in his Equation 1 is almost exactly the differential of this model. The adaptation characteristic has also been used in a similar subtractive network adapted to mammalian conditions (Bridgeman 1971).

As Grossberg notes, noise suppression is also characteristic of additive networks under some conditions (Ratliff 1965), and has been quantitatively analyzed (Bridgeman 1978). Grossberg is right that contrast of pattern responses in linear additive models does not change as a function of suprathreshold background luminance, but contrast does change with luminance when physiologically realistic boundary conditions are imposed on activity levels in the additive model.

Grossberg notes that inhibitory networks will carry stimulusgenerated disturbances far beyond the anatomical spread of individual inhibitory connections. This important characteristic is also shared by additive models (Bridgeman 1971), and forms one of the bases for rejecting a strict detector scheme in higherlevel sensory coding; if activity from a restricted region is recoded by inhibition over a wide region of a network, most of that information is present in areas not directly affected by the original input. In addition to the spatial spread, there is also a temporal spread of information which has been simulated parametrically (Bridgeman 1971; 1978). Temporal spread of information does not support comparisons with short term memory because of the very short persistence of activity held in these networks and because the bottom-up nature of the storage is more comparable to iconic storage than to short term memory.

The relationship of lateral inhibitory models with spatial

frequency analysis has also been noted previously (Bridgeman 1977); if each neuron in the network has the same lateral inhibitory coefficients, the network will be tuned to be maximally sensitive to a single spatial frequency.

Other issues are broader than the generality of the mathematical properties of inhibitory networks and deal with the appropriateness of these and similar simulation programs. First, the use of shunting inhibition rather than subtractive inhibition is a physiological assumption which can be tested empirically. The original model for shunting inhibition is an axo-axonal synapse in the spinal cord; conventional postsynaptic inhibition, which is subtractive, is more common in the cortex. A possible source of shunting inhibition in cortex is the recent discovery of the widespread distribution of neuropeptide transmitters. These transmitters often act on target neurons not by changing postsynaptic potentials directly, but by changing postsynaptic sensitivity to other neurotransmitters. This biochemical scheme complicates mathematical simulations because two separate systems with different dynamics are involved.

One of the paradoxes of model building is that even if a model is successful in simulating a number of psychological characteristics in realistic ways, there is no guarantee that other models might not do the same. One way to restrict the number of possible models is to use our knowledge about the physiological substrate of the networks to show that each parameter of the network has a physiological concomitant. This homeomorphic approach has been used in simulating additive networks (Ratliff 1965; Bridgeman 1971), and needs to be done more thoroughly for the Grossberg alternative.

The most problematic aspect of all simple models applied to complex phenomena, including my own, is the comparison between network activity and organismic behavior or experience. This problem is illustrated with Grossberg's efforts to solve the "filling in" problem by literally having modulations of network activity fill in areas to which the model's elements are insensitive. This effort assumes the need for a rather literal isomorphism between the activity and the network and the resulting perceptions. But filling in and related issues may be pseudo-problems because a literal isomorphism is not required. The only essential feature is a second-order isomorphism (Shepard & Chipman 1970) where there is a one-to-one relationship between network activity and experience. As applied to the Grossberg simulations presented here, second-order isomorphism means that it is not necessary to fill a homogeneous area. with network oscillations for it to "appear" filled in. The asymmetry of the Laplacian edge might do the job equally well with fewer added assumptions. Many of the problems of binocular summation and rivalry might be handled in a similar way. As long as no homunculus is looking at a realistic picture painted on the inhibitory network, information can be represented in nonliteralistic ways.

Finally, there is a question whether large-scale effects such as those simulated here can be appropriately discussed at the level of interactions between a relatively small number of neurons in a single layer. I call this objection "Uttalism" in honor of Uttal's (1973) strenuous objections to these sorts of extrapolations. At the beginning of Part II, for instance, Grossberg states the problem as that of relating the psychological data to the membrane equations of neurophysiology. This misses the point, because one of the great powers of brain organization is that algorithms at a higher level can become independent of the membrane equations which must support them at a lower level. The now-commonplace analogy of the computer program independent of the computer's hardware applies here. Lower-level models, of course, must be consistent with neurophysiology, but by the time we get to correlates of experience involving millions or billions of neurons interacting over hundreds of milliseconds, the neurophysiological scale of analysis may be no more useful than describing the pattern on an oriental rug by specifying the color of one knot after another; when you are

finished, you still don't know what the rug looks like. We know, for instance, that many of the properties of short-term memory, such as a variety of cognitive "chunking" effects, are inconsistent with storage in a single-layer network. The network described in Section 28 is too microscopic a level to support plausible explanations of short-term memory. Again, accounting for the McCullough effect (Section 43) in a simple topographic model is difficult because the effect can last for weeks, to be recalled only by the specific stimuli which first elicited it. New information can come in without erasing the old.

Grossberg has made a major contribution to mathematical modeling of sensory systems, but should not be expected to answer all questions in a single stroke. David Marr's (1982, Chapter 1) point that algorithms are just as important as neurophysiology on the one hand and psychophysics on the other must eventually be integrated into models simulating cognitive processes.

## Functional and computational aspects of perception

Hans Buffart

Department of Experimental Psychology, University of Nijmegen, 6500 HE Nijmegen, The Netherlands

1. Methodology. Marr (1977) emphasizes the distinction between functional and computational theories. A functional theory is "an abstract formulation of what a system processes and why." A computational theory explains how the system is constructed and why it works on the basis of the given "hardware." Grossberg's paper represents a computational theory about the functional behaviour of the visual system. However, without a functional theory a computational theory eludes falsification.

1.1. Modeling cognition in terms of networks. Which theoretical principle determines which network is sensitive to cognitive factors? Brightness- (van den Brink & Keemink 1976) and size-perception (Rock 1977) may be influenced by interpretations, but binocular rivalry is hardly influenced (Levelt 1968).

Which theoretical principle justifies an interpretation of a cognitive phenomenon in terms of some network-property? "Gestalt switching" in ambiguous patterns may be interpreted as a two-stage process (Grossberg 1980b) and can thus be described by a dipole field. But even in the absence of perceptual ambiguity perception is a two-stage process (Leeuwenberg 1982; van Tuijl & Leeuwenberg 1979). Perceptual ambiguity is only an extreme case, and sometimes it is not even experienced (Buffart, Leeuwenberg & Restle 1981).

1.2. Alternative modelling of retinal interactions. In his target article Grossberg ascribes Weber's law to a feedforward network. This is, indeed, an explanation, but the assumption that Weber's law is due to a chemical process transforming lightenergy into electrical activity is equally admissible (Cornsweet 1970). The chemical model fits the classical data of König and Brodhun (1889); it predicts (Buffart 1978, 1981) the shift of the relative threshold, r, due to the overall luminance in spatial, f(van Nes 1968; van Nes & Bouman 1965) and temporal (de Lange 1957) frequency detection. It also explains why under some conditions this threshold's dependency on spatial frequency remarkably changes from exp[af] into exp[3af].

The latter functions arouse doubt whether the interaction coefficients between retinal cells can be described by Gaussian distributions. Much the best set of orthonormal functions describing the data in the frequency domain is the set of Laguerre functions. This means that the interaction between cells would be better described by a Lorentz distribution. In the first approximation this coincides with the Gaussian distribution, but it has the computationally inconvenient, yet theoretically important property that its spread is infinite.

1.3. Alternative interpretation of psychophysical data. Hochberg (1964) and Levelt (1968) proposed that binocular interaction is due to a permanent rivalry at the micro level even in the case of perceptually nonrivalrous images. In a crucial experiment Fox and McIntyre (1967) produced evidence for this hypothesis. Consequently, brightness summation does not occur, and although it is strongly advocated (Blake & Fox 1973; Curtis & Rule 1978; Engel 1967, 1969) a comparative test (de Weert & Levelt 1974) has shown that averaging or weighting models describe the data better.

The computational modelling of Grossberg is evidently so powerful that networks can be designed showing permanent rivalry and brightness weighting. But this again emphasizes the need for a functional model.

2. A Julesz-Levelt model for cyclopean perception. If a functional model is falsified, all related networks are falsified, but the reverse is not true. A network is more specific about electrophysiological factors. Phenomena in cyclopean perception, explained by Grossberg on the basis of networks, can, for instance, also be explained by a functional model (Buffart, 1981) unifying the ideas of Julesz (1971) and Levelt (1968).

Briefly, the model is as follows. The left (l) and right (r) retinae are only sensitive to luminance increase (+) and decrease (-). Every point,  $\vec{r}$ , in the cyclopean space, R2, has four states, j, representing the type (l+, l-, r+, r-) of signal processed. These exclude each other and there are restrictions (Figure 1) on the mutual transitions (the rivalry principle). Note that the seemingly evident on-off interaction during binocular rivalry is absent. A two-dimensional vector, called a disparity-detector,  $\vec{d}$ , determines the retinal place,  $\vec{r} + \vec{d}$ , from which the transition-inducing signal originates. After defining the state, x, of the whole system one defines mathematically the probability density, p(x,t), to find the system in state x at time t. The dynamic development of such a system is described by the so-called master-equation:

$$(d/dt) p(x,t) = \int_{y} [w(x,y,t)p(y,t) - w(y,x,t)p(x,t)]$$

Reformulating this into an equation for the dynamic development of the probability-density P(j, d, r, t) in the point  $\dot{r}$  a psychophysical meaning is assigned to the transition-probability per unit time, w. Here three independent interactions are supposed to take place in the cyclopean space: First, an autonomous spatial interaction, mostly interpreted as cooperation. Second, an autonomous disparity interaction (Fender & Julesz 1967). Third, the permanent rivalry interaction (Figure 1), for which wis identified with the retinal output representing the luminance



Figure 1 (Buffart). The transition-schema of the four cyclopean states: monocular, on-off transitions; binocular, on-on and off-off transitions. A state may only be reached due to a retinal signal of the same type.

change per unit time. The resulting equation looks like a competitive network, and may serve as a functional justification of the computational approach. It appears that several phenomena are due to functional aspects of the cyclopean system.

For instance, the model shows depth due to disparity, it explains displacement, depth due to brightness and in the presence of rivalry (Kaufman, Bacon & Barroso 1973), the role of bias in ambiguous stereograms (Julesz 1971b), and filling in (Gerrits & Vendrik 1970a, 1970b, 1972, 1974; Julesz 1971b). The system is multistable in the case of binocular rivalry, and transitions are mainly caused by stimulus changes in the nondominant eye (Levelt 1968). The first-approximation model for binocular brightness is a modified version of the centroid model (Schrödinger 1926). Data on binocular as well as monocular brightness perception are fitted. Vernier acuity, stereo acuity, and Panum's area may be identified with the three types of interactions. In the first order the autonomous disparity interaction is an Ornstein-Uhlenbeck process fitting the data of Fender and Julesz (1967), and the autonomous spatial interaction is a Wiener process from which the U-shaped curves for optimal apparent motion may be derived.

Note that phenomena caused by specific computational properties can never be explained by a functional theory.

## Universal coding and network structures for vision: Is Grossberg correct?

## Terry Caelli

Psychology Department, University of Alberta, Edmonton, Alberta, Canada, T6G 2E9

For almost twenty years Grossberg has focused on the difficult problem of developing a general representation for neural activity which can be quantitatively applied to a wide variety of sensory- and memory-specific phenomena. This representation is based on a few critical assumptions about neural activity, albeit in specific applications they become quite complex in their cascaded forms. These assumptions, I propose, seem to be:

(A1) Internal representations (including neurophysiological) of visual events are defined by the temporal responses and interactions of cells which form natural equivalence classes with respect to their function.

(A2) The activity of each cell is defined by its "direct" response to external stimulation and interaction between cells of other classes. These interactions are restricted to: excitatory, inhibitory, feedforward, feedbackward, and shunting processes.

(A3) "Perception" occurs by the temporal course of such activity and, in particular, visual "percepts" are determined when such activity reaches equilibrium conditions.

It is these assumptions which drive Grossberg to consider a large range of visual (and other) phenomena in terms of the generalized coding equation

$$\frac{dx_i}{dt} = \sum_{j=1}^n A_j \cdot x_i + \sum_{k=1}^m (B_k - x_i) \cdot I_i + \sum_{l=1}^r C_l \cdot I_l \quad (1)$$

where  $A_j$ ,  $B_k$ ,  $C_l$  represent excitatory, inhibitory (or quenching), and threshold vectors which determine the time varying response of cells  $x_i$ . It should be noted that such coefficients are adequate for the determination of inter- and intracell facilitation and inhibition since the summation occurs (n, m, r) over any specified class of units (including the cell itself if required).

The question is: Do these assumptions, and the associated network equation (1), satisfy Grossberg's claim to have generated a most general language to describe visual function? This can be answered in two ways. One, can we prove that all known models for specific visual function are examples of this formulation? Two, can we analytically derive such a representation from known neurophysiological and psychophysical results?

As Grossberg and others (Caelli 1982) have already pointed out, any filter model or general linear systems approach to spatial vision can be represented in discrete form by (1) with appropriately defined equivalence classes and weighting functions (A, B, C in (1)) which define filter profiles. Equally, all existent models for "cooperative" or "competitive" visual phenomena rely on (a) some form of interelement interaction, (b) summation, and (c) usually some threshold device. Again, these processes are precisely encoded in (1), and implicit in the assumptions of the Grossberg representation. The feedforward and feedbackward constraints on such systems are encoded in (1) via the time series characteristics of the A, B, C vectors where, for example we may further consider A(t) = F(B(t - k)), etc.

Concerning the second question, perhaps the most general (and generally accepted) model for a given cell's electrical activity is (Leake & Anninos 1976).

$$\phi_i = \eta_i + \sum_{j=1}^m C_{ij} g(\phi_j)$$
(2)

where  $\eta_i$  = the external input to cell *i* (afferent volley including the resting potential),  $\phi_i$  = the output of cell *i* (for example, average membrane potential),  $g(\phi_j)$  = some function of the input or output of cell *j* relevant as input to other cells, and  $C_{ij}$  = the connectivity between cells *i* and *j* reflecting complex dendritic processes ( $C_{ii}$  would represent recurrent collaterals). This representation (2), again, is compatable with (a) so long as the coefficients *A*, *B* are seen as weighting functions and summation (*n*, *m*) occurs over all elements.

So we may well conclude, from the answers to the two questions posed above, that the Grossberg formulation is general in both senses of the word. This is not the limitation of his theory. No, where Grossberg's work (including the present treatments of depth and stereopsis) is incomplete is in the generation and prediction of precise psychophysical data, besides the general existence of psychophysical functions. Grossberg makes the comparison between models in physics and visual perception and correctly concludes about the latter that we do not have a general theory or even language for visual function. However, we should not think that physics has. Physical laws and general representations exist to describe physical phenomena (for example, special and even general relativity). However, when it comes down to specific explanations with respect to the prediction of events as sophisticated as visual phenomena, physics is equally deficient. In fact, Grossberg's distinction between functional and spatial scales and, in particular, the relationships between computational networks (giving functional scales) and psychophysical scales (spatial scales) is important. But, again, detailed evaluation of precise psychophysical data is required before this general language can be validated and accepted.

Grossberg does not need to reject other models of precise visual function in order to prove that his representation is valid. For example, though the very activity of (1) is in general nonlinear, this does not imply that from the next level such nonlinear networks cannot be represented by a linear system. Analogously, the highly nonlinear activity of subatomic particles does not preclude a linear molecular chemistry. Equally, (1) does not necessarily imply that aspects of receptive field function or geometry cannot be modelled in linear systems (linear filter) terms. Just the opposite: since A, B, in (1) and  $C_{ij}$  in (2) represent weighting functions, such components have a filter representation, and vice versa.

## When "filling in" fails

## Stanley Coren

Department of Psychology, University of British Columbia, Vancouver, B.C., Canada V6T 1Y7

Grossberg has offered an interesting and ambitious model which he believes explains a number of lightness and form interactions. Of particular interest are the Craik-O'Brien and Cornsweet effects (Cornsweet 1970; O'Brien 1958), which show how a sharp local change in luminance can alter the apparent brightness of a fairly wide region of the field. Perhaps conceptually the simplest example of such an effect is shown in Figure 1(a), where we have two linear luminance gradients, equal in mean intensity and spatially adjacent to one another. The resultant percept is shown as Figure 1(b), which is a sensation equivalent to two uniform areas differing in brightness.

Actual measurements were taken on a pattern similar to this, using six experimentally sophisticated observers. Each matched the apparent brightness of the two regions using a variable reflectance rotor. The patterns themselves were produced on a rotor, and were  $0.5^{\circ}$  of visual angle, and each gradient varied from a minimum reflectance of 40% to a maximum of 45%; thus at the boundary the ratio of the reflectances was 1.13. The mean matched reflectance difference was 1.11, which is not statistically different from that expected on the basis of edge computation, and the two regions appeared quite uniform. Thus, in accordance with Grossberg's reasoning, we have demonstrated both the filling-in and the edge-computation processes.

With this in mind let us now simply extend this pattern, so that instead of having two linear gradients as shown in Figure 1(a) we now have five linear gradients, as shown in Figure 1(c). Since nothing about such an extension apparently affects any of the theoretical mechanisms suggested by Grossberg, we should now expect the perception of five uniform bands differing in brightness by the sequential product of the brightness transitions, which comes to a ratio of 1.6 between the leftmost band and the rightmost.

When the same observers now made brightness matches on such a pattern, they produced a reflectance ratio of 0.98, which is not significantly different from 1.0, representing no perceived brightness difference; this is significantly below the expected 1.6 predicted from the theory. However, perhaps even more important, the apparent uniformity of the individual bands is now gone, and the apparent brightness profile looks like that shown in Figure 1(d). The filling-in mechanism seems to have



Figure 1 (Coren). The luminance pattern shown as a produces the apparent brightness pattern shown as b, complete with filling in. The luminance pattern shown as c produces the apparent brightness pattern shown as d, with no filling in.

failed completely, and the creation of brightness differences from the local luminance transitions now also seems inoperative, although, from a conceptual standpoint, patterns 1(a) and 1(c) should not alter the operation of the theoretical mechanisms in any way.

What is going on in this situation? Perhaps the simplest explanation would maintain that although relative luminance changes at an edge or contour may be combined with a filling-in process between local transitions, some absolute brightness information is still available. When the local luminance transitions predict values which are at variance with the global luminance assessment, the relative information is disregarded or deemphasized. Certainly in Figure 1(c) there is a large discrepancy, since the local transitions would predict an apparent brightness difference of 60% between the rightmost and leftmost bands! With the abandonment of the local differencing strategy, the filling-in strategy is also no longer needed, hence the bands no longer appear to be uniform.

The alert reader has probably noticed that the term "strategy" was slipped into the above discussion, thus implying some cognitive factor which serves as a global modifier of the final result. This is quite correct. There is already evidence suggesting that basic brightness interactions, such as simultaneous brightness contrast, may be altered by cognitive factors (Coren 1969; Coren, Porac & Ward 1979). In fact, whether one obtains brightness contrast effects or brightness assimilation may depend upon some sort of cognitive selection process (Festinger, Coren & Rivers 1970). Since both of these processes also involve some sort of filling in between the contours, it would not be surprising to find that the filling-in process itself may be subject to situational and cognitive factors as well. Certainly such a mechanism could be used explain the results obtained here. In any event, the above demonstration indicates that some sort of global mechanism must be involved to determine when filling in and edge computation occur. Without such an additional mechanism, not only filling in, but also the Grossberg model, must fail.

## Grossberg's "cells" considered as cell assemblies

## G. J. Dalenoort

### Institute of Experimental Psychology, University of Groningen, 9750 AA Haren, The Netherlands

Grossberg has presented some interesting hypotheses on perceptual processes in self-organizing neural networks. They are based on a mathematical formalism, which he has called an "onoff geometry" of which he has been the architect. This formalism is intended as part of a general theory for perceptual and cognitive processes. I share Grossberg's belief in the possibility of such a general theory, and I do not share the expectation of Sperling (1970), although I do agree with his observation. I cannot think of any reason why we should not be able to formulate a general set of equations, as in physics. We must not expect that we will be able to solve it in general, but, again as in physics, we must construct various approximations, and thus obtain relatively simpler equations for specific phenomena and questions we want to answer. And even if the equations would still be too complex (nonlinearities seem to be the rule), it still seems likely that the underlying local principles and mechanisms must be relatively simple, so that computer simulations may be feasible. We must not forget that the physicist too must be satisfied with an understanding of principles and mechanisms for only idealized and simplified cases, and that he will not even consider writing down the Maxwell equations for, say, a television set. Nevertheless, he is able to write them down, and even to solve them, at least approximately, for every separate part of a

complex system. In physics the construction of common-core theories, and the reduction of specific phenomena to the common core, has never been a simple and straightforward affair. We should not despair of the admittedly more complex system of the brain. I consider Grossberg's proposals an important step in the effort, albeit a number of my questions have not yet been incorporated in his model so far. I shall here briefly indicate some of them, which can in principle be incorporated in Grossberg's theory.

His basic building blocks are 'cells'  $v_1, v_2, \ldots, v_n$  (Sec. 21), which are defined in a functional manner (Eq. 1), as I see it. We know that the building blocks of the brain are neurons, and there are sufficient reasons not to identify Grossberg's cells with neurons. But then the question arises, What are these cells? I think the best answer is that they are fuzzy sets of neurons, with excitation patterns that change according to context. This idea originates from Hebb's (1949) hypothesis on cell assemblies, and I have proposed (Dalenoort 1982a) to continue to call them thus, albeit the concept must be brought up to date considerably. It is then also conceptually easy to interpret Grossberg's equation for the changes of interaction strengths between his cells in terms of synaptic modification, which is still the most serious candidate for the process underlying learning; most current neural network models are based on some form of it. Hebb has also postulated that the condition for synaptic modification is more or less simultaneous activity of the neurons concerned, which relates to theories of conditioning. It is remarkable that Hebb's hypothesis on cell assemblies has been neglected so long, and I consider it of fundamental importance that it can be tied to Grossberg's model in a natural manner. For me, it is the "missing link" between Grossberg's functional cells and the neural level (see further Dalenoort 1982a, 1982b).

In attempts to construct networks of neurons which are connected such as to form cell assemblies, and that can do simple computations (Dalenoort 1982b, and more recent work), I have come to the conclusion that a concept cannot be carried by a simple, single cell assembly, but by a cluster of interconnected cell assemblies. Different subsets provide a sufficient number of assembly states; they are excited according to the context. Such complex assemblies must have a quenching threshold (cf. Grossberg, Sec. 28), such that excitation levels under it can "mumble on," but ones above it will quickly and autonomously rise to the maximum level. The quenching threshold is then a collective property of the neurons involved, that themselves can be simple threshold elements. Context can then be envisaged as maximum excitation of a few current concepts (cell assemblies), and under-quenching-threshold excitation of related concepts which are thus primed, and will be more easily excitable in full.

There are good reasons to assume that the neurons of one cell (assembly) are distributed over the network. But this might be difficult to reconcile with the on-off geometry of Grössberg's Figure 10, because the cells  $v_i$  are functional and not localized in space. Therefore it might be preferable to introduce an "activity control center" that keeps the average total activity constant, at least above quenching threshold, instead of a proliferation of mutual inhibitory connections. (cf. Amari 1982). This total activity is attractive to equate with what is usually called arousal. It is reasonable to assume that a maximally excited (complex) cell assembly will take its share of the total allowed activity, so that the number of concepts that can be simultaneously at their maximum is restricted. If we interpret temporal excitations as short-term memory, we would then have an explanation of its restricted capacity.

The interpretation of Grossberg's cells in terms of cell assemblies also seems to solve another problem. How can functional scales serve as a memory for abstract concepts, such as "rectangle." It seems that such semantic properties can better be tied to structural cell assemblies than to dynamical activity patterns alone, given their large variety, corresponding to the large variety of rectangles, which are all categorized as "rectangle." Grossberg's proposal of quantization processes is attractive for, among other phenomena, the filling-in phenomenon, but less so for categorization, and other higher-level processes. Filling-in seems a peripheral process (where "cells" may be more localized), and categorization a more central process.

Finally two comments, of a linguistic nature. I prefer not to use the expression "to store in short-term memory" (Sec. 21) for the models of networks, becauss there is no separate store: the system is considered as self-organizing. I propose to reserve it for the models of the "human, information-processing paradigm" (which, as I see it, is largely restricted to modeling the above-quenching-threshold activity, which indeed is serial in nature). Second, I propose to speak of the "noise-saturation problem" (Sec. 21) since Grossberg solves it, whereas a dilemma implies a choice.

There are many other issues that would deserve mentioning, e.g. controlled versus automatic processing, primitive concepts (also visual ones), spreading activity, etc. In principle, they can all be accommodated in this framework of on-off geometry and cell assemblies, which therefore, to me, provides the outlines for a general theory.

## ACKNOWLEDGMENT

This commentary is partly the result of discussions with Professor S. Grossberg during a stay at the Center of Adaptive Systems, Boston University, in September 1982, which was made possible by support from the Netherlands Organization for the Advancement of Pure Research (Z. W. O.).

## On psychophysical linking hypotheses, the direction of pattern induction, and the representation of distance and size

### John M. Foley

Department of Psychology, University of California, Santa Barbara, Calif. 93106

Unlike some modern theorists, Grossberg takes seriously the myriad of distortions, illusions, induction effects, and global context effects that characterize vision. He undertakes to give a unified account of a larger range of these phenomena than anyone since the Gestalt psychologists. His characterization of the classes of phenomena that a general theory of vision must encompass is a good one, and his proposed competitive feedback network with several structural scales appears to be a promising basis for visual theory. Some critical assumptions, however, are not made clear, and some of the phenomena that the theory is said to account for, particularly those of distance and size perception, are not adequately dealt with.

Any theory of perceptual phenomena must specify not only how the visual system responds to stimulation, but also how that response is related to the experience of the perceiver (what are sometimes called psychophysical linking hypotheses). Grossberg's theory is not explicit about these hypotheses, and consequently it cannot be tested by perceptual experiments. Central to the paper is the notion that if a perceived pattern extends through a region of the visual field, a corresponding pattern of activity must extend through regions of the network associated with that part of the field. In some passages it appears that, for each structural scale, perceived lightness at one point in space is assumed to depend on activity at one point in the network, but in other passages a more complex relation is specified (e.g., Section 38). When the distance dimension is considered, the correspondence between positions in the neural network and positions in visual space is abandoned altogether. Distance is coded in terms of an energy balance across

functional scales (Section 41). Although the binocular depth mixture phenomenon (Foley 1976) suggests a colorlike code for distance, the idea of different representations for distance and frontal extent still seems counter-intuitive in view of the very close relation between these dimensions. There is some hint (Section 3) that Grossberg thinks the kind of representation he uses offers advantages in dealing with the mapping of sensory inputs into other subsystems. The theory would benefit if these assumptions and their rationales were made explicit.

The theory postulates a process for pattern matching with many desirable features. Among these are its directness and its allowance for a flexible criterion of matching, mutual deformation of two patterns, and hysteresis. This pattern matching is produced by excitatory and inhibitory feedback among pattern sensitive cells. Grossberg's proof that the simplest plausible feedback signal that will suppress noise has a sigmoid signal function is important. There is both physiological and psychophysical evidence for signal functions of this form (Legge & Foley 1980; Wilson 1980), but no functional role has been previously ascribed to them.

A central concern of the theory is the propagation of activity patterns through homogeneous regions between edges. This filling-in process is a fundamental property of the competitive feedback network, which is the backbone of the entire theory. The phantom gratings described by Tynan and Sekuler (1975) and Weisstein, Maguire, and Berbaum (1977) are taken as evidence of this perceptual filling-in process. I agree with Grossberg that these induced gratings suggest interaction among cells sensitive to patterns. However, the induced gratings are generated in a direction parallel to the inducing grating, whereas in Grossberg's network the propagation of disinhibition is at right angles to the inducing edge. The much more robust induced grating effect which occurs when the test region is illuminated is also parallel to the inducing grating (McCourt 1982). This induced grating provides another challenge to the theory in that it is 180° out of phase with the inducing grating. The importance of a nonzero luminance in the test field seems to fit with Grossberg's Proposition III (Section 39) that propagation of inhibition occurs only if there exists an input source on which the FIRE can feed, but the FIRE spreads in the wrong direction to account for these induced pattern effects. A second property of the filling-in process is that the propagation of activity between two edges is quantized in the sense that there can exist only an integral number of disinhibitory peaks between any pair of edges. This provides an explanation for hysteresis during deformation. However, there does not appear to be any direct evidence of this kind of perceptual quantization.

In the treatment of the perception of distance and depth the main idea seems to be that larger disparities activate larger structural scales, a conjecture made earlier by Marr and Poggio (1979). This is contradicted by the finding that the maximum disparity that will produce fusion or depth is independent of the luminance spatial frequency of the pattern (Pulliam 1981). There is a more fundamental problem as well. Although in principle, if direction is known, horizontal and vertical disparities together are sufficient to compute distance, in practice vertical disparities do not produce a depth signal, and horizontal disparity is not sufficient to determine perceived distance or depth (Foley 1980). This point is clearly acknowledged (Section 18), yet most of the discussion of distance perception appears to proceed on the assumption that disparity (presumably horizontal disparity, since no reference is made to vertical disparity) determines perceived distance. No provision is made for the fact that binocular distance perception involves the integration of disparity with an egocentric distance signal (Foley 1980).

There is a further problem with the representation of size. Perceived size is postulated to depend on the number of functional wavelengths generated between the representation of two edges of an object, which in turn depends on the disparity and therefore on the object's distance. This implies that perceived size is quantized and, therefore, that the size-distance invariance principle (the real-image analogue of Emmert's law) holds only roughly at best. In fact, if an observer is required to directly match a frontal extent to some fraction of its distance, size-distance invariance holds rather precisely (Foley 1968). Although these comments have focused on phenomena that pose problems for the theory, there is a wide range of phenomena that the theory gives good promise of encompassing.

## Experimental test of a network theory of vision

## David H. Foster

Department of Communication and Neuroscience, University of Keele, Keele, Staffordshire ST5 5BG, England

A frequent problem in evaluating neural-network models is deciding whether they should be tested against neurophysiological or against behavioral data. Grossberg's network model is clearly concerned with behavior. In the full feedback version of his network there are, however, seven adjustable parameters, and with this potential flexibility in specification some difficulty in producing testable hypotheses might be anticipated. In Section 32 Grossberg emphasizes an important qualitative feature of his model: the generation of constrained patterns of spatially periodic activity in response to appropriately bounded inputs. These patterns are initiated at the edges of the input field and propagate within its interior. The failure commonly to perceive these spatial patterns is subsequently attributed to the randomization of spatial phase at high spatial frequencies and to the production of interference effects. Nevertheless, in Section 35 Grossberg does offer a specific qualitative prediction from his theory, an experimental test of which I report below. First, I want to consider some other issues raised by Grossberg.

**Cats, homunculi, metrics, and impletions.** Grossberg uses Attneave's cat (Attneave 1954) to provide some motivation for his edge-initiated filling-in process. He poses the question in Section 5, "Why does interpolation between points of maximum curvature with lines of zero curvature produce a good facsimile of the original picture?" The answer is that it may not always do so, and when it does, it may be trivial. Consider a contour which consists of large sections on which curvature variation is small and small sections on which curvature variation is large. Then a good approximation (in terms of pointwise error) can be achieved by replacing the sections of small curvature by straight lines and the sections of large curvature by points, as was done in Attneave's cat. But for a contour characterized differently, for example, a circle, the approximation does not work.

Grossberg's argument against the applicability of a Fouriertheoretic approach in Section 15 appears in one place homuncular. He argues that after spatial frequency processing a high-pass filtered version of a rectangular function would look like a Cornsweet profile rather than a rectangle. But the requirement here is that the two outputs should be the same, which they are, not that they should look like their inputs.

In Section 22 Grossberg criticizes those models in Artificial Intelligence that use a metric to compute pattern matches. He objects that a metric requires a separate processor to compute the scalar distance between two patterns before deciding how to tack the results of this scalar computation back onto the mainstream of computational activity. But something like this twostage operation may well underlie mental rotation performance investigated by Shepard (Shepard & Metzler 1971) and others (e.g. Just and Carpenter 1976). That is, matches between the patterns are computed at successive stages, and it is these computations which govern the evolving sequence of mental

transformations. His criticism that "a metric does not encode a pattern" misses the point that metrics are used to compare the results of those encodings. That a metric "does not amplify the matched patterns because it is minimized rather than maximized" is merely a matter of convention: A simple algebraic manipulation converts it to a measure of similarity. Nor need a metric be restricted to making "local matches between corresponding input intensities." Metrics may be used implicitly by the nervous system in considerably more complex ways, operating upon whole patterns, for example in determining apparent motion between sequentially presented disparate figures (Foster 1978; Mori 1982), and in determining the discriminability of local features (Foster 1980). On the other hand, it is not easy to see how Grossberg's network explains the easy recognition of reflected and point-inverted patterns, and the easy recognition of patterns differing by dilatations.

The operational novelty of Grossberg's network lies in the edge-initiated filling-in process generated in response to spatially bounded inputs. On the basis of the algorithm described in Section 33, the level of activity generated for a spatially uniform and therefore edgeless field should be much less than that for a large but bounded field. I know of no psychophysical evidence that suggests that the interiors of these large fields should differ so perceptually.

An experimental test. In Section 35 Grossberg proposes a simple but critical experimental test of his theory. He suggests that the high spatial frequency components of the patterns of activity induced by a spatially bounded input can be adapted out by using a spatial frequency adaptation paradigm. A low spatial frequency pattern of activity should then be detectable when fixating a bounded display which is large enough and has a uniform interior. I performed such an experiment in collaboration with Dr. M. J. Musselwhite of this department.

We used a square  $10^{\circ} \times 10^{\circ}$  10 cycles per deg square-wave grating, oriented vertically, as the adaptation field, and a square  $2^{\circ} \times 2^{\circ}$  uniform field as the test field. Both test and adaptation field were presented on zero background. The luminance of the uniform field and the maximum luminance of the grating, both white light, were each 300 cd per m<sup>2</sup>, about 2000 Td. This retinal illuminance of the grating was sufficient to produce significant adaptation at the fundamental spatial frequency (Maudarbocus & Ruddock 1973). The adaptation field and test field were presented monocularly in a standard cyclic mode: 2min adaptation to the grating, and either 2- or 4-sec exposure of the test field. During the adaptation period, the point of gaze was moved continuously over the central two-thirds of the grating; during the test period, the test field was fixated centrally. Subject DHF was practised in performing this type of experiment, and MJM, although not practised, had experience in related psychophysical experiments.

The results of this experiment were as follows. With the 4-sec test field, both DHF and MJM observed no low spatial frequency structure. Interestingly, when the 2-sec test field was used, one of us (MJM) observed two low-contrast vertical bands; significantly, this effect was very much reduced when a fixation point was introduced into the centre of the test field, suggesting that the effect might have been an artifact of edge after-images. Independent of whether a fixation point was used, DHF failed to detect any low spatial frequency structure with the 2-sec test field.

It might be argued that the conditions used in this experiment were not optimal for manifesting the hypothesized low spatial frequency patterns of activity. Yet the effects of the 10 cycles per deg adaptation field should have been sufficiently localized in the spatial frequency domain not to extend beyond 2 octaves, i.e. not below 2.5 cycles per deg, and the test field was sufficiently large to allow at least the generation of patterns with periodicity of 2 cycles per deg.

In summary, this experimental evidence for Grossberg's theory is not compelling. It might be that the conditions of the experiment were chosen inappropriately, but, in the absence of explicit values offered by Grossberg, some choice had to be made.

## Experimental demonstration of "shunting networks," the "sigmoid function," and "adaptive resonance" in the olfactory system

### Walter J. Freeman

(i)

## Department of Physiology-Anatomy, University of California, Berkeley, Calif. 94720

My work and Grossberg's offer a striking case history of what happens when two investigators, one theoretical working from the top down, the other experimental working from the bottom up, meet on mathematical ground. Our commonality holds in particular at the later stages of sensory processing and the earlier stages of perceptual processing. I will comment on three basic features of Grossberg's model: the "feedforward shunting network," the "sigmoid" input-output function, and the feedback network with "adaptive resonance." I will use these to illustrate the developments, modifications, and reinterpretations required in order to realize a working physiological model within the general theory.

The experimental work has been done over the past two decades on the vertebrate olfactory system (Shepherd 1972; Freeman 1975, 1981). In many respects the olfactory bulb is the simplest, most accessible, and best understood part of the brain devoted to higher information processing, yet it manifests transformations of sensory input that have been described in other sensory systems. This accords with Grossberg's intent that his theory hold more broadly than for vision.

A "shunting network" operates in the outer layers of the bulb, which receives axon terminals from the olfactory receptors. This operation has been identified as "presynaptic inhibition" (Minor, Flerova & Byzov 1969); however, it is not merely presynaptic, and it is not inhibition in the classical Sherringtonian sense, because it is multiplicative and not additive.

The operation is mediated by a pool of interneurons in the outermost cellular layer. They are mutually excitatory and form what I have called a  $KI_e$  set (Freeman 1975). A transient excitatory perturbation of these neurons induces a surge of activity that long outlasts the transient and attenuates the transmission of receptor input to the bulb. The agent of attenuation is nonsynaptic and probably involves the release, accumulation, and slow clearance of a substance such as potassium in the extracellular space in this synaptic layer.

Strictly speaking, this subsystem is not a "feedforward" network, because the degree of attenuation depends on the induced interneuronal activity and not directly on the input. Yet the overall operation yields the desired properties of range compression and signal normalization, in a manner formally related to Rushton's (1965) prescription for accommodation in the visual system. Clearly such a mechanism is essential at or near the first central station in every sensory pathway, including the olfactory one, in which the input to a glomerulus might be carried by from 1 to 20,000 axons, depending on odor type and concentration.

Analogously, in simulating the operations of the bulb with nonlinear differential equations (Freeman 1979b), it is necessary to employ a  $KI_e$  set with an output that attenuates the simulated input to the inner bulb, in order to avoid saturation and instabilities of the inner subsystem.

The KI<sub>e</sub> set has other functions as well. It clips and holds the input from a sniff. It provides a degree of contrast enhancement

## Commentary/Grossberg: Quantized geometry of visual space

by the extension of the attenuation effect from each local domain into its surround. It has a stable mutually excitatory state that provides a steady excitatory bias to the inner bulb, maintaining the inner bulbar subsystem in a quasilinear domain. It compensates for surges in activity levels of inhibitory interneurons in the inner bulb that are induced by inhalation (Gonzales-Estrada & Freeman 1980). The KI<sub>e</sub> set also has a zero stable state. In theory the transition from the high to the low state can be induced by a single inhibitory pulse, and can be reversed by a single excitatory pulse. This property might serve as a means for rapidly switching the bulb "off" and "on." However, no means to experimentally demonstrate the requisite inhibitory centrifugal pathway has yet been found.

It is remarkable that these many janitorial functions of range compression, bias control, and the taking of local spatial and temporal integrals and derivatives should be executed by a single population of interneurons, prior to the real work of pattern recognition. These operations are more clearly or complexly manifested in other sensory systems; inevitably they are the first to be encountered by physiologists working inwardly from receptors. On the whole they are well understood as types of sensory preprocessing. As Grossberg notes, the extrapolations to "feature detectors" and "frequency extractors" should be regarded with scepticism.

The existence of the "sigmoid function" for olfactory neural sets was predicted from the properties of bulbar electrical activity and was demonstrated experimentally by computing the probabilities of neural axonal pulses conditional on amplitudes of dendritic potentials (Freeman 1975). An equation describing this function has been derived in part from the Hodgkin-Huxley system (Freeman 1979c). It may well provide for the bulb the properties of noise suppression and signal enhancement described by Grossberg.

Three points deserve brief mention. First, this function represents a collective property and does not hold in this form for single neurons. Second, it is a major determinant of the global stability of the olfactory bulb. Third, the maximal slope of the curve is displaced to the excitatory side of the rest point of the abscissa. This means that input does not merely excite the bulb; it increases the global interaction strength. Thereby the inner mechanism changes with each inhalation from a quasiequilibrium state to a limit cycle state and then back again.

The inner mechanism is formed by excitatory and inhibitory neurons with mutual excitation, mutual inhibition, and negative feedback comprising a KII set, which is related (loosely speaking) to Grossberg's "dipole" (an unfortunate choice of terms, considering physiologists' half century of experience with cortical dipole fields of potential). The output of this set to the next cortical stage is a "wave packet" having a carrier frequency of 40–80 Hz and a duration on the order of 0.1 sec. The repetitive state transition with respiration is the key process leading to the self-organization property of the bulb manifested in these bursts of activity.

Herein is revealed the phenomenon of "adaptive resonance." Information is carried by the wave packet in the patterns of amplitude and phase modulation of the carrier wave in its spatial dimensions. These patterns develop in part in respect to the initial conditions at the onset of the state change induced by inspiration, but in larger part in respect to intrinsic synaptic connections formed during prior experience (Freeman & Schneider 1982). These spatially patterned connections, together with centrifugal controls, serve to define the "expectancy" that is being "matched" in some sense (not correlation) with the spatial pattern of input (Freeman 1979a).

In serial pictures of bulbar spatial patterns of activity reconstructed by computer graphics (Freeman 1973) one can "see" Grossberg's FIRE ignite and spread with each inhalation. Significantly, these "waves" do not propagate; they are standing waves that wax and wane over the time-envelope of the wave packet, while the filling-in takes place.

## Commentary/Grossberg: Quantized geometry of visual space

There are numerous differences and uncertainties to be dealt with between theory and data. The experimental difficulties of precise description and measurement are taxing. However, there is no question that here is a robust convergence of theory and experiment that tells us we are on the right path.

## To have your edge and fill-in too

## W. Eric L. Grimson

### Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. 02139

The basic tenets of Grossberg's target article appear to be that (1) there is significant evidence that different visual properties interact to create a global visual perception; (2) current theories of early visual processing are restricted to analyzing single visual properties and cannot provide a coherent explanation of these effects; and (3) the model proposed by Grossberg provides the coherent basis needed. While there is probably little argument with the first claim, there is some contention as to the tightness of the interaction. In other words, can early visual processing be considered as a system of roughly independent modules which interact loosely to create a global perception, or is the processing so tightly interconnected that the simplest possible description of the process is in terms of its interactions? To use Marr's terminology (Marr 1976a), can early visual processing be modeled by Type 1 theories, or must one resort to Type 2 theories? I will argue that a coherent explanation of a broad range of visual perceptions is, in fact, provided by the Type 1 theories of the computational vision approach of Marr and others, contrary to Grossberg's presentation of that approach. Moreover, the model proposed by Grossberg and the models developed within the Marr approach are differentiated by the fact that the latter attempt to provide explanations at the level of mechanism, algorithm, and computational theory, thereby providing an equally coherent and perhaps more comprehensive explanation and understanding of the visual system, while the Grossberg model is restricted to the level of a mechanism.

Do alternative theories of early visual processing provide an explanation of the cited effects, or do they demonstrate, as Grossberg claims, a lack of coherence and correctness? While it certainly does not constitute a final answer, I will argue that the computational approach of Marr and many others is not inconsistent with the discussed data. Grossberg characterizes this approach as a zero-crossing-based one; however, the definition of zero-crossings that is presented, and consequent inferences based upon it, do not coincide with the viewpoint presented by Marr and coworkers. First, while the zero-crossings of an image filtered with a Laplacian of a Gaussian  $(\nabla^2 G)$  are an important component in the computational theories of early vision, they do not constitute the entire initial representation of the image. Rather, they provide an initial encoding of the changes in the image from which a rich symbolic representation, called the primal sketch, can be computed (Marr 1976b). Second, the term zero-crossing may be potentially misleading or misunderstood. As Marr and Hildreth (1980) clearly state, a step change in intensity at some scale is represented by the combination of a zero-crossing, which marks the location of the change, and an amplitude associated with that change, measured as the slope of the directional derivative taken perpendicular to the zerocrossing segment. This amplitude captures the sharpness and contrast of the underlying intensity change. This is not guesswork, as suggested several times by Grossberg, but follows naturally from the definition of the spatial-frequency-tuned channels and the behaviour of the signal near a zero-crossing in such a channel (see Marr and Hildreth 1980, p. 195) and is

clearly included in both the proposed computation of the edge and in the suggested neural model, contrary to Grossberg's opinion of Section 16.

While only the positions and contrast signs of the zerocrossings may occasionally be sufficient, (for example in stereopsis, (Marr & Poggio 1979; Grimson 1981)), it is clear that the amplitude of the zero-crossing is an important, robust, and easily computable attribute of an intensity change. As has previously been noted, a zero-crossing representation of the image is extremely rich in information. Logan's theorem (Logan 1977) shows that band-pass filtered one-dimensional signals are almost completely characterized by just the positions of the zero-crossings, and extensions by Poggio (1982) suggest that the same may be true in two dimensions. Moreover, a computer algorithm by Nishihara (Winston 1979), based on the positions of the zero-crossings and their amplitudes, is able to reconstruct a close approximation to two-dimensional signals.

One consequence of the richness of zero-crossings (including their amplitude) is that a lightness reconstruction can be executed from them, contrary to Grossberg's claim. For example, Horn's algorithm (Horn 1974) based on Land's retinex theory (Land & McCann 1971), while developed in a slightly different context, clearly illustrates one method for computing lightnesses (which, for certain classes of images, correspond to reflectance) that is compatible with a zero-crossing representation, and that clearly explains effects like the Cornsweet and Craik-O'Brien effects (see Frisby 1979 for a discussion of this). Hence, in those cases where the Horn algorithm is applicable, lightness reconstruction follows from a zero-crossing-based representation.

Moreover, explanations within the context of this representation have been offered for many other standard visual illusions. Examples for which at least partial explanations have obtained include the Poggendorf illusion (Glass 1970), the Mueller-Lyer illusion, and variants involving perceived motion of the illusion undergoing real time blurring (Zucker 1980), the Harmon-Julesz Lincoln image (Hildreth 1980), the change in perception of a checkerboard from edges to diagonals with change in viewing distance (Hildreth 1980), the cafe wall illusion, and so on.

A second class of effects raised by Grossberg as an example of the limitations of zero-crossing-based computations is figureground completion. The examples include the perception of a 5% density Julesz random dot pattern, and the Julesz version of the Shipley experiment. This is a second type of filling-in, similar to the lightness reconstruction, but now based on fillingin depth information. Contrary to Grossberg's claim that "all global filling-in and figure-ground effects hereby become inexplicable in their [Marr and Poggio's] theory," a rigorous computational theory and algorithm for interpolating visual surfaces from depth information based on a zero-crossing representation have been proposed and tested (Grimson 1982a, 1982b). This theory requires no homunculus (a term frequently applied by Grossberg), operates directly on a depth representation (whether from disparity information obtained from stereo, or other depth information), and is based on an analysis of the image irradiance equation (Horn 1977; Grimson 1982b, 1983) and on how the shape of a surface can be inferred from both the locations of the zero-crossings and those locations that do not contain a zero-crossing. The theory has also been given a solid physical motivation, by relating the model obtained on first principles from the image irradiance equation to the physics of thin plates. The perceptions of the 5% random dot pattern and of the Julesz-Shipley experiment follow straightforwardly from this theory and have been demonstrated computationally. Further, since the interpolation process applies in principle across the entire depth representation, any feature evident in one image but occluded in the other will be assigned depth values based on the interpolation between known depth points; hence

it will inherit a depth percept strongly influenced by its nearest binocular neighbour. In addition, the vivid subjective contours associated with sharp changes in disparity (such as those observed in random dot stereograms) are easily localized by the "discontinuities" in the interpolated surface.

One of the major models that has been built upon a zerocrossing representation, and indeed, the model which provided the initial motivation and development of the zero-crossing representation, is the Marr-Poggio stereopsis theory (1979) (with several important modifications suggested by Mayhew & Frisby 1981 and extensive testing reported in Grimson 1981). Many of the effects discussed by Grossberg fall naturally into the domain of this model and most of the criticisms leveled by Grossberg at what he calls Sperling-Dev models do not apply to the MPG model. One already discussed is the interpolation of disparity data to obtain figure-ground effects and to compute complete surface representations. A second is that the Kaufman observation that depth can depend on fixation has been demonstrated in the MPG model (Grimson 1982b). A third is that while Sperling-Dev type models are restricted to a finite set of depth planes, the accuracy of the MPG model is not so limited. Since it has been demonstrated that zero-crossings can straightforwardly be localized to subpixel precision (Hildreth 1980; see also Crick, Marr & Poggio 1980), the computed disparity and depth have similar resolution.

Within the MPG model, which is built directly on processing of a zero-crossing representation within a set of roughly independent spatial frequency tuned channels, effects such as the coexistence of depth and binocular rivalry, multiple spatial scales, and disparity hysteresis follow naturally. Moreover, the model has been demonstrated to be consistent with human perception over a range of images (although not all; see for example Mayhew & Frisby 1981). Examples (see Grimson 1981) include the ability to compute depth when (1) one or both halves of a random dot stereogram are blurred, (2) band-pass filtered noise is added to one image, (3) one of the images is decorrelated at random, (4) one of the images is diagonally decorrelated, (5) one of the images is compressed, (6) one of the images is reduced in contrast, and so on. Moreover, in situations in which humans cannot perform stereopsis, the algorithm has also been demonstrated to be incapable of computing depth, for example, if one half of the stereogram is high-pass filtered and the other half is appropriately low-pass filtered. A second example is the case in which one of the random dot pairs is contrast reversed. We know that human observers cannot fuse these images, and the MPG algorithm also fails. Although the performance of Grossberg's proposed scheme on this range of images is not indicated, and although a precise definition of what constitutes a matching pattern in his scheme is not indicated, one can infer the performance of the scheme from Figure 17. Given the nonpreservation of contrast sign in the structural scales shown in Figure 17(b), one would expect that reversing the contrast of one monocular pattern would result in similar representations. In this case, the pooled binocular edges of Figure 17(c) would presumably still be formed, implying that Grossberg's model would incorrectly (from the point of view of the human system) fuse the contrast-reversed stereogram.

Given that primal-sketch-based theories seem to provide just as coherent an explanation of many different visual perceptions as the scheme proposed by Grossberg (and perhaps have been more extensively demonstrated on actual images, wherein the need to define a precise algorithm ensures that all the details of the process have been explicitly accounted for), how do the schemes differ? Marr, Poggio and others have frequently stressed (see for example (Marr 1982)) the need for explanations of visual processing at several levels, singling out in particular the levels of mechanism, algorithm, and theory. An explanation at the level of computational theory, and vice versa. While it is

clearly important to have an understanding of the mechanisms involved in executing visual computation, it is equally as important to have an understanding at the other levels. Thus, while the proposal made by Grossberg may lead to a better understanding of the underlying mechanisms involved, it is not incompatible with algorithmic and computation-theoretic models, and it is moreover limited in the explanations it can provide at those levels. Two examples illustrate this point. The first concerns lightness reconstruction, such as in the Cornsweet effect. Grossberg's explanation (Section 37) is mechanistic in basis and relies on an explanation of how inhibitory troughs interact within a complex network. It does not say why the effect occurs in a computational sense. On the other hand, the Land-Horn computational theory provides a clear explanation for why the human system "incorrectly" computes lightness in the Cornsweet effect, namely that slow gradients of intensity are discarded in computing lightness, since physically they generally correspond to illumination gradients, and not to gradients in the surface photometry. The second example concerns using computational theories and algorithms to predict and understand mechanisms in the visual system. In particular, the model of the retina due to Richter and Ullman (1982), demonstrating that the X-ganglion cells in the inner-plexiform layer appear to be performing a  $\nabla^2 G$  convolution and that the Y-ganglion cells appear to be performing a time derivative of a  $\nabla^2 G$  convolution, was motivated in part by the theoretical and algorithmic models of zero-crossing computations, and may provide, in part, an explanation of why that mechanism is evident in the human system. More recently, in a study directly motivated by the zero-crossing theory, Richter (work in progress) has obtained preliminary evidence suggesting that some of the simple cells of area 17 respond to zero-crossings in the output of the ganglion cells (where zero-crossings are to be distinguished from intensity step changes). Again, the study followed directly from predictions made by a computational theory, and provides a mechanism-level explanation of the corresponding aspects of the computational theory. In summary, to obtain a complete understanding of the processing of the human visual system one requires a synthesis of models and explanations at all levels, computational theory, algorithm, and mechanism, no one level is sufficient.

## Physiological models and geometry of visual space

### Tarow Indow

School of Social Sciences, University of California, Irvine, Calif. 92717

To be honest, I had not been aware of Grossberg's work until I was sent this target article. Hence, I read the text with great interest, and it reminded me of my experience of almost thirty years ago. That is the impression I had when I studied articles by Selig Hecht, e.g., "Vision: II. The nature of the photoreceptor process" (Hecht 1934).

Models for perception in physiological terms. The Hecht approach is a most systematic mathematical systematization of visual processes at that time. As another example, we can mention N. Rashevsky's (1968) approach. Compared with the latter, however, the Hecht approach is more concrete and much tighter in its connection with experimental data. As early as 1918, Hecht proposed a basic equation for the photoreceptor process. Suppose there is photosensitive material S of amount a. When stimulated by light of intensity I (t = 0), S is changed into its photochemical product which initiates the series of events ending in the outgoing impulse to the nerve. By denoting the amount of that product by x, he assumed  $dx/dt = k_1 I(a - x)$  –

 $k_2x^n \ n = 1$  or 2 and derived a number of equations which fit experimental data on reaction time, dark and light adaptations, flicker, intensity discrimination, visual acuity, etc. All are directly related to x. For example, when the equilibrium state x to I and that to  $(I + \Delta I)$  have a constant difference, then  $(I + \Delta I)$  is assumed to be just noticeably different from I. In other words, all the subsequent physiological processes are assumed to be linear to x. Setting aside this problem, though I was a student, I felt keenly that this type of approach cannot be completely confirmed by psychophysical experiments alone. The final verification has to wait developments of neurophysiological experimental techniques.

It is interesting to see that the Grossberg Equation (1) is of the same form as Hecht's equation except for its third term, and that both make use of the equilibrium state (Equations [2] and [22] in Grossberg's article). In this case,  $x_i$  represents the potential of the ith cell at some stage which is due to depolarization or hyperpolarization. It is true that we feel perceptual phenomena are "accounted for" if these are related to their underlying physiological processes. However, once we use equations which have physiological terms, there is no other way to achieve the final verification than to directly record the assumed processes. Grossberg started from discussion about "theoretical diversity" as "multiplicity of theories," which seems to be "inherent in nature of psychology." I feel that another fatal burden is inherent in the nature of psychology. That is, all models giving explanatory feeling cannot be self-contained. The problem is boiled down to the question of what is more productive, venturing into physiological hypotheses or staying within the selfcontained level. As an example of models belonging to the second category, I would like to mention the Harvich and Jameson opponent process model for color vision. Chromatic response functions consist of stimulus values measured in a psychophysical experiment (cancellation) and, in theory, various quantitative predictions on phenomenology of color are possible from these. The model assigned neurophysiologists a task to look for the corresponding physiological phenomenon at some stage of the process, but, no matter whether this is discoverable or not, the model is a compact description of a perceptual process in terms of the corresponding stimulus intensity.

Geometry of visual space. Another example of self-contained models in perception may be the geometry of visual space. If the global structure of visual space is formulated in terms of a geometry, it is a model to describe perceptual phenomena by themselves or by their correspondence with stimulus configuration. As is well known and referred to in Grossberg's target article, R. K. Luneburg (1947, 1950) proposed a hypothesis that visual space is hyperbolic and of constant curvature to "account for" the discrepancy between parallel- and equidistance-alleys, etc. This is a kind of explanation different from that offered to 'account for" perceptual phenomena in terms of physiological processes. For example, the two alleys are not the same "because the visual space is non-Euclidean." However, it is a different question to ask why the visual space is described by such and such a geometry, including the Euclidean case, too. As I have been working on this problem (Indow 1979, 1983), I expected a great deal from the title of Grossberg's article and by his promise "to explain why so many visual percepts seem to occur in a curved visual space." However, what is given in Section 41 is too abstract and not really informative. Edgedetection and figure-ground separation are discussed, but it is hard to see how to derive from this approach real geometrical properties like straightness, perpendicularity, parallelness, congruence, etc. It will take a long time before we can reach the stage of having a concrete image of the neurophysiological processes underlying these perceptual aspects. To systematize perceptual processes in their own terms will be useful to hasten the arrival of that stage.

## The role of analog models in our digital age

## Bela Julesz

## Bell Laboratories, Murray Hill, N.J. 07974

It is a dilemma what to do after reading Grossberg's interesting article on neural modeling, particularly of modeling global stereopsis of random-dot stereograms (RDS), a kind of problem that has preoccupied me throughout my scientific career. Shall I clutter the literature by repeating the content of dozens of articles of mine from 1960 till 1971 that were omitted by Grossberg (during which the notions of "false target elimination," "cooperativity," "hysteresis," "multiple stable states," "globality," "plasticity," "filing in," etc. were first introduced and observed in the field of stereopsis, culminating in my spring-coupled-magnetic-dipole-model of stereopsis to account for these notions), or refrain from commenting at all? After all, Grossberg did refer to several of my contributions, including my book (Julesz 1971b) that contained all these findings; he even referred to my analog dipole model in a sentence: "My dipoles are not the classical dipoles which Julesz (1971b) used to build an analog model of stereopsis.'

Since Grossberg reviews neural network models of stereopsis, including his own, the omission of my analog model (and my computer simulated recursive and cooperative model of stereopsis, called AUTOMAP-1 [Julesz 1962] published nine years before all the models cited) must mean that Grossberg either did not regard the dipole model as "neural," or "mathematical," or interesting.

Let me stress that I consider Grossberg's model and his mathematical insights to be of great importance for psychobiology, but at the same time I want to explain why I regard analog models, such as my own, as an alternative if not a better way to communicate mathematical ideas and complex perceptual phenomena to workers in brain research. In doing so, I claim that explicit differential equations describing neuronlike elements in pools are not more "scientific," "neural," or "mathematical" than implicit morphological models. As a matter of fact, my analog model of stereopsis has a deep "neural" content, and is superior to existing algorithmic models on two accounts. First, it is matched to human heuristics (i.e., its workings can easily be grasped by the intelligent layman); and, second, it can account for *both* cooperativity and plasticity, which most other models cannot do (or have these properties deeply hidden in their deep structures).

In order to keep this commentary short, I ask the reader to consult the original papers describing my analog dipole model of stereopsis. I introduced this model in order to account for the many cooperative and plastic phenomena in stereopsis, particularly the Fender-Julesz hysteresis effect under binocular retinal stabilization (Fender & Julesz 1967). (A RDS has to be brought within 6 min arc alignment for fushion, but after fusion it can be pulled apart up to 120 min arc on the retinae before fusion breaks.) The plasticity of being able to fuse RDS pairs of different sizes (Julesz 1960) could easily be explained by the model. The first serial version of this model I described in my MIT lectures in 1969 and in Berlin, April 1970 (Julesz 1971a). My book contains an entire chapter devoted to this model. Since I showed that a model of stereopsis must be parallel (Julesz 1964), since a 2% unambiguous bias can pull an ambiguous RDS into the biased state (whereas in a serial process attaining first a 96% correlation instead of 100% would yield a good match, de facto, stereopsis seeks 100% matches), we introduced a parallel spring-coupled model (Julesz & Chang 1976). This parallel model, with its neural implications, is further discussed for the expert in the Handbook of Sensory Physiology (Julesz 1978a) and for the intelligent layman in an American Scientist article (Julesz 1974).

In essence, my spring-coupled-magnetic-dipole model is the

generalization of the zipper (Julesz 1974). A left and right magnetic dipole array (with adjacent magnets in each array coupled by springs) glide over each other. The black and white dots in the RDS polarize the corresponding magnetic dipoles in two possible ways, and the spring coupling is switched on only after polarization. The left and right dipole arrays have opposite polarization, so when they slide above each other correlated domains contain facing magnetic dipoles that become in-terlocked. After such a "hyper-dipole" is formed, one can slide the arrays farther away (in search of corresponding areas with larger binocular disparities) without causing the already interlocked hyper-dipoles to unlock. I regarded this "generalized zipper" as a rather unique device. It permits us to search for correlated areas in a 2-D manifold. After an adequately large correlated area has been found, it will interlock, and resist unlocking as we seek other matches with larger disparities by sliding the dipole arrays (or deforming them). If by chance sporadic patches interlock, they are usually small, and the sliding of the arrays will exert tension through the springs that will unlock the small interlocked patches (that constituted false matches).

In the articles cited I give many illustrations of how the model can easily account for the Fender-Julesz hysteresis effect, or for the fusion of 15% expanded half-stereopairs. Particularly important is the explanation of why the weakening of stereoscopic local interaction reduces perceived global depth. For instance, Frisby and Julesz (1975) demonstrated that random-line stereograms in which the left and right corresponding line segments (needles) were perpendicular to each other still gave global depth percepts, but the perceived depth was much less than in stereograms where the corresponding needles shared the same orientation. Most models based on pools of disparity-tuned neurons cannot account for such a reduction in perceived depth as the local binocular interaction between corresponding elements decreases. For the spring-coupled-dipole model the weakening force between the interlocked dipoles permits the neighboring springs to turn the dipoles. And it is the turning of interlocked dipoles (more precisely, the horizontal projection of this turning) that corresponds to perceived depth.

Such examples and similar ones are given in the articles cited to demonstrate the power of explanation achieved by the analog dipole model. So far so good. But how can magnetic dipoles and their rotations have neurological significance? This is also explained in the articles (Julesz 1974, 1978). One rotating interlocked dipole corresponds to "reading out" the output of different binocular disparity-tuned neurons that inspect the same retinal locus, while the joint parallel rotation of many adjacent interlocked dipoles corresponds to reading out the output of pools of binocular disparity detectors tuned to the same disparity, but inspecting many different retinal loci.

It is these two "hypercolumns" or "processing pools" that are the two hyperunits of global stereopsis of RDS. Indeed, Gian Poggio (1980) recently found in Area 17 layers IVB and IVC of the behaving monkey cyclopean units that fire for dynamic RDS (at 100 frames/sec rates) without any monocular density (contrast) difference, except binocular disparity.

Let me stress that I derived my dipole model deliberately, in place of mathematical models (and to replace my less satisfactory AUTOMAP model based on computer simulation). Originally, my background was in applied mathematics, particularly in nonlinear oscillator theory, before my metamorphosis into a psychologist. For years before this metamorphosis I was interested in pulling the frequency of a powerful oscillator (magnetron) by a much weaker one (klystron). Obviously, my first cooperative model of stereopsis was formulated between coupled oscillator arrays, where the frequency pulling between corresponding left and right oscillator pairs had corresponded to the sensing of depth at that locus. However, to my amazement, this rather elegant model was not comprehensible either to the physicist or to the psychologist. Only now, when "limit cycles," "bifurcations," etc. are part of the physicist's curriculum would such a model have wider appeal.

So, for the "far-interaction" I decided to use magnetic dipoles and for the "near-interaction" spring couplings. I found that suddenly I could communicate with almost everyone who in childhood had played with magnets and springs. I think I succeeded in translating the complex phenomenology of global stereopsis into a heuristically matched model, as attested by the many references to this model by both physicists and psychologists. As a matter of fact, this model is one of the simplest cooperative models known, simpler than Ising-models, since the "order parameter" is just binocular matching between already polarized left and right dipole arrays.

It is interesting to note that the cooperative model of stereopsis by Sperling (1970), which he developed independently from my dipole model about the same time or slightly after, consists of two parts: an "energy well" model in which convergence movements of the eye distort the shape of the well enabling a ball to roll down into the pit, or get caught in an earlier dip. This global analog model is supplemented by a neural model, and the correlation between the two models is not worked out in every detail. Nevertheless, it is the "energy well" model of Sperling that had the greater influence on brain research, while his neural model is much less cited. Let me note that the Fender-Julesz hysteresis effect, which occurs in the total absence of convergence eye movements, has always been a difficult problem to reconcile with the Sperling model in which "vergence energy" has a substantial role. Sperling himself wanted to know about further reconfirmation of this hysteresis effect. This has been reconfirmed by Diner (1978) in his Ph.D. thesis, in which he used binocular retinal stabilization and found hysteresis both for temporalward and nasal pulling (while Fender & Julesz 1967 used only temporalward disparities).

In summary, I wanted to stress that complex mathematical ideas do not have to wait for sophisticated mathematical tools. The conceptual models already existed years ago enabling researchers to account for many complex perceptual phenomena, at least qualitatively, and to predict a few new ones. That the novel mathematical insights enable Grossberg to elucidate global perceptual phenomena by neural computations is an important step. Whether the quantitative tools are powerful enough to predict novel global phenomena (experimentally not yet observed) remains to be seen.

## On the need for discipline in the construction of psychological theories

### Donald Laming

University of Cambridge, Department of Experimental Psychology, Cambridge CB2 3EB, England

My comment is chiefly metatheoretic and is succinctly expressed by an analogy:

Imagine a large jigsaw puzzle, possibly a very large puzzle. No one knows how large the puzzle is because some unknown proportion of the pieces are missing – we can each of us think of interesting and relevant experiments of which the outcomes are not yet known. The picture which usually accompanies such a puzzle is also missing – so that each person's vision of what that state of nature will ultimately prove to be is restricted to what he can conjecture from assembling the available pieces. Grossberg's target article presents his synthesis of those pieces which relate to visual perception.

There are two stages in such a synthesis that are, to varying degrees, arbitrary. First, with a very large jigsaw puzzle it is a useful strategy to gather pieces with some identifiable charac-

## Commentary/Grossberg: Quantized geometry of visual space

teristics – green pieces, say, that might all be part of a tree – in the hope that they all belong to the same region of the puzzle. Even so here; different theorists make slightly different selections of experimental findings on which to ground their theories. Second, unless the selected puzzle pieces actually interlock, it is a matter of conjecture how they should be disposed with respect to each other, and each theorist brings his own particular intuitions to guide his conjectures. This much is apparent from the number of theoretical essays and the variety of mathematical formalisms which Grossberg lists in his opening paragraph.

Now, if one is able to gather a high proportion of the pieces from some small region of the puzzle, there is little uncertainty about how those pieces should be assembled. Such a situation is identifiable in practice because the shape of the theory is, as it were, forced by the experimental data, as is common in the physical sciences. And the correct assembly of even a very small part of the puzzle is a useful exercise, not least because it points the way to crucial experiments that need to be carried out. But if only a small proportion of the pieces is available, the manner of their assembly will depend much more on prior intuitions about what the state of nature looks like than on experimental evidence. And the attempt to assemble those pieces may actually be counterproductive; his prior intuitions may cause the theorist to overlook other related pieces, or even to deny the relevance of such pieces presented to him. No one knows that proportion of relevant empirical findings is at present to hand in the field of visual perception; that is a question about how much more remains to be discovered and is, of necessity, a matter of opinion. But, in view of the dangers of premature theorization the history of psychology is littered with it - I think we should demand of any theoretical essay, such as Grossberg's, a high density of experimental evidence linked by a parsimonious set of theoretical constructs. Grossberg's article, to my mind, presents the contrary complexion.

Grossberg might reply, with some justification, that his work here and elsewhere presents a detailed exploration of the properties of cooperative and competitive networks and is valuable, at the least, on that account. But why explore *these particular ideas?* Well, the visual pathway is known to be composed of discrete cell bodies interconnected by discrete neural processes – naturally modelled as a network. Other known physiological properties, such as "saturation," are imposed on those networks and exploration proceeds within these constraints. The search is on for those kinds of network, subject to the chosen constraints, which best match the phenomena of visual perception.

There are some hidden assumptions here which need to see the light of day. First, even though visual perception is mediated by an assemblage of neural units, each of which may be modelled by a network, it is by no means necessary that particular properties of those networks be germane to the psychological process of perception. There are very large numbers of neural units involved and the phenomena of visual. perception might depend rather on the behaviour of aggregates of neural units in which, as in the Central Limit Theorem, the idiosyncracies of the individuals are lost; the discrete nature of the charge on an electron is of no consequence in electrical engineering. Second, it is not necessary that all of the visual pathway be essentially engaged in the mediation of perception all of the time. For the most part visual perception is veridical, suggesting a basically linear process, so the saturation of neural units which Grossberg models as a sigmoid signal function (Figure 14) may not be essentially to perception at all; it might, instead, simply be a physiological device for suppressing the contribution of certain units, allowing the perceptual process to be carried by other units still operating within their linear dynamic ranges. Third, there is a minor obsession with "the filling-in dilemma," as though visual perception consisted essentially of the realization in the head of a morphologically similar model of the physical world. This, I assert, is not

necessary, and the filling-in dilemma is an artificial problem.

If I may be permitted to make a selection of pieces from the jigsaw puzzle somewhat different from that favoured by Grossberg, it can be shown that visual perception is differentially coupled to the physical world (Laming 1973, pp. 145–150). The crucial observation in this demonstration is that the psychometric function for the detection of an increment  $\Delta L$  added to a continuous background L is approximately a normal integral with respect to the square of  $\Delta L$  (Leshowitz, Taub & Raab 1968; Nachmias & Kocher 1970), so that the psychophysical discriminator must see a quantity proportional to  $(\Delta L)^2$ . Such a quantity can be produced only if the increment be first stripped from the background in the course of visual processing, and this means differential coupling.

Now apply this insight to the perception of the Cornsweet figure (Figure 5). Only changes in luminance are available as a basis for perception. The abrupt transitions are easily perceived, while the ramps present small gradual changes which are made even smaller by square law processing and are submerged in the background noise (an instance of "low-signal suppression"; Swets 1961). So the observer perceives only the abrupt changes in luminance and (presumably) cannot distinguish the Cornsweet figure from a similar figure with real changes in luminance and no ramps. Of necessity (Figure 14[c]) the two figures look alike. If the sensory information available to the observer from the Cornsweet figure be reflected outwards to the physical level of description, traversing the processing system in reverse direction, it will be found to correspond to the similar figure with real changes in luminance; and this, I insist, is a sufficient explanation of the phenomenon.

## Toward a unified theory of visual perception

## Daniel S. Levine

## Department of Mathematics, University of Houston, Houston, Tex. 77004

Stephen Grossberg's target article represents a major effort to tie together theories of various perceptual phenomena by the same investigator and his coworkers. Such diverse visual phenomena as Weber law modulation, depth perception, edge enhancement, certain illusions in orientation perception, and phantoms are explained as part of a unified theory with a neurophysiological basis. Other popular theories for different aspects of vision, such as Fourier analysis for spatial perception and cooperative-competitive interactions between disparity detectors for depth perception, are partially supported and partially supplanted.

In Section 13 of the target article, Grossberg reviews a range of visual experiments that indicate "the mutual interdependence and lability of depth, form, and lightness judgments." It had been known since the work of Emmert in 1881 that monocular cues are in some cases insufficient to estimate an object's length, so that binocular cues, i.e., depth cues, will influence such a judgment. Other instances are shown where apparent depth influences perceived brightness, or vice versa.

The interaction between percepts commonly treated as separate ones points to the inadequacy of commonly accepted theories of those same percepts. The influence of size and brightness cues on depth perception suggests that while disparity of the locations of the images of an object on the two retinae is a major determinant of perceived depth, it is not the only determinant. Such mutual influence of percepts, Grossberg states, also implies feedback, and it is therefore incompatible with a purely feedforward theory like the Fourier theory of spatial vision. The Fourier theory is also linear and therefore hard to reconcile with previously established nonlinear phenomena such as reflectance processing.

Grossberg accomplished the modeling of these data by means of a shunting competitive, or on-center off-surround, network of neuron populations obeying the neuron membrane equations. Various examples of such networks have been used in his models and those of his coworkers (Ellias & Grossberg 1975; Grossberg 1973, 1976a, 1976c; Grossberg & Levine 1975; Levine & Grossberg 1976) and related networks have appeared concurrently in the work of several other neural modelers such as Sperling and Sondhi (1968), Wilson and Cowan (1972, 1973), and Amari (1977). There were two primary reasons for introducing this class of networks, one neuroanatomical and the other psychophysiological. The neuroanatomical reason was that similar networks have long been known to exist in both the horseshoe-crab eye and the vertebrate retina, and much evidence points to their existence in the visual and somatosensory cortices. The psychophysiological reason was that such networks are efficacious for contrast enhancement and noise suppression. In Grossberg's terms, the feedback competitive dynamics with shunting interactions overcome the noise-saturation dilemma: how to prevent parts of the pattern with small intensity, or network noise, from being amplified and flooding the network. and differences between parts of the pattern with large intensity from being blurred.

One of the two largest schools of thought about complex visual processing emphasizes the detection of specific features, whereas the other school emphasizes the perception of spatial frequencies. Various attempts have been made to synthesize the feature-detection and spatial-frequency emphases, and Grossberg's is one of the best such attempts. His papers from 1973 to 1976, including his joint work with me and with Ellias, stressed festure detection. Specifically, in much of this work, the populations making up the network were columns in the visual cortex, each sensitive to a given position or line orientation. How detectors of various features were chosen during development, and thus how sensory codes were established, was also considered.

In more recent work, including the present article, Grossberg has indicated that the same set of networks can be utilized to explain spatial frequency data as well. Specifically, a realistic oncenter off-surround anatomy will have interpopulation interactions depending on distance, because the density of cell connections decreases in approximately a Gaussian manner with distance. The "on-center off-surround" property means that inhibitory interactions have a wider spread than excitatory interactions. Both of these spreads interact with the external input to determine natural spatial scales of the network. Thus, under certain conditions, a network's response to a uniform rectangular input can include spatially periodic patterns, analogous to "phantoms" which are sometimes seen in real life.

Grossberg's network provides insight into two questions that are of concern to artificial intelligence theory as well as to biology. The first question is, How does a network decide whether two input patterns from different sources, such as from the two eyes, match? Most artificial intelligence theories relate matching or mismatching to a metric which measures the absolute difference or squared difference between intensities of two patterns at given locations and then sums these distances. Such a metric performs its function but requires a separate processor that is apart from the network. Grossberg shows that matching or mismatching can arise out of the dynamics of the competitive network: "A given cell can receive two different inputs, yet these inputs may be part of perfectly matched patterns, hence the cell activity is amplified. A given cell can receive two identical inputs, yet these inputs may be part of badly mismatched patterns, hence the cell activity is suppressed" (Section 22). Such matching or mismatching is shown to subserve binocular fusion or rivalry, and transitions from one to another.

The other question is, How is edge processing supplemented by processing of the interiors of a pattern? A contrast-enhancing network will tend to enhance edges and suppress locally uniform patterns. Yet if an edge is a boundary between light and dark regions, the visual system needs to be able to tell which is which. The functional spatial scales that arise from excitatory and inhibitory spreads are shown to be a means for filling in the space between edges.

Grossberg suggests that functional scales lead to a "quantization" (i.e., discretization) of perceived visual space. Perceived length of an object is interpreted as a number of "quanta," so that any input changes, such as brightness or depth changes, which affect the functional scales will also affect length judgments. A variety of illusions where the same object appears to be of different size in different contexts, such as the corridor and Ponzo illusions, can also be explained in this manner. Grossberg asks the question of how such quantization of space is compatible with the fact that depth seems to change continuously, not in jumps, as an object is moved toward the observer. A partial answer is suggested in terms of energy balance but the question is still an open one.

Grossberg's paper is closer than any other I have read to a comprehensive theory of vision. The middle sections of the article which review the structure developed earlier for the contrast-enhancement network could be somewhat shorter for better presentation of the many important new ideas contained herein. Some aspects of visual processing are not discussed here, such as interactions between different retinal and cortical levels (except for monocular-binocular interactions); these questions are partly covered by some of the earlier papers referenced here and are partly still open ones. Grossberg's overall theory poses a formidable challenge to both the experimentalist and the theorist in visual psychophysiology. Anyone possessing data on size, depth, and form perception should investigate whether those data are accounted for by the theory herein and, if a counterexample should be found, suggest an alternative theory which accounts for as large a body of data as Grossberg's.

## Failures of Grossberg's theory to compute depth, form, and lightness

Steven E. Poltrock<sup>a</sup> and Marilyn Shaw<sup>a,b</sup> <sup>a</sup>Bell Laboratories, Murray Hill, N.J. 07974; <sup>b</sup>Rutgers University, New Brunswick, N.J. 08903

Grossberg's theory lends itself to several interpretations that differ in degree of theoretical importance. First, it can be interpreted as Grossberg proposes it, as a theory of perceptual experience. According to this interpretation, differentials in neural activity, as described by shunting networks, exactly define the percept. This position can be assailed simply by finding situations in which the activity of the network fails to match the percept. A weaker interpretation is to assume that shunting networks describe an early stage of visual representation, providing a data base for further perceptual computation. In this case, mismatches between network activity and perception can be attributed to later processes. A third and still weaker interpretation is that shunting networks provide a mathematical tool, similar to probability theory, that could be useful to the psychological theorist. Grossberg's analysis of shunting networks helps the investigator select a network with characteristics to match specific functional theories of perception.

Our criticisms focus on the first, and strongest, of these interpretations. Suppose, as Grossberg asserts, these computational systems provide a theory of experienced perception. Are the requirements that we normally impose on a psychological theory satisfied by the shunting networks? We expect a theory to provide explanations of observed phenomena and predictions

## Commentary/Grossberg: Quantized geometry of visual space

of the outcome of future research. Grossberg attempts to explain perception of depth, length, lightness, and the interaction among these domains. We raise questions about the ability of the theory to account for critical phenomena in each perceptual domain. Indeed, we suggest that, in some instances, his explanations are inadequate for the phenomena he considers. Furthermore, Grossberg fails to provide any clear predictions from the theory. When a prediction is offered (for example, for perception of phantoms), reasons why the prediction might fail are advanced.

**Perception of length.** The theory stipulates that length is coded as the number of excitatory peaks in a structural scale; this code is called a functional scale. The peaks are generated by a filling-in resonant exchange (FIRE) driven by the luminance in the interior of an object. In general, any given object will be represented by a different number of peaks in different structural scales. The perceived length of an object is determined by the number of peaks in the most active structural scale, which is determined by binocular matching. Thus in a monocular presentation length is ambiguous because self-matching prevents the differential excitation of structural scales that is needed to determine which functional scale is appropriate.

Tachistoscopic length. We now consider what appear to be consequences of this theory of length perception. Consider how the network would respond to binocular, tachistoscopic presentation of a white line (or bar) on a black field. The binocular presentation will insure that an appropriate scale is used to measure length. The network will respond by sharpening the edges of the line (or bar), then feedback will initiate the filling-in process, thereby defining the interior. However, the offset of the stimulus will remove the luminance required to drive the filling-in process. Thus it seems that the observer should judge the line (or bar) to be shorter (narrower) than the same stimulus presented for longer duration. In addition, the stimulus will appear dark in the center and brighter near its ends (edges).

While these predictions follow quite naturally from the theory, we would be surprised if they are confirmed by experimentation. Of course, before such an experiment could be conducted we must have quantitative descriptions of temporal characteristics of the FIRE, and we must know whether a mask is required to quench the FIRE. To permit a rigorous test of the theory these characteristics should be established on the basis of prior experimentation and model fitting. At present the theory does not appear to be described in sufficient detail to decide even the form of the shunting network in advance.

Black-white asymmetry. As we understand the theory, an asymmetry in the perception of dark and light stimuli is implied. A vertical black bar presented on a white background cannot generate a functional scale because the low luminance of the black stimulus cannot drive the FIRE. Thus the observer lacks any information about the width of a black stimulus. An observer should be incapable of comparing the widths of black and white stimuli, as the white stimulus will always have width defined by the functional scale, but width will be undefined for the black stimulus. If the theory truly requires such an asymmetry in perception of black and white forms, then the theory is incorrect.

**Perception of depth.** The theory represents depth by the relative activations of structural scales. Objects at greater distance are assumed to differentially activate fine scales, and near objects activate coarse scales. The differential activation is a result of binocular matches of disparate stimuli. Grossberg argues that a primary advantage of his theory is its ability to represent depth and the interaction of depth with length and brightness. However, a theory that claims to account for perceptual experience must explain how both binocular and monocular depth cues result in appropriate activation patterns.

In apparent contradiction, Grossberg claims that monocular stimulu (1) self-match, thereby activating all structural scales equally, and (2) contain texture information that causes differential activation of structural scales similar to the patterns caused by binocular matches. Unfortunately, we do not understand how texture gradients supplant self-matching in the determination of scale activation. More generally, it is unclear how the theory can accommodate the range of effective monocular cues to depth. As a particular example, consider the profound effect that interposition has on perception of depth. When placed in conflict with binocular disparity, interposition typically dominates the perception of depth (Schriever 1925). How can interposition differentially activate structural scales, overwhelming the pattern induced by disparity?

Grossberg suggests that monocular illusions of size and depth, such as the Ponzo illusion, result from activation patterns in the structural scales that are equivalent to those produced by an object in depth. The texture of the stimulus is presumed to be computationally equivalent to the corresponding object. This claim is vulnerable to both a logical and an empirical argument. First, his principle of scale equivalence implies that the converging lines in the Ponzo illusion should be experienced as receding in depth. In fact, the depth is correctly perceived at the same time that the length of the cross bars is misjudged. Second, it has been observed that Zulu tribesmen, who lacked experience with linear architecture, failed to perceive these illusions (Deregowski 1973); This seems to pose a serious problem for the theory because it claims that the activation pattern arises naturally from the computational processes required for binocular vision. Therefore, these cultural differences should not be observed.

Grossberg suggests that the theory can account for Gogel's (1956, 1965, 1970) observations that a monocularly viewed object acquires the depth of the closest binocularly viewed object. He explains that a FIRE spreads from the binocularly viewed object to differentially excite the structural scales of the monocularly viewed object. This explanation seems in conflict with earlier arguments that object boundaries act as filling-in boundaries to prevent the spread of FIRE. Indeed, the activity would have to spread through a black region separating the objects, violating the requirement that luminance is required to drive the resonance exchange.

**Perception of brightness.** The theory represents brightness by the relative energy in a functional scale corresponding to an object. Brightness constancy is explained as resulting from normalization activity in a structural scale. Thus the theory predicts, correctly, that stimulus brightness is judged relative to other stimuli at the same depth. However, it is unclear how the theory would account for certain findings regarding depth and brightness.

First interposition is an effective means for determining perceived depth and thus perceived lightness (Gilchrist 1977). (Though Gilchrist distinguishes between brightness and lightness [shades of gray], Grossberg uses these terms interchangeably.) Furthermore, Hochberg and Beck (1954) showed that perception of brightness depends on the perceived source of illumination and angle of view of an object, and cannot be explained by the pattern of retinal stimulation. An upright trapezoid illuminated from above and viewed monocularly appeared to be a dark, horizontal square until a stick was waved behind it to establish its true orientation, and thus its brightness. The retinal stimulation corresponding to the trapezoid was unchanged, but the depth information provided by interposition of the trapezoid and the stick were sufficient to alter the percept. Thus the theory must provide an explanation of how interposition selects an appropriate scale.

**Conclusions.** In attempting to evaluate the adequacy of shunting networks as a theory of perceptual experience we noted several inconsistencies between the theory, as we understand it, and perception of length, depth, and brightness. However, we anticipate that the theory may be readily modified to account for at least some of these phenomena. This adaptability of the theory is attractive but poses a problem for the research psychologist. Grossberg seems to avoid settling on a specific, testable form of the network, relying on different shunting networks to explain different perceptual phenomena. Thus the theory cannot predict perception but is modified to match perception. We wonder whether any perceptual experiences are excluded, given the flexible way in which Grossberg adapts the mathematics of shunting networks.

Possibly there is no ready answer to how monocular cues such as interposition, height in the picture plane, and aerial haze modify structural scale activity to affect perception of depth. Perhaps it is more appropriate to restrict the theory so that the shunting networks compute an early stage of visual representation that differs from, but constrains, the perceptual experience. Adopting this perspective might suggest reasonable limits on the domain of perceptual phenomena to be explained by the theory. Alternatively, Grossberg's development of the mathematics of shunting networks may prove most useful as a general tool for the building of specific models.

## On non-quantum quantization

## Robert Rosen

Department of Physiology and Biophysics, Dalhousie University, Halifax, N.S., Canada B3H 4H7

Most pioneering work in theoretical science has an essential epistemological component, and it certainly has radical epistemological implications. This is clearly true of Grossberg's work, and at several levels. In these brief comments I shall consider only one limited aspect: Why do the paradigms of physics, which are supposed to be universal, prove to be of such little use in dealing with integrative organic phenomena? In particular, as Grossberg himself says, why should quantum theory be so linear, while phenomena of self-organization are so stubbornly nonlinear? Is there something essential missing from the physics, or are we missing something in the biology?

There are a number of alternative approaches to quantum theory, but they are all mathematically equivalent. Thus their physical content is also identical. Tracing back from Schrödinger's original formulation, which is intuitively perhaps the most transparent, we find that it has its roots in Hamilton's Principle, which is a form of Least Action law going back to D'Alembert. In Hamilton's formulation, we find an essential restriction; although it can encompass systems for which the impressed forces are time-dependent (i.e. nonconservative systems), it cannot encompass those for which these forces are not gradients of some scalar potential. Likewise, the forces of constraint on the system must be holonomic if Hamilton's Principle is to be valid. In other words, the apparatus of quantum theory is valid only in situations entirely governed by scalar potentials.

Physicists have been fortunate that such situations are common enough, and important enough, so that this limitation can be ignored by them. Biologists are not so fortunate; the systems with which they must deal essentially violate these basic restrictions. Moreover, biologists must concern themselves with transients, and not just stationary or steady-state behavior.

Look, for example, at Grossberg's basic dynamical equations (1). Not only are they nonstationary (nonconservative), but they cannot, even in principle, arise from the scalar potentials so dear to physics. Indeed, in physical terms, such a system of first-order dynamical equations would, in mechanics (either classical or quantum) translate into a situation in which the system possesses no degree of freedom (recall that an unconstrained mechanical system with n generalized coordinates possesses n degrees of freedom; such a system loses a degree of freedom whenever a nonholonomic constraint is imposed on it). Therefore, such a system is as far away from obeying Hamilton's

Principle as it possibly can get. As such, it cannot possibly be described by any formalism derived from quantum theory.

Seen in this light, Grossberg's suggestion that the characteristics of biological self-organization should be taken as fundamental is a challenging one. Could it be, for example, that the familiar paradoxes of quantum theory, which have led many thoughtful physicists, starting with Bohr, to deny that the quantum-mechanical formalism has any single self-consistent physical interpretation, arise because the interaction of a system with a meter involves nonpotential forces, and thus violates the conditions under which quantum mechanics itself is valid? If so, then the properties of biological meters, as in the visual system, may enter in an essential way into the very heart of physics. Studies like those of Grossberg, which bear precisely on these points, will provide a basic conceptual prerequisite for dealing with this matter, and physicists ignore such possibilities at their peril.

## Field of feature detectors or features detected by a field?

## Robert L. Savoy

## The Rowland Institute for Science, Inc., Cambridge, Mass. 02142

The author requests that his article "be viewed as a summary of an ongoing research program, rather than as a completely tested visual theory" (Section 18). In keeping with that request, I will focus on several questions that the theory raises, my intent being to highlight those aspects of the theory that seem to be the most novel, the most surprising, and/or the most testable. Given the range of topics that Grossberg touches on, and the necessary brevity of the present commentary, I will only be able to discuss a select group of such issues.

In Section 24 Grossberg says that "the noise suppression property in the network (22) [feedforward, distant dependent system] implies an edge detection and spatial frequency detection capability in addition to its pattern matching capability." It seems to me from the subsequent discussion that until the  $D_{ki}$ and  $E_{ki}$  are specified explicitly, this is largely equivalent to: It responds to spatial nonuniformities. In the case of an edge, this is a useful property . . . an edge is in some sense the canonical (and most extreme, naturally occurring) spatial nonuniformity. We want a system's response to an edge to be particularly strong. But why are spatial frequencies special?

The answer, in Grossberg's model, is that the particular choice of  $D_{ki}$  and  $E_{ki}$  in a given network defines a "structural scale" that can exhibit preferential spatial frequency response. One could say that Grossberg (like other theorists) is postulating a group of different spatial interactions to obtain different frequency channels. This property is not inherent in Equation 22. Assuming various  $D_{ki}$  and  $E_{ki}$  distributions, as in Figure 11, is logically analogous to assuming various receptive field sizes of center-surround cells.

But Grossberg's model, while getting a variety of spatial scales by fiat, represents the activity of those scales in a completely different way than is envisioned in a feature detector model. In fact, nothing short of the whole field of "cells" in Grossberg's network could be called a "feature detector" if the feature being considered is an oriented bar.

The idea is that, in the usual, hierarchical feature detector conceptualization, center-surround cells (one class of feature detector) are used to build up the "simple" cells of oriented line detectors (another class of feature detectors). This conceptualization has both advantages and disadvantages, which will be compared with those of Grossberg's model in a moment.

Grossberg's system could be said to have "center-surround" feature detectors based on the choice of the  $D_{ki}$  and  $E_{ki}$  as in

Figure 11. But subsequent detection of features such as a bar is due to the behavior of the field as a whole; it is not channeled through a single "higher" feature analyzer. Grossberg correctly observes that in such a system, "features" are not the appropriate unit of analysis. The behavior of the model is inherently too global for that.

What are the consequences of these distinctive approaches? In the case of Grossberg's field, there is the immediate potential for explaining many facts about perception that are inextricably global. "Filling-in" phenomena are, perhaps, the most intriguing, i.e., the filling-in of depth planes discussed by Grossberg here, but also pattern completion and illusory motion effects discussed elsewhere. Other global phenomena include lightness, complex random dot stereograms, and sensitivity to many cycles of a grating pattern. While it may be possible in yet higher stages of a "feature detector" model to incorporate such global effects, Grossberg is attacking them directly at a more immediate level.

On the other hand, some appealing aspects of the "feature detector" view are lost. For instance, in adaptation to a spatial frequency grating, the fact of its one-dimensional character is a priori arbitrary, but shows up in the cortical physiology. Thus, while one might not have predicted frequency-specific adaptation on a priori grounds, the knowledge of fatigue in neurons *plus* the finding of simple and complex receptive field cells make a nice, concise story. It is the studies of orientation-specific frequency adaptation more than any other single psychological experiment that suggest that grating patterns are tapping into something that is physiologically "real" in the system.

Note that grating patterns are special in terms of Grossberg's model as well, but for different reasons. They are not special because the linear orientation taps into a field of linear feature detectors repeatedly, but rather because the periodicity at a fixed frequency repeatedly excites one particular field of center-surround cells. Furthermore, when the long-distance interaction made possible by the "filling-in" mechanisms is added to the system, a kind of positive feedback based on the match between physical periodic stimulus and the underlying structural scale  $(D_{ki} \text{ and } E_{ki})$  can arise. This could lead to a global fatigue of that spatial scale system.

So, a grating (in contrast to a single bar) is a special stimulus for both systems, but for different reasons. For a field of feature detectors, it is because a grating is many copies of the right feature – the fact that it is periodic is something of a coincidence. (Note that some people would argue that higher feature detectors want more than a single bar – but the argument can still hold.) In Grossberg's model it is the spatial periodicity that is fundamental.

It seems to me, assuming the above interpretation of the models is correct, that this distinction has considerable predictive import. For example, radially periodic (circular annuli) stimuli might resonate with Grossberg's structural scales, but would not consistently excite a given class (frequency and orientation) of linear feature detectors. Checkerboard patterns (or checkerboard arrays of variously oriented gratings) might also distinguish the two.

A different sort of question, primarily quantitative, is raised by the "pattern matching" properties of the purely feedforward system (Grossberg's Equation 1, Section 21). Consider the sum of two patterns that have no spatial overlap . . . that is, each pattern is zero where the other one is positive. Equation (1) says that, in all cases, the peak response to the sum of such stimuli will be less than the maximum peak response for either of the individual stimuli presented alone, to wit, "the sum of mismatched input patterns from two separate input sources will be inhibited by network (1). This is because the mismatched peaks and troughs of the two input patterns will add to yield an *almost uniform* total input pattern, which will be quenched by the noise suppression property." [Section 22, emphasis added.] What is meant, quantitatively, by "almost uniform"? Consider,

for concreteness, a square wave grating of frequency f, as the first stimulus. If the second stimulus were another square wave grating of frequency f, but of opposite phase, the sum of the two would be exactly uniform, and Grossberg's network (1) would attenuate such a stimulus drastically, as desired. But now make the second stimulus a rectangle grating of frequency f, but with a duty cycle much less than one-half. If this stimulus is added in a phase such that the narrow "on" regions are centered in the broad "off" regions of the first stimulus, how much should the sum be attenuated? A little attenuation by total energy normalization is probably okay, but irrelevant. The key issue is, When is a sum of two patterns "almost uniform" enough to be quenched to zero by noise suppression? This quantitative question is crucial in Grossberg's subsequent model of binocular cooperativity (Figure 18) because the first stage is a strictly feedforward network. This stage is subsequently used to control rivalry or positive feedback for fusion. In short, the quantitative details of pattern matching for interesting stimuli are crucial for the success of Grossberg's approach.

Note that more elaborate systems (e.g., the feedback network of Equation 27) allow for (realtime?) modulation of the "quenching threshold." It seems to me that this feature could be helpful in explaining the subjective experience of unconscious searching to find a stereoscopic match, e.g., in looking at complex random dot stereograms, where, at first, nothing but a confusion is seen, but then slowly, often over a period of 20–40 seconds, a region of stereoscopic fusion appears and then spreads to the rest of the field.

Finally, let me mention a few questions about the quantized geometry of responses. Grossberg argues that one of the problems of Sperling-Dev models is that it is wrong to quantize the activity of sheets of depth planes in an all-or-none manner. If such were true, we would see discrete jumps in depth. Yet Grossberg quantizes, through FIRE-created periodicities, responses that seem to correspond to a variety of phenomena (length, lightness) whose quantal jumps we would surely notice. Grossberg addresses this issue to some extent, but I'll remain equally skeptical of either kind of underlying quantization until it can be demonstrated visually. Grossberg suggests ways to look for such things, although warning that the effects will be weak.

Viewing the issues of pattern matching, quantization, and global feature analysis as a whole, in the specific context of vision and stereopsis, I suppose my main concern is more an intuitive one than a principled theoretical one. I worry that there is a mismatch between the very robust, continuously variable subjective impressions of visual sensations and what appear to be rather triggerlike quantum transformations, some based on thresholds or quenching, in Grossberg's model here. Quantitative correlation to experimental studies is one way to address this concern. Fortunately, in Grossberg's present paper, much of the modeling is closer to testable psychophysics than is presently the case in some of his other work. Indeed, Grossberg suggests several tests himself, and I have perhaps indicated a few more. I hope this will lead to detailed, quantitative connections between constructs in the model and the variety of lightness, depth, and size effects that Grossberg discusses, as part of his ongoing work.

In closing, let me add that, as usual, studying a paper by Grossberg brings unique rewards. He judiciously selects a small number out of the myriad of psychological effects available and pushes these, theoretically, as far as he can. His data summaries focus attention on theoretically relevant portions of findings. This alone often illuminates a muddled field of investigation. Furthermore, it has been my experience, even in cases when I would summarize the data differently, or when I would choose different data to study, and even when I'm not convinced by a particular aspect of the model, that following Grossberg's analysis is invariably good exercise for any modeler's brain. It has often supplied me with conceptual tools for thinking about brain models that I have seen nowhere else.

## False dilemmas: Confusion between mechanism and computation

### Kent A Stevens

### Department of Computer and Information Science, University of Oregon, Eugene, Ore. 97403

Grossberg motivates his theoretical stance largely by posing "dilemmas" that certain alternative theories cannot explain, but that his can. The dilemmas, however, derive mostly from misunderstandings or misrepresentations of those alternative theories. My comments on this will lead into a more substantial issue: the restricted explanatory power of Grossberg's theoretical ideas.

First consider Grossberg's indictment of the edge detection theory of Marr and Hildreth (1980): "Because their theory is linear, it cannot tell us how to estimate the lightness of objects, and because their theory is feedforward, it cannot say how apparent depth can influence the apparent size and lightness patterns of monocular patterns." Grossberg sees the inability of their theory to account for lightness and depth phenomena as an insurmountable dilemma. For there to be a dilemma, of course, these phenomena must be integral to edge detection, therefore requiring an explanation by any proposed theory of edge detection. But Marr and Hildreth believe otherwise; their theory is cast within a computational framework (Marr 1976, 1977) which assumes a sense of modularity in visual processing (discussed momentarily). Hence we find Grossberg criticizing a theory for not accounting for phenomena that its authors regard as beyond its scope.

In much the same vein, Grossberg claims that the Marr and Poggio (1979) stereopsis theory is similarly flawed. This theory is condemned for its inability to account for: effects of perceived depth on perceived size and lightness (e.g. Emmert's law, Gogel's equidistance tendency, the Gilchrist effect), surface interpolation (the "filling in" of a smooth surface between sparse loci where stereo disparity is known), the perception of planarity and rigidity, and other spatial phenomena. Most of the dilemmas hinge on Grossberg's incorrect claim that Marr and Poggio's (1979) theory of human stereopsis is an attempt to account for depth perception from binocular vision; in fact it concerns the computation of stereo disparity. Curiously, while Grossberg points out the fallacies that come from confusing disparity with distance (Section 14), he then cites Marr and Poggio and others for elucidating the distinction between disparity and distance (Section 18).

Grossberg, referring to Marr and Poggio (1979), concludes that "all global filling-in and figure ground effects hereby become inexplicable in their theory." But again, those effects are beyond the scope of a theory of disparity computation. Moreover, it is not as if Marr and Poggio have no theoretical account for filling-in (surface interpolation). Marr [1978] provides a concrete computational framework for the representation of visible surfaces and for the processes that integrate 3-D information from different sources (e.g., stereopsis, motion, shading, and texture gradients). In the proposed representation, both distance and surface orientation are made explicit (with interpolation across smooth regions) and discontinuities in either distance or surface orientation are also explicitly marked. Phenomena such as Emmert's and Gogel's seem more appropriately attributed to processes operating on such a 3-D representation than to stereopsis. Grossberg should also study the computational theory of surface interpolation that Grimson (1981) has developed and implemented in conjunction with the Marr-Poggio stereo matching algorithm.

Let us turn now to the modularity hypothesis. Marr, along with many others, expects that distinctly different information processing problems (e.g., edge detection, stereopsis, lightness, motion detection, surface color) have essentially independent solutions (Marr 1976, 1977) and for that reason, are amenable to separate study, at least initially. Of course that hypothesis might be wrong – the modularity we see might be wishful thinking – in which case to understand any one aspect of vision seemingly would require understanding every other aspect. But then, the success of recent computational theories is very promising and suggests otherwise.

It should go without saying that the principle of modularity is rather subtle. For instance, it does not imply strict physical separability at the level of neural mechanisms. Roughly, it means that independent visual information, such as color and motion, are treated as independent by the visual system, despite the fact that the percepts derive from neural processes that share many common mechanisms and pathways. Some interaction phenomena, such as color-contingent motion aftereffects. are attributed to limitations of the neural mechanisms to process-independent types of visual information completely independently. Other interactions, such as between apparent size and distance (Emmert's law), or between distance and proximity (Gogel's equidistance tendency), or between lightness and surface orientation (Gilchrist) are functional, and reveal the visual system's ability to reason about space, illumination, and reflectance.

Grossberg does not seem to make any such distinctions, nor does he differentiate between a theory of information processing and a theory of the neural mechanisms that implement a process. What one finds, instead, is an explanation of visual phenomena directly in terms of feedback networks of neurons.

Vision certainly involves feedback of various sorts, such as light and dark adaptation, the control of vergence movements during stereopsis, and the influence of object recognition on perceived figure-ground. Where the information being fed back is simple (e.g., may be characterized as a single continuous variable, as might be the case with light adaptation), an explanation of the feedback mechanism might suffice. However, as Marr and Poggio (1977) eloquently argue, complex information processing requires satisfactory descriptions at several levels, of which a mechanism description is but one. They distinguish the computational theory (What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?), the level of representation and algorithm (What is the representation for the input and output, and what is the algorithm for the transformation?) and the level of implementation (How can the representation and algorithm be realized physically?).

Grossberg's descriptions of visual computations, unfortunately, are primarily at the level of mechanism, of patterns of neural activity within networks. There is no notion, for instance, of symbolic information processing. For those of us interested in understanding vision, the real problems seem to lie here.

## Adaptive resonance theory: Problems with prediction

### Mark Wagner

## Psychology Department, York University, Downsview, Ontario, Canada M3J 1P3

Grossberg's theory is quite impressive. Only someone with a broad theoretic vision, a keen mind, and a large amount of intellectual courage could produce a theory of such breadth and sophistication. In particular, Grossberg's attempt to relate phenomena at higher levels of scale (such as subjective contours) to phenomena at lower levels of scale (such as receptive fields) is laudable. Some of the greatest recent advances in physics involve a renormalization process (somewhat like Grossberg's) in which lower-level phenomena are summarized to yield phenomena at higher levels of scale. Psychobiology (whose object of study is the complex pattern of neuronal firing as it relates to

higher-order perceptions) is a natural place to apply such renormalization procedures.

Several objections need to be met before adaptive resonance theory can be accepted as *the* theory of visual perception, however. First of all, I do not feel that Grossberg's theory can truly be said to unify all of the many visual models mentioned in the introduction into an integrated visual theory, as the author suggests. Many of these theories are designed to address aspects of visual perception to which Grossberg's theory is inapplicable. For instance, I doubt that the precise localization of points in visual space entailed by the hyperbolic geometry of Blank (1978) and Luneburg (1947) can be derived from Grossberg's formulation. Adaptive resonance theory may possess wide application, but it is not all-inclusive.

Second, the form of the theory is perhaps too general. The theory could "predict" any number of alternate sets of empirical findings. For example, the model could predict either noise suppression or noise enhancement. Suitable parameter values are chosen after the fact to produce noise suppression. The model could predict either signal noise suppression or signal noise enhancement. A specific form of the positive feedback function f(w) is chosen to produce signal noise suppression. The same pattern occurs throughout the paper. As the mathematics grows to fit new phenomena, the overly general model is constrained (somewhat arbitrarily) to conform to what is already known. This flaw is not fatal provided that once all constraints are specified, the model can still make further predictions of empirically testable phenomena.

Third, in an attempt to predict such additional phenomena, Grossberg sometimes incorrectly reinterprets the meaning of certain perceptual laws in order to make them fit the model. Two examples may clarify this point.

a) The derivation of Weber's law (found in Section 23) does not really produce Weber's law. As the problem is set in Equation 10, the total input I is equivalent to uniform background luminance and  $\Delta I$  is a particular stimulus added to this field. The 'Weber's law" derived is thus the absolute threshold for stimulus detection above background luminance. Weber's law, as it is generally used, refers to the amount of additional stimulation needed to detect a just noticeable difference in a suprathreshold stimulus. That is, instead of saying that  $I_1 = K + \Delta I$  (the background luminance plus some additional luminance at a particular point), the equation should read  $I_1 = K + i + \Delta I$ (where *i* is a suprathreshold stimulus impinging on  $v_1$ ). The total input before  $\Delta I$  is applied to  $I_1$  is I = nK + i. Following essentially the same argument as Grossberg, I find that (i + i) $\Delta I$ /( $I + \Delta I$ ) = constant, which is not at all like Weber's law. (Essentially, I just substituted  $i + \Delta I$  for  $\Delta I$  in the equations). An alternate interpretation of Weber's law would be to say that I as a whole is the stimulus. In this case  $\Delta I$  should increment I as a whole (i.e.,  $I_1 = I_2 = \ldots = I_n = K + \Delta I$ ). This formulation will not produce Weber's law either.

b) The derivation of hysteresis (found in Section 34) is also not correct. Hysteresis commonly refers to the tendency for a perception to persist once established. A stimulus which normally would not produce a certain percept if presented immediately produces the percept when the stimulus is gradually approached from another stimulus which does produce the percept. In Grossberg's theory, a large number of physical lengths will produce the same perceived size. Any stimulus from length L to length  $L + w \Delta L$  will produce  $M_L$  disinhibitory peaks and thus the same size estimate. Whether  $L + w \Delta L$  is approached gradually from L or is presented directly does not matter. The same number of peaks will be produced, and thus the same size estimate will be made. If anything, Grossberg's formulation argues against hysteresis.

Finally, Grossberg's theory makes several startling predictions for phenomena which have not been extensively investigated (e.g., ghost images, luminance-depth interactions, a lag between filling-in and the construction of contours, etc.). The ultimate acceptability of Grossberg's theory rests on whether or not these and other predictions can be born out in subsequent empirical investigation.

## Author's Response

## Interdisciplinary aspects of perceptual dynamics

Stephen Grossberg

Center for Adaptive Systems, Department of Mathematics, Boston University, Boston, Mass. 02215

Before replying to the interesting specific issues raised by the commentators, I will set the stage with some general remarks.

1. Fourier analysis as a perceptual representation. My theory attempts to provide a unified framework for physically understanding at least three types of data and theory that were heretofore separate: Fourier theory of pattern perception, edge and local feature extraction theories, and cooperative theories of depth perception.

The Fourier theory is of special importance for understanding my target article. For simplicity, consider a function f(x) of a single spatial variable x. Its Fourier representation

$$f(x) = \int_0^\infty [a(y)\cos(xy) + b(y)\sin(xy)]dy \qquad (1)$$

expresses f(x) as a sum of cosines  $\cos(xy)$  and sines  $\sin(xy)$  of different frequencies y weighted by frequencydependent amplitudes

$$a(y) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(z) \cos(yz) dz$$
 (2)

and

$$b(y) = \frac{1}{\pi} \int_{-\infty} f(z) \cos(yz) dz$$
 (3)

Such a decomposition shows how a function f(x) whose mass is localized in a small region of x-space can be represented as a superposition of spatially periodic functions with different spatial frequencies and amplitudes. The spatially periodic functions oscillate at values of x that can extend far beyond the region of x-space where  $f(x) \neq$ 0. The many successes of the Fourier approach is generating and interpreting perceptual experiments are by now well known.

Difficulties of interpretation arise when one views the Fourier approach as a physical theory wherein a Fourieranalyzed input pattern is represented by periodic perturbations within a perceptual space. Such a representation is implicit in the manner that Fourier theory is used to analyze an input scene before resynthesizing the scene into a percept from suitably weighted spatially periodic functions. Such a representation is also explicitly suggested by various data, such as the forward masking experiments of Weisstein et al. (1977). In these experiments, dots or oriented bars are used to mask later bullseye or oriented grating patterns. Even a dot that is  $0.17^{\circ}$ in diameter can mask a complete 5° bulls-eye, not just the bulls-eye center. In another version of this experiment, a vertical bar mask is followed with a 1.5° circular patch of vertical grating whose position with respect to the bar is varied. Pattern-specific masking occurs at spatial separations as great as 4.25°. Thus the grating is masked even when no spatial overlap occurs between the mask and the target. Weisstein and Harris (1980) provide an elegant review of these and related results.

Such findings raise various basic issues. If a small spot can influence a region 5° or larger in a frequency-specific way then why don't we see this influence of the spot while it is on? Why does the spot's surround look spatially homogeneous? The Fourier theory provides a simple rationale: The spot selectively desensitizes Fourier components in the surround, but the sum of these components is spatially uniform. When a bulls-eye later probes some of these desensitized components, its threshold contrast is raised.

Given these appealing formal conclusions, one needs to ask how frequency-specific, spatially periodic reactions to an aperiodic input pattern are physically generated. One approach is to look for neurons whose receptive fields have many periodic sidebands that are sensitive to at least 5° of visual space. Where such cells are synthesized entirely by bottom-up or top-down influences (Weisstein 1980), I find it hard to reconcile their broad spatial scale with the existence of sharp perceptual boundaries. Instead, I have suggested that matched informative visual features can trigger multiple-scale filling-in reactions that are spatially contained by filling-in barriers. This approach does not require that individual receptive fields exist with 5° multiple periodic sidebands, and it shows how a broad region can be filled-in that has sharp perceptual boundaries. Because the matching processes underlying filling-in include binocular edge matches among multiple structural scales, relationships between perceived depth and the existence of phantom contours (Weisstein & Maguire 1978) and between the perception of an illusory grating and its ability to mask a real grating (Weisstein, Matthews & Berbaum, 1974) become qualitatively more understandable.

Some commentators asked why my filling-in patterns are not more easily visible (Arend, Foley, Savoy). One answer is: For the same reason that you cannot easily see the periodic components in the Fourier theory. In this respect the two theories, as representations of perceptual space, stand or fall together.

Other aspects of the two theories differ markedly. Instead of sines and cosines in (2) and (3), the kernel that I use derives from equation (23) in Part II. Equation (23) correlates the (preprocessed) input pattern with a weighted difference-of-Gaussian kernel. Changing the structural scales in (23) changes the amplitude of the resultant "edge." Such edge estimates generate a functional scale via a FIRE based upon the same structural scales that determine the edge amplitude. A relationship between the spatial frequency and the amplitude of a functional scale thus obtains because both are mediated by the same structural scales. Analogously, the relationship between the spatial frequency y in equation (1) and the amplitude a(y) in equation (2) is mediated by the same function  $\cos(y\cdot)$ . Spatial periodicity, however, is no longer a structural property. It is a coherent dynamic property that is initiated by edge-matches.

2. Evaluating a real-time perceptual theory. Several commentators made remarks that raise questions concerning the proper approach to evaluating any real-time perceptual processing theory, not only my own. Because such theories are not common in psychology, I will summarize the framework in which my theory has been growing before turning to specific issues.

My theory was born in 1958 out of an effort to understand data about human and animal learning. Every stage of the theory has been derived from psychological postulates concerning the adaptive self-organization of individual behavior in response to environmental pressures (Grossberg 1982e). The entire theory thus rests upon a psychological foundation. The fact that its computational structures always look like neural networks says something important about the relevance of adaptive constraints on individual behavior to the evolutionary design of neural mechanisms.

By analyzing general environmental pressures to which a behaving organism needs to adapt, the theory has identified a small number of general psychological principles that have been used to derive its formal dynamical laws. The generality of these environmental pressures helps to explain why the same dynamical laws have been helpful in explaining data about many different types of behavior. A parsimonious synthesis of a huge data base is already published in the literature. This synthesis has not been readily assimilated by most investigators because of its interdisciplinary use of psychological, physiological, and mathematical tools.

Through the analysis of adaptive interactions between a behaving organism and specialized properties of its environment, special-purpose processors capable of supporting these interactions have been derived. The *minimal* network is sought, both to express an atrophy-due-todisuse hypothesis and to justify every computational detail of the processor. Using this "method of minimal anatomies," a network is expanded only when it becomes clear how and why it fails to computationally support an important environmental constraint. Then the minimal network is embedded into a new network that is minimal with respect to a richer environment. This procedure suggested the name *embedding field theory* for this approach in 1959–60. I still like the name.

From a neural viewpoint, the existence of a small number of dynamical laws corresponds to electrical and chemical properties shared across neural subsystems. The existence of special-purpose networks corresponds to the functional heterogeneity of neural subsystems, including the proper juxtaposition of anatomical connections, electrical signaling rules, chemical transmitters, and the like. Obviously it is far more difficult to synthesize an explicit minimal network for a complex body of psychological data than it is to identify the qualitative processes that must enter such a network. Thus in testing a processing theory, it is crucial to distinguish whether data support or contradict aspects of a particular minimal network – notably a qualitative property computed by the network – or the dynamical laws themselves.

From this perspective, Part I of the target article summarizes some of my reasons for believing that a new

qualitative property was needed to deal with certain perceptual data. Part II describes a new minimal network wherein this property is computed using known dynamical laws that have elsewhere been derived behaviorally. Part II also indicates by its remarks about hypercolumns and dipole fields (Section 43) that this minimal network is known to be too minimal to compute all the perceptual phenomena related to depth, lightness, and form that one might desire. The minimal network is necessary, I claim, but not sufficient to achieve this goal. Nonetheless, the explicit meaning and design of the network and the restricted manner in which it can be embedded in known dipole field structure creates a strong teleological pressure for expanding the network when it is confronted by more perceptual data. Because of this fact, the filling-in notion that has emerged within my framework has already opened new paths to rapid progress, as the next section of this response will illustrate.

3. Brightness perception, filling-in barriers, and stabilized images. Issues about brightness perception are raised by Arend, Coren, and Laming. I will suggest an explanation of their data that illustrates how a properly designed minimal network exerts constraints on its own embedding into a richer computational structure.

Figure 18 of Part II of the target article says little about how the input pattern is preprocessed. The same arguments that suggest frequency-specific edges as a basis for binocular matching also suggest that the edges corresponding to each structural scale parse the monocular input patterns into frequency-specific monocular brightness domains. Figure 1 of the Response depicts such a network. In Figure 1a, each monocular input pattern bifurcates into a specific branch (edge processor) and a nonspecific branch (brightness processor). Such a bifurcation is a general design feature of my theory (Grossberg 1970a, 1971a, 1972a, 1972b, 1975, 1978e, 1980b). The offedges send topographically organized inhibitory signals from the edge processor to the brightness processor. These signals generate monocular barriers to the lateral diffusion of input activity that occurs in the brightness processor. The inhibitory signals accomplish this by acting as inhibitory shunts of the boundaries that exist between adjacent cellular compartments (Figure 1b). Both the input pattern and the edges send inputs to the compartments. Michael Cohen and I (Cohen & Grossberg 1983b) are now using these ideas, notably the concept that off-edges are inhibitory inputs as well as boundary signals, to explain brightness data of such workers as Bergström (1966, 1967a, 1967b, 1973), Hamada (1980), Levelt (1965), Sakata (1981), and Von Békésy (1968). The brightness processor in Response Figure 1a interacts with the binocular filling-in mechanism just as in Part II of the target article.

Gerrits and his colleagues have summarized clinical and experimental evidence that support this expanded interpretation of the interaction between edges, brightness, and filling-in. Gerrits and Timmerman (1969) found that when a visual stimulus entirely surrounds a retinal scotoma, patients report that filling-in occurs very rapidly, and that offset of the stimulus rapidly elicits a perception of darkness. By contrast, Gerrits et al. (1966) found that when a black disk (or an artificial scotoma) was fixed on the eyeball, it took several seconds for a sur-



Figure 1. Modulation of perceived brightness by edges and resonant matching. (a) A monocular input pattern sends signals both to a brightness processor and to an edge processor. Frequency-specific off-edges define the boundaries of brightness domains which laterally average the input pattern and the edge input within each domain. On-edges that match in the binocular matching field can lift these brightness estimates into the perceptual domain via a FIRE. (b) A blow-up of the compartments comprising a brightness processor. Off-edge inputs determine whether activity can flow across the boundaries of a cellular compartment to form a single brightness domain from a series of such compartments.

rounding stimulus to fill-in the disk. When the light was switched off, the field outside the patch immediately looked dark but it took several seconds for the filled-in percept to vanish. Gerrits and Vendrick (1970b) analyzed the difference between the filling-in times of real and artificial scotomas in terms of the nonexistence of edge barriers to filling-in around a real scotoma. The long time needed for the filled-in percept to vanish when it is protected from stimulus reset by bounding edges suggests that a feedback process maintains the filling-in percept. If a feedforward process with a slow decay rate maintained the filled-in percept, then the percept should also decay slowly when the stimulus is shut off around a real scotoma. Instead the percept rapidly vanishes. [See also Campion et al: "Is Blindsight an Effect of Scattered Light, Spared Cortex and Near-Threshold Vision" BBS 6(3) 1983.]

Gerrits and Vendrick (1970b) also point out that in a stabilized image experiment contours disappear before brightness fades, and that spreading of brightness or darkness takes place as soon as the contours disappear. In the light of this familiar result, I am not sure how to interpret Foster's claim that he knows no evidence indicating that "the level of activity generated for a spatially uniform and therefore edgeless field should be much less than that for a large but bounded field."

The rules depicted by Figure 1 (Response) embody a "domain-of-integration" effect such that "the integral must not extend over the edges," and there are "limits to the domain of interactions even within the same depth plane," to use Arend's expressions. The perceived brightnesses of the luminance profiles in Arend's figure have a simple qualitative explanation in this framework. For example, in the multiple scallop of his figure (f), the offedges create boundaries to filling-in and lower the average activation level within their brightness domain. Hence a series of increasing steps is generated with rather sharp boundaries. Coren's interesting figure involves both changes of spatial scale and number of peaks between his luminance profiles (a) and (c). I am not sure from his description of the patterns whether or not the fact that the five-peaked pattern is averaged by a different set of brightness domains than the two-peaked pattern is sufficient to explain his data. In our brightness theory, a change of scale accompanied by an increase in the number of peaks does flatten the percept. Laming desires that "the increment be first stripped from the background." This demand is met in (Response) Figure 1. Despite Laming's remark that "this, I insist, is a sufficient explanation of the phenomenon," his model does not meet any of the data of Gerrits and his colleagues, nor, it would seem, does it fit most of the data of Arend and Coren.

The reader who values parsimony might now ask: Why do you need edge-driven resonant filling-in if you already have edge-bounded brightness domains? One answer is the following. A binocular mismatch of edges may suppress an edge and thereby prevent it from bounding a monocular brightness domain, but it does not suppress the activity of the brightness domain that is initiated by the input pattern. To achieve binocular selection of spatial scales and its stepchild, binocular rivalry, one still needs to prevent the perception of activities that represent mismatched data.

4. Dipole fields: Adaptation level, habituation, normalization, adaptive filtering, and the McCullough effect. The issue that Weber's law can be explained by an intracellular chemical process whereby light-energy is transformed into electrical activity is raised by Buffart. Wagner claims that I derived a Weber law under the wrong physical hypotheses. Freeman discusses a "surge of activity that long outlasts the transient [input] and attenuates the transmission of receptor input to the bulb. The agent of attenuation . . . probably involves the release, accumulation, and slow clearance of a substance such as potassium . . . the degree of attenuation depends on the induced transneuronal activity." Freeman also makes other interesting remarks about his measurements of normalization properties, shunting interactions, supraquenching-threshold switching, and standing wave FIRES. Dalenoort mentions the relevance of synaptic modification to learned changes in interaction strengths. Levine and Savoy remark about the reality of feature detectors. Julesz discusses cooperative and hysteretic effects of spring-coupled-magnetic-dipoles. Indow mentions the Hurvich and Jameson (1955) opponent process model for color vision. Bridgman discusses how difficult

it is to explain the McCullough effect because of its persistence and feature specificity.

Dipole fields have all of these properties. None of the other approaches listed by these commentators, except possibly Freeman's, can make the same claim. Dipole fields thus constitute a processing design wherein processing insights of many other theories can be rationalized and generalized. This conclusion does not deny the utility of other approaches. Rather it suggests how dipole field properties can unify specialized models and can clarify processes that are lumped together or not present in these models. Figure 2 (Response) depicts a dipole field that was used to help explain data about reinforcement and motivation (Grossberg, 1982c, Figure 12). I include this figure to emphasize that the processing principles embodied in the dipole field design extend beyond visual phenomena. These principles are concerned with the stable self-organization of developing codes (Grossberg 1980b). Since the dipole field concept has been in the literature for a decade (Grossberg 1972b), my remarks about Figure 2 that are aimed at clarifying commentators' assertions will be brief.

The chemical gates in pathways  $4 \rightarrow 6$  and  $5 \rightarrow 7$  have the type of Weber law properties that **Buffart** desires. These properties led Gail Carpenter and myself to use a gated dipole model to fit parametric data quantitatively about the transduction of light into electrical signals within vertebrate cones (Carpenter & Grossberg 1981,



Figure 2. A motivational dipole field. This figure illustrates that the same types of concepts and mechanisms that arise in a motivational dipole field are also implicated by the perceptual data and properties mentioned by several commentators. (From Grossberg 1982c)

1983). In Figure 2, the gates are part of positive feedback loops  $4 \rightarrow 6 \rightarrow 8 \rightarrow 4$  and  $5 \rightarrow 7 \rightarrow 9 \rightarrow 5$  that can maintain an STM (short-term memory) response after a brief input terminates. This response habituates due to an imbalance between gate accumulation and depletion that arises due to the activity of the loops, as **Freeman** would desire. The intracellular gating actions and the intercellular inhibitory interactions 13 both contribute to the adaptational and switching properties of the system.

The pathways 11 and 12 contain LTM (long-term memory) traces. The LTM trace in 11 is conditioned when sampling signals in 11 are temporally contiguous with an active STM reverberation at 8. The LTM trace computes a time average of presynaptic signal and postsynaptic STM activity (Grossberg 1968). Since persistent pairing of 11 and 8 can cause conditioned decreases as well as increases in synaptic strength (Grossberg 1969a, 1971b, 1972c), the LTM trace is not Hebbian, contrary to Dalenoort's claim. When several conditionable pathways (such as 11) converge upon individual nodes (such as 8), they form an adaptive filter. When such an adaptive filter is modulated by a postsynaptic competitive process, feature detector tuning curves emerge (Grossberg 1976a, 1976b). Essentially all feature development models in the literature use a combination of adaptive filtering, competition, and normalization. Cooperativity and hysteresis in a dipole field are due to the manner in which positive feedback (as in  $8 \rightarrow 4 \rightarrow 6 \rightarrow 8$ ) generates inertia against lateral inhibition (as in  $10 \rightarrow 8$ ).

Each gated dipole (e.g., nodes 4, 5, 6, 7, 8, 9 and connections) realizes an opponent process. In the case of a red-green dipole, pattern-contingent long-term aftereffects can occur. In Grossberg (1980b), I suggested that persistence of the McCullough effect can be partly attributed to an interaction between gate habituation and contrast enhancement in STM. The imbalance due to asymmetric habituation can also be encoded in the LTM traces due to the property that a decrease in postsynaptic STM activity can cause a decrease in LTM strength. Such an LTM change is not a case of associative learning due to CS-UCS (conditioned stimulus, unconditioned stimulus) contiguity. Rather it is a case of retuning feature detectors to their trigger input patterns. Thus habituation within a nonclassical opponent process interacting with a non-Hebbian adaptive filter can explain McCullough effect properties as manifestations of a basic processing design.

Wagner's claim that a Weber law was incorrectly derived in Section 23, Part II, seems to be based on an insufficient understanding of the mathematical formalism. My derivation is carried out for the standard case of an increment threshold on a background (Barlow 1972, p. 136; Boynton 1968, p. 22; LeGrand 1957, p. 255).

Because of the broad scope of my target article, the commentaries have also touched upon many topics. Each commentary has its own conceptual center of mass. I will therefore discuss individual commentaries in the remaining pages.

5. Bridgman: Multistability, subtractive inhibition, shortterm memory, and Uttalism. Many of Bridgman's general remarks are explicitly or implicitly discussed in the first four sections of my Response. Herein I will discuss some of his specific remarks to clarify distinctions that are important for a full understanding of my theory.

"Quantum" does not equal "multistability." Multistable spatial patterns can be stored in STM (Grossberg 1973) even if they are not quantized standing waves. Automatic tuning of sensitivity does not occur in a subtractive feedback model of lateral inhibition, if by automatic tuning one means automatic gain control. Automatic gain control requires multiplication, as between x, and  $f(x_k)$  in equation (27) of Part II (target article). Short persistence of activity is not the rule in my competitive feedback networks, unless they are actively reset by gate habituation, antagonistic rebound, or an inhibitory input burst. The positive feedback terms, such as  $\sum_{k=1}^{n} f(x_k)D_{ki}$  in equation (27) of Part II can maintain the STM reverberation. Bridgman seems to have in mind networks in which all the interactions are inhibitory. My use of the term STM fits in well with a variety of data about sensory, cognitive, and motivational information processing (Grossberg 1978e, 1980b, 1982b, 1982d, 1983a).

Bridgman suggests that subtractive inhibition "is more common in the cortex." Results of workers like Freeman (1979b) and Rall (1977) suggest the importance of shunting dynamics, as does the ubiquity of membrane processes in nerve cells. This observation does not deny that cellular tissues may maintain themselves within a sensitive operating range. However, invoking additivity from the start may prevent one from understanding the designs by which this crucial property is achieved. All the circuitry, both intracellular and intercellular, that goes into these designs would seem inexplicable if additive dynamics were erroneously invoked. This observation is a simple version of the general fact that different assumptions about local processing can generate qualitatively different global theories about the same class of behavioral phenomena.

Bridgman criticizes my approach for its "Uttalism." Ironically, my approach is one that is unusually undeserving of such a criticism. All of my theorems characterize the dynamics of arbitrary numbers of cells starting out in arbitrary initial states and responding to wide classes of inputs. Such global theorems about nonlinear n-dimensional systems are rare not only in psychological and neural modelling, but also in the mathematical literature about dynamical systems. The struggle to prove global theorems of this type has always forced design insights that went beyond the lessons of available computer simulations, and I have repeatedly used these design insights to propel my physical theory forward. For example, instabilities during the development of cognitive codes that were not reported in other authors' computer simulations using small numbers of units were clearly visible to my mathematical analysis. These insights led to my adaptive resonance theory (Grossberg 1976a, 1976b, 1980b).

My general mathematical results are usually complemented by working out important examples mathematically or numerically. I believe that one should try to keep both the forest and its trees simultaneously in view. This is true in my filling-in theory as well as in past theoretical work. For example, I demonstrated how to design associative networks in which arbitrary spatial patterns could be learned without bias by arbitrarily many, simultaneously active sampling populations that may be activated by arbitrary data preprocessing and may possess arbitrary LTM decay and associative sampling

rules (Grossberg 1969a, 1972c, 1982a). This general theorem is balanced by the study of such special topics as: self-organizing temporal processors, whose simplest instantiations look like vertebrate command anatomies and whose more complex versions help to explain aspects of continuous speech processing (Grossberg 1970b, 1978e, 1983a); and the parametric analysis of serial learning to provide new explanations and predictions about bowing, skewing, serial generalization gradients, inverted U properties, and the self-organization of planned behavior (Grossberg 1969b, 1974, 1978e, 1983a; Grossberg & Pepe 1970, 1971). Analogously, I proved a general result about how to design competitive networks so that their decision schemes would always lead to stable storage in STM even if these schemes are so complex that one cannot explicitly compute them (Grossberg 1978c, 1980a). This theorem is balanced by parametric studies of how changing each parameter of a competitive network alters its transformation from input pattern to stored pattern (Grossberg 1973, 1981; Ellias & Grossberg 1975; Grossberg & Levine 1975; Levine & Grossberg 1976). In the present instance, Michael Cohen and I proved a general stability result about STM in networks with arbitrary structural scales, decay rates, saturation levels, and input patterns (Cohen & Grossberg 1983c) before using this theorem to guide our simulations of filling-in (Cohen & Grossberg 1983a).

6. Buffart: Functional theory, brightness averaging, and permanent rivalry. I have offered a computational theory but not a functional theory, Buffart suggests. I do not agree. If one insists upon a real-time dynamical description of a perceptual process, then a functional theory of perception is one that describes the dynamical processes underlying perception, including how and why they respond to the input patterns that they receive. Otherwise expressed, knowing the deep structure of a functional theory is tantamount to knowing its dynamical process. This knowledge implies nothing about the hardware implementation of the process. The fact that a good real-time theory sometimes suggests a possible implementation is not a disadvantage of such a theory. Rather it is a mark of its physical correctness. More is more, not less

The dynamical laws of the present theory simultaneously express both functional and computational information. The interactions of the minimal network with prescribed input environments also express both functional and computational information. The existence of a FIRE in response to certain environments but not others is a global property of the minimal network's abstract information processing. Any attempt to separate the functional and computational aspects of the FIRE would trivialize both its meaning and its properties. To understand a concept such as FIRE, one must enter into the internal logic of its process, just as one must do to fully understand quantum mechanics or relativity theory.

I consider functional theories that have not reached an understanding of process to be less mature as physical theories than theories which have achieved such an understanding. Such a remark would be considered too obvious to make in a discussion of theories in physics or chemistry. Some people, perhaps due to the sheer difficulty of characterizing perceptual processes, seem to feel

that it is necessary to justify lesser, but nonetheless useful, efforts. If a theory does not penetrate to the processing level, such individuals give the theory a nice new name - functional or computational theory - and say that it is good to have distinct functional and computational theories. To me this just turns desperation into a virtue, and gives metaphorical theories too much credibility. In a balder example of the same social phenomenon, the early Skinnerians made the semireligious commandment: "Thou shalt not theorize, especially about intervening variables," because it was so hard to explain their data at that time. This attitude may have simplified the goals of committed Skinnerians, but it did not enhance the acceptance or vitality of Skinnerian psychology within the greater scientific community. [See forthcoming BBS special issue on the work of B. F. Skinner and special issue on Foundations of Cognitive Science, BBS 3(1) 1980.]

I agree with many of Buffart's descriptions of data and assume that he has offered them as interesting results that are dealt with by his important theory, rather than as challenges to my theory, since my theory also helps to explain them. I do not, however, agree with his assertion that brightness summation does not occur. Brightness averaging, brightness summation and rivalry can all occur, and for good reasons, in a binocular dipole field. In such a field, the critical insights about brightness averaging versus summation are: uninhibited monocular edges only fill-in their own monocular lightness domain; inhibited or nonexistent monocular edges cannot lift their monocular lightness domain into resonant perception; and matched monocular edges can fill-in using both of their monocular lightness domains. The gating process underlying rivalry is active even in fusion. This gating process prepares the percept to be rapidly reset by stimulus offset leading to antagonistic rebound.

7. Dalenoort: Activity control center, arousal, categorization. It is noted by Dalenoort correctly that long-range shunting inhibition is not the only intercellular normalization mechanism. An "activity control center" can also normalize total activity. He does not, however, point out that the two approaches join different dynamical rules to their different global anatomies. I will illustrate the connection between these mechanisms with a historical example.

In Grossberg (1970a), I classified minimal networks capable of extracting the reflectances in an input pattern. Although this problem was imposed upon me by computational issues in learning theory, not surprisingly, the resultant networks look strikingly like retinas, since they contain analogs of receptors, horizontal cells, bipolar cells, amacrine cells, and ganglion cells. During this classification, it became clear that these minimal networks need to normalize their total activity as a preprocessing stage in the reflectance computation. In order to instantiate this functional property computationally, I observed that in an additive theory, a nonspecific inhibitory interneuron does the job, whereas in a shunting theory, an on-center off-surround anatomy is suggested. Such a nonspecific inhibitory interneuron is the "activity control center" of Dalenoort and of Amari (1982).

I do not agree with Dalenoort that it is harder to realize an on-center off-surround anatomy than a nonspecific

inhibitory interneuron when the nodes  $v_i$  represent nonlocalized cell populations. Long-range shunting inhibition from every cell to all others can do the job. The choice, I believe, is constrained by deeper properties. In a retinal model of a reflectance processor, for example, shunting interactions at the bipolar cell level support sustained responses that obey the shift property (Section 23, Part II). Subtractive inhibitory interneurons can, by contrast, generate transient on- and off-reactions (Grossberg 1970a, 1976a).

Unlike Dalenoort, I do not equate the normalized total activity of a network with arousal. Instead, my theory distinguishes several types of arousal. One type is activated by unexpected events and thereupon resets STM via selective antagonistic rebounds. I locate this arousal source in the reticular formation or hippocampus and relate it to the N200 evoked potential. [See forthcoming BBS special issue on Event-Related Potentials and Cognition.] Another type of arousal acts like incentive motivation and biases the attentional focus toward motivationally compatible cues. I localize this arousal in the hippocampus and relate it to the CNV evoked potential. Other arousal sources modulate the quenching thresholds of sensory, cognitive, or motor representations, thereby transforming subliminal patterns into supraliminal percepts, concepts, or actions. The PGO (pontogeniculo-occipital) wave illustrates such an arousal source (Grossberg 1978e, 1980b, 1982d, 1983b).

Dalenoort doubts that filling-in on a single network level can explain categorization. So do I. In my theory, categorical and hysteretic properties of stable adaptive resonances are supported by coordinated bottom-up (filtering) and top-down (expectancy) processes in response to prescribed input environments (Grossberg 1978e, 1980b, 1983b).

8. Foley: Hypercolumns, fusion, and size. I agree with Foley that aspects of my theory need further development. I have preferred being explicit about where and why I feel that uncertainties exist, rather than overstating my case. Some of the uncertainties illustrate incomplete aspects of perceptual theory per se. Other uncertainties are due to limitations in our knowledge about how perceptual and motor processes interact. On the sensorymotor side, we are working out a detailed model of adaptive visuomotor interactions for a relatively simple. but still very demanding sensory-motor system (Grossberg & Kuperstein 1983) and are generalizing these results to more complex sensory-motor interactions. On the perceptual side, we are moving in several directions at once: toward a better theory of brightness processing; toward more realistic preprocessing of monocular patterns through retinal models; toward a generalization of our one-dimensional binocular theory into a two-dimensional theory (cortical hypercolumns); toward the specialization of general results on cognitive feedback to the binocular case. Despite these incomplete aspects of the theory, I do not agree that the theory "cannot be tested by perceptual experiments." As a simple example, equation (23) in Part II can be used in just the same way that difference-of-Gaussian models are now used (Wilson & Bergen 1979), and may explain data that go beyond these models, as in experiments wherein perceived suprathreshold contrast changes as a function of image luminance.

Foley writes that "induced gratings are generated in a direction parallel to the inducing grating, whereas in Grossberg's network . . . propagation . . . is at right angles to the inducing edge." This confusion is due to the one-dimensional nature of the simulations. In a twodimensional dipole field, horizontal and vertical orientation detectors are mutually inhibitory (Grossberg 1980b). Filling-in will therefore tend to occur among like orientations, as in the data. This qualitative property does not, however, solve the difficult problem of designing a hypercolumn organization that computes these coherent properties in a physically meaningful way.

Foley also writes that my use of matches in multiple structural scales as a basis for depth perception "is contradicted by the finding that the maximum disparity that will produce fusion or depth is independent of the luminance spatial frequency of the pattern (Pulliam 1981)." Pulliam's data are more complex than this remark suggests. Pulliam constructed pairs of vertical sine wave gratings each of whose vertical bands is horizontally displaced by an amount that varies sinusoidally from top to bottom. Thus two spatial frequencies are used to define each image. The images are grouped into pairs in which the sine wave displacements of the vertical bands are 180° out of phase in each image. Thus stereoscopic fusion of such an image pair produces a sinusoidal variation in disparity from top to bottom. A horizontal sinusoidal depth modulation is thus seen superimposed on a vertical luminance grating.

Pulliam (1981) found that "the channels which detect displacement modulation seem to be differentially tuned to luminance spatial frequency [the spacing of vertical sinusoids] as well as to displacement spatial frequency [the spacing of the wiggles within each vertical band]. Channels tuned to high displacement spatial frequencies tend to be tuned to high luminance frequencies, and channels tuned to low spatial frequencies tend to be tuned to low luminance spatial frequencies" (p. 73). The data summarized by Foley just indicate that given a fixed displacement spatial frequency, the maximum disparity that produces fusion is independent of luminance spatial frequency. Foley did not mention that "these functions decrease with increasing disparity frequency" (Pulliam 1981, p. 76). In other words, the maximum disparity that will produce fusion or depth decreases as the disparity frequency increases. This is what one expects from my theory.

The subtlety of the Pulliam experiments is due to their simultaneous manipulation of two spatial frequencies in orthogonal directions. If the maximum disparity that produces fusion at a fixed disparity frequency does not depend upon the orthogonal luminance spatial frequency, then why does the maximum displacement frequency at which depth can be seen increase with the maximum luminance spatial frequency? Pulliam (1981, p. 76) suggests an answer that is also compatible with my theory (Section 41, Part II) when he reviews results obtained with square wave luminance gratings with sine wave disparity modulations: "Evidently, the higher frequency components of the square wave can 'carry' the high frequency depth information, even though the [low frequency] fundamental component by itself would not.

As Foley notes, I have suggested several processes that may contribute to the determination of perceived size. Size may be influenced by the number of functional wavelengths in each of the active structural scales, or by the spatial extent of coherently active network regions, or by interactions of these properties with motor feedback mechanisms. None of these options implies "that the size-distance invariance principle . . . holds only roughly at best." A functional wavelength can be commensurate with the receptive field size of its network, and the activity spectrum across several structural scales is an even more sensitive representation than one across a single structural scale. Tests of length uncertainty due to the quantization of functional scales must be prepared to detect very small differences, notably hysteretic differences that may arise when a monocularly viewed object's length is slowly increased and then decreased.

9. Foster: Percepts, metrics, and tests. That a Cornsweet profile and a rectangle (target article, Figure 7, Part I,) become Cornsweet profiles after Fourier processing rather than the rectangles that are perceived is a fact defended by Foster. He suggests that only the identity of the outcomes, not their shapes, is important. This position is incompatible with other uses of the Fourier theory in which the outcome of a Fourier analysis does look correct and is therefore used to support the theory. Foster's position dissociates perception from any possible measurement of the collective activities of cellular feature detectors. Given the growing neural evidence for spatial frequency detectors in visual cortex (Pollen & Ronner 1981, 1982), I prefer to salvage the truth in Fourier theory by seeking a modified formalism, rather than internally weaken its intuitive foundation. The data reported by contributors like Arend also seem to require a more powerful theory.

Foster points out that "metrics may be used implicitly by the nervous system" to compute pattern matches at successive stages. I do not doubt this in the least. Rather, I am suggesting how and why matching operations are explicitly computed from the perspective of a deeper computational viewpoint.

I very much appreciate Foster's effort to test the theory. One problem with his test is that the displays are square. The best paradigms for seeing a low-spatial frequency functional scale are probably those that imitate the circumstances where phantom contours are seen. Phantoms are seen if a rectangular uniform region is placed perpendicular to, but not parallel to, a moving grating (Weisstein, Maguire & Berbaum 1977). A square figure has both parallel and perpendicular elements, and therefore does not satisfy the constraint on one-dimensionality that I stated in Section 35, Part II. Weisstein and Maguire (1978) have shown that phantoms are seen if the occluding figure appears to have a perceived depth that places the figure in or behind the plane of the moving stripes, but not in front of them. This property is compatible with my theory's description of filling-in as being restricted to binocularly consistent structural scales. A

stationary display would also have to satisfy this constraint.

It may be that a stationary display can only activate a functional scale commensurate with a single receptive field, rather than the "wave packet" of functional scales that a moving grating can induce. If such a functional scale is too small to perceive, then the adaptation paradigm that I suggested will fail. However, if the experiment can be made to work, then its impact will be profound.

10. Grimson and Stevens: Modules, models, and metaphors. A theoretical movement is espoused by Grimson and by Stevens that has made some interesting contributions to machine vision using biological insights as an intuitive guide. Some adherents of this movement advocate their philosophical beliefs with an enthusiasm that borders on religious intolerance. It is therefore difficult to discuss the often elementary concepts in their models without first stripping them of their scientifically irrelevant hyperbole.

For example, Stevens begins his commentary by saying that I have posed false dilemmas, and then goes on to build his case on false issues. He accuses my theory of not being "modular," whereas anyone who has read my articles with any care - including the present target article - realizes that I work in a hierarchical framework in which many subsystems coexist. Worse than that, I cannot think of any theorist who does not believe that subsystems exist. The modularity issue is a false one. Stevens also incorrectly accuses my theory of having "no notion . . . of symbolic information processing," of not discussing "what is the goal of the computation . . . what is the algorithm for the transformation?" In reality, I began making such theoretical contributions more than a decade before the present philosophical juggernaut coined its jargon. Stevens cannot understand that I have always done what he says I should be doing because he recognizes only one form of "information processing." Ironically, key aspects of this type of information processing are special cases of my own concepts, but Stevens seems unwilling or unable to understand this.

The unsettling confusion between means and ends. between wanting to understand human vision but hating to study human processes, is evident throughout Stevens's commentary. How should one interpret statements like "different information processing problems . . . have essentially independent solutions . . . the principle of modularity is rather subtle . . . it means that independent visual information, such as color and motion, are treated as independent by the visual system . . . "? If color and motion are independent, then why do color-contingent motion aftereffects occur (Hepler 1968; Sekuler 1975; Stromeyer & Mansfield 1970)? If not even color and motion are independent, then is the independence-of-modules hypothesis incorrect? Or is it just a philosophical slogan to be used as a weapon at scientifically inconvenient moments?

Grimson goes beyond Stevens's exhortations by reviewing some results of the Marr school. I say "Marr school" because specific scientific results, philosophical dogma, routine mathematical methods, and incantations of magical phrases like Primal Sketch are always packaged together by its adherents in a way that sometimes tends to obscure what is new in the scientific contributions themselves. I believe that the contributions of this vigorous approach deserve and require an analytic assessment of how each contribution dovetails with the rest of perceptual theory and data.

To start, I must note that Grimson does not reply to my criticisms about zero-crossings. As I stated in Section 16, Part I, the issue is not whether "the location of the change, and an amplitude associated with that change" are computed. The issue is how they are computed and represented in memory. Similarly, we all know that an algorithm for interpolating visual surfaces from depth information . . . and how the shape of the surface can be inferred" has been suggested. The issue is how this algorithm works and what it implies about human vision. In Grimson's commentary, a number of strong claims about this algorithm and related results of the Marr school are made. Unfortunately, these claims do not stand up to scrutiny. I will first review Grimson's surface interpolation algorithm to suggest why this algorithm begs the perceptual issues about filling-in that my target article has raised. I will then review problems of the Horn-Marr (Horn 1974; Marr 1974) lightness algorithm and their relationship to the Richter and Ullman (1982) retinal model. A similar analysis can be given of the other claims, and in much the same way.

Grimson's goal is to reconstruct, or fill-in, a smooth surface that interpolates a given set of zero-crossing points. He wants the surface to be as flat as possible between these zero-crossings. He calls this property the surface consistency constraint. As any student of calculus knows, such a surface should have the smallest possible second derivatives to keep it as linear as possible while also going through the prescribed points. In order even to define these derivatives, Grimson decides for convenience to interpolate the zero-crossings with a smooth surface, despite the fact that perceptual boundaries are not smooth. Denoting a smooth surface by z = f(x,y), its second derivatives are  $f_{xx}$ ,  $f_{xy}$ , and  $f_{yy}$ . To avoid worrying about whether these derivatives are small but negative or small but positive, it is routine to consider their squares  $f_{xx}^2$ ,  $f_{xy}^2$ , and  $f_{yy}^2$ . To measure whether the derivatives are small all over the surface, one routinely considers sums of the squared derivatives integrated across (x,y)space. A good surface f will have a small integral.

Two elementary measures of the flatness of f(x,y) are

$$\theta_1(f) = \iint (f_{xx}^2 + f_{yy}^2) dx dy$$
 (4)

$$\theta_2(f) = \iint (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2) dx dy$$
 (5)

A measure such as  $\theta_1(f)$  omits the mixed partial derivatives  $f_{xy}$ . Hence a curved surface like  $f_1(x,y) = xy$  can minimize  $\theta_1(f)$ , since  $\theta_1(f_1) = 0$ . Consequently Grimson decides to minimize  $\theta_2(f)$ .

In Grimson (1981, 1982a, 1982b), many pages are spent trying to justify the choice of  $\theta_2(f)$ . None of these arguments says anything about vision. They all review routine mathematical results on whether  $\theta_2(f)$  has a unique minimizing solution. To anyone who has taken an undergraduate course in calculus of variations, it is obvious that **Grimson** has chosen the functional  $\theta_2(f)$  by fiat based upon elementary ideas about linear interpolation. Grimson's claim that "the theory has also been given a solid physical motivation by relating the model . . . to the physics of thin plates" just means that a thin plate also tried to stay smooth yet bend as little as possible subject to constraints upon it. By contrast with Grimson's functional, the potential energy

$$\theta_{3}(f) = \iint \left[ (f_{xx}^{2} + f_{yy}^{2})^{2} - 2(1 - \mu)(f_{xx}f_{yy} - f_{xy}^{2}) \right] dxdy \quad (6)$$

of a thin plate is derived from a physical model of the plate – just the type of model that the Marr advocates criticize me for developing. Hamilton's Principle in classical mechanics then physically justifies the claim that the plate f will minimize the potential energy  $\theta_3(f)$ . No such physical model is offered by Grimson. Instead, Grimson's formal theory boils down to finding approximately linear interpolations between sets of points using standard methods.

When one studies how Grimson's algorithm actually finds this linear interpolation, it becomes apparent that the algorithm begs the filling-in question in a most serious fashion. His "random dot wedding cake" is illustrative (Response Figure 3). How does Grimson's algorithm decide not to interpolate points of similar disparity that are spatially separated in (x, y) space by points of different disparity? How does the algorithm decide instead to join different layers of the wedding cake? First, the zerocrossing points are represented as points in the threedimensional (x,y,d) space, where the height  $d_i = d(x_i,y_i)$ above the *i*th zero-crossing location  $(x_i, y_i)$  denotes the disparity at this location. This construction assumes a homunculus within the Marr-Poggio theory because the computation of disparity within that theory does not take place in (x,y,d) space. A computer program constructs (x,y,d) space from the outcomes of binocular matches. This is, however, a minor difficulty compared with the ones raised by the next steps of the algorithm.

To get his algorithm started, Grimson has to interpolate the zero-crossings with a smooth surface. As Grimson (1981, p. 186) writes, "Determine a feasible initial surface approximation (any surface approximation which contains the known stereo depth values . . .)." In other words, Grimson assumes that global filling-in has already occurred and then tries to make it more linear. I contrast this fact with Grimson's assertion that "the perceptions of the 5% random dot pattern . . . follow straightforwardly from the theory. . . ." Of course no problem arises in interpolating even a 5% density of random dots if one assumes that an interpolation has already been given by an undisclosed mechanism.

How does the algorithm progress from the hypothesized filling-in to a flatter filling-in? Here Grimson uses routine methods of steepest descent as implemented by Lagrange multipliers (Grimson 1981, Chapter 7). These methods are mathematically routine, but as visual processes they are highly improbable. Given a trial surface  $f_0$ , the algorithm computes a new surface  $f_1$  such that  $\theta_2(f_0)$  decreases most rapidly in the direction from  $f_0$  to  $f_1$  without leaving the zero-crossing points. To perform this computation, the algorithm uses a finite mesh of n(x,y)points over which a discrete approximation to the surface will be found. The algorithm also has to satisfy m additional constraints if there are m zero-crossing points. Thus every step of the algorithm takes place in an n + mdimensional space (Grimson 1981, p. 175), not the threedimensional (x,y,d) space. Grimson notes that, from a mathematical viewpoint, this search is local in the sense

that each surface is a point in the space of all surfaces. He does not mention, however, that this search is global as a visual process. The search procedure implies that the visual process has the following properties. It can simultaneously sense all its estimated values across a surface  $f_0$ , can compute the second derivatives of all surfaces f close to  $f_0$  at all n mesh points, can figure out the rates with which all these surfaces would change  $\theta_2(f)$ , and can then pick out among all these surfaces the new surface  $f_1$  that will most quickly decrease  $\theta_2(f)$  without disrupting the mzero-crossing estimates. Then this process repeats itself iteratively. Although no visual process is ever defined, one can safely remark that its action cannot possibly be isomorphic with the rules of this algorithm.

Given that the algorithm converges, what does the wedding cake teach us? In Figure 3, every (x,y) point is associated with a single disparity value, and only with this disparity value. Thus the wedding cake is subject to all of my criticisms of Sperling-Dev models (Section 14, Part I) despite **Grimson's** claim that "most of the criticisms leveled by Grossberg at what he calls Sperling-Dev models do not apply to the MPG model. One already discussed is the interpolation of disparity data to obtain





Figure 3. A random dot wedding cake. This figure shows how the Grimson surface reconstruction algorithm computes a Sperling-Dev surface that interpolates given  $(x_i, y_i, d_i)$  points as linearly as possible, where  $(x_i, y_i)$  designates the 2-dimensional position of the *i*th zero-crossing and  $d_i$  measures its disparity. This algorithm works by assuming that a filling-in surface has already been provided and then uses a standard steepest descent procedure to flatten out the initial surface. (From W. E. L. Grimson, From Images to Surface, M.I.T. Press, 1981, p. 192. Reprinted with permission.)

figure-ground effects and to compute complete surface representations."

The individual models of the Marr school also deserve detailed analysis as contributions to the neural modeling literature, where they are so intended. For example, Grimson writes, "Horn's algorithm (Horn 1974)... clearly illustrates one method for computing lightnesses... which clearly explains effects like the Cornsweet and Craik-O'Brien effects...." Marr (1974) interprets the Horn algorithm as a model of the retina by assigning each algorithmic stage an anatomical interpretation. I will indicate below how the Marr (1974) retinal model and the later Richter and Ullman (1982) retinal model, which **Grimson** also mentions, are related to my own efforts to model retinal dynamics.

In Grossberg (1970a, Sections 7-14), I classify minimal networks capable of computing the reflectances in an input pattern, and in Grossberg (1972d) I interpret these results retinally and relate them to Land's retinex theory. My networks are defined in terms of three successive computational stages. The first two stages normalize and partially filter the input pattern. I show how these stages can be accomplished by a subtractive competition in which signal thresholds are already embedded, or by a logarithmic transduction feeding into a subtractive stage followed by signal thresholds, or by a shunting competition followed by a signal threshold stage. This classification of possible instantiations illustrates the clear demarcation between abstract computation and physical realization that I was already actively pursuing in the 1960s.

The Horn (1974) and Marr (1974) articles also invoke three stages to compute lightness. Marr's first two stages are my logarithmic transduction feeding into a subtractive stage followed by signal thresholds. Thus these two stages do not represent a new model at all. In particular, the bipolar cell response to inputs  $I_i$  is defined by

$$J_i = \log(I_i) - \sum_{k \in K} w_{ki} \log(I_k)$$
(7)

which implies

$$J_{i} = \log \left[ \frac{I_{i}}{\prod_{k \in K} I_{k}} \right]$$
(8)

(Marr 1974, equation 3, p. 1379). This form factor does not meet intracellular data about bipolar cells, notably the S-shaped form of the shift property (Werblin 1971). In Grossberg (1972d, equation (7), p. 52, and equations 9–10, p. 54), I realized that retinal data suggested the choice of a shunting competitive model for bipolar cell responses. I have summarized some of the reasons for my choice in the target article, Section 23, Part I, including the S-shaped shift property.

The third stage of my model differs in a fundamental way from the work of Horn and Marr. I realized in the 1960s that one does not have to construct an output pattern that is a faithful reproduction of an input pattern whose background illuminations have been discounted. In fact, the existence of opponent-process ganglion cells at the output end of the retina strongly argues against such a reconstruction process. The output pattern can be an abstract representation of the input pattern. All one needs at the output end are tuning curves that are sensitive to pattern reflectances. In Grossberg (1970a), formal considerations led me to suggest a third stage that is built up from subtractive opponent processes. In Grossberg (1972d, Section 3), I pointed out how this procedure could be realized by bipolar cell and amacrine cell interactions with opponent-process ganglion cells.

By contrast, Marr (1974) accepts Horn's requirement that the third stage invert the first stage of the process. Horn's third stage, however, is an excitatory feedback interaction among the thresholded output signals from the second stage. Marr (1974) realizes that this property creates difficulties because "retinal ganglion cells are not pre-synaptic to any other retinal cells" (p. 1382) and because "it is taken as a basic constraint on the . . . reconstruction algorithm that the whole computation should be carried out using variables whose signs do not change" (p. 1381). Because it was already well known that amacrine cells can inhibit ganglion cells, Marr (1974) does not take into account the fact that the output cells of the retina are ganglion cells; he seeks his interpretation of the algorithm's third stage among the many types of amacrine cells. Unfortunately, no matter what interpretation is used, an opponent-process ganglion cell cannot be synthesized from sums of excitatory activities.

Grimson's commentary mentions the retinal ganglion cell model of Ullman and Richter (1982) as an example of how "computational theories and algorithms [are used] to predict and understand mechanisms in the visual system." Grimson notes that this model is based on a " $\nabla^2 G$ convolution" and a "time derivative of a  $\nabla^2 G$  convolution" to further argue for the importance of zero-crossing computations. I believe that it is essential to distinguish the  $\nabla^2 G$  convolution from the more impoverished zerocrossing computation. Even a  $\nabla^2 G$  operation just approximates a difference-of-Gaussians, and a convolution of a difference-of-Gaussians is just  $\sum_{k} I_{k}(D_{ki} - E_{ki})$  in the notation of Section 24, Part II. This form factor has been a component of retinal ganglion cell models, including my own, for more than a decade. A time derivative of this form factor crudely approximates motion-sensitive interactions that are built up from excitatory and inhibitory interactions (Barlow & Levick 1965; Grossberg 1970a; Miller 1979; Wyatt & Daw 1975). Form factors that include the major Richter and Ullman (1982) computations have also been previously studied. These form factors include the Richter and Ullman expression

$$C(1 - e^{-\alpha(t - \beta)}) - S(1 - e^{-\gamma(t - \delta)})$$
(8)

for asynchronous center (C) and surround (S) interactions (Grossberg 1970a, Sections 9–11); and their expression

$$x_{1}(t) = S(t) - \int_{0}^{t} I(t-v)x_{2}(v)dv$$
 (9)

$$x_{2}(t) = \int_{0}^{t} E(t - v) x_{1}(v) dv$$
 (10)

for a dyadic negative feedback loop (Grossberg 1970a, Section 3). I used such form factors in Grossberg (1970a) to study the dynamics of sustained and transient responses before most of the data simulated by Richter and Ullman (1982) were published. These historical facts do not deny the usefulness of the Richter and Ullman (1982) study. They do, however, illustrate the inappropriateness of the proprietary tone with which the  $\nabla^2 G$  convolution and its time derivative are discussed by Grimson.

Such is also the tone of Marr and Poggio's (1976, p. 287) incorrect statement that "There has hitherto been no careful study of a cooperative algorithm in the context of a carefully defined computational problem. . . ." This inaccurate observation was directed at the algorithm of Dev (1975) that Marr and Poggio (1976) used, as well as at the entire field of neural modeling. There is this proprietary tone in the judgment: "Much has been published recently on possible cooperative processes in nervous systems, ranging from the 'catastrophe' literature . . . to various attempts of more doubtful credibility" (Marr & Poggio 1976, p. 287). This tone is also involved in expressions of the belief that the phrase Primal Sketch somehow turns some familiar elementary notions into something more than that, indeed for Grimson into a "rich symbolic representation." In a similar tone, priority is being claimed for a theoretical method that others were effectively using long before the Marr school. Such a proprietary attitude is not appropriate for reasoned scientific discussion. I hope that it changes soon.

Grimson says that I do not precisely define a matching process and that "reversing the contrast of one monocular pattern would result in similar representations." Both assertions are contradicted by the equations in Appendix A.

11. Julesz and Rosen: Springs, magnets, and networks. I am glad that the commentary of Julesz reviewed some of his seminal contributions to visual perception. Such results cannot be contemplated and admired often enough. The reminiscence that his original cooperative model of stereopsis "was not comprehensible either to the physicists or the psychologists" illustrates an unfortunate fact of life about the communication difficulties that await new visual theories. Even Julesz needed to use eighteenthand nineteenth-century concepts to effectively communicate his twentieth-century ideas. That these concepts were already realized in hardware certainly expedited the communication process.

Processes underlying perception will also find their way into hardware, and may even one day help to model the dynamics of nonperceptual phenomena. Such progress will be expedited if, to use **Rosen's** phrase, a frontal attack on the "radical epistemological implications" of perceptual data is supported along with approaches based on traditional concepts.

12. Laming: The unstable world of linear positive feedback. Many of the general concerns expressed by Laming have already received a reply in the foregoing sections of this Response. One of his assertions does require a separate reply, however: "for the most part visual perception is veridical, suggesting a basically linear process."

Obviously many visual processes involve internal positive feedback. Helmholtz himself acknowledged this with his doctrine of "unconscious inference," a doctrine, moreover, that challenges the statement that "visual perception is veridical." When positive feedback is made linear in a visual context, an unstable model is the result (Section 28, Part II). My "parsimonious set of theoretical constructs" thus mathematically contradicts Laming's belief in a linear visual process. Laming does not seem to consistently hold this belief in veridicality, however, since he also says that "a sigmoid signal function . . might . . . simply be a physiological device for suppressing the contribution of certain units, allowing the perceptual process to be carried out by other units still operating within their linear dynamic range." This statement is compatible with my results about sigmoids, although it is far from a working understanding because a sigmoid signal function in an improperly designed network can just as easily saturate units as enable them to operate linearly. How such saturation can be prevented is now well understood. The answer to Laming's question "why explore *these particular ideas*" can only be found by taking the time to explore them.

## 13. Poltrock and Shaw: Zulus, scale selection, and dipole fields. A number of interesting questions about the theory are raised by Poltrock & Shaw, but then these commentators go on to draw strong conclusions that do not follow from them. I will illustrate this assertion with two examples.

**Poltrock & Shaw** note that "Zulu tribesmen, who lacked experience with linear architecture, failed to perceive" the Ponzo illusion. They claim that this result poses a serious problem for my theory because "the activation pattern arises naturally from the computational processes required for binocular vision." The Zulu results must be discussed in terms of feature development mechanisms, namely, bottom-up adaptive filtering of input patterns during a period of developmental plasticity (Grossberg 1976a, 1976b), not in terms of the lateral interactions that react to the filtered patterns at the next network level of an adult network.

Poltrock & Shaw correctly note that monocular stimuli may self-match and may also contain scaling information that differentially activates some scales more than others. These commentators call this an "apparent contradiction." I do not see why. Self-matches are not subject to disparity restrictions; so they can, other things being equal, activate more scales than binocular matches. Hence the perceived depth of a monocularly viewed pattern may be more ambiguous than its perceived depth due to binocular viewing. This fact does not prevent the size scales within a monocular scene from also modulating network activity in the manner indicated by Figure 12 of Part II (target article). A quantitative understanding of such interactions will only come from massive numerical simulations of multiple scale reactions to numerous monocular and binocular stimuli. Such simulations are now being undertaken. The other issues raised by these commentators are also worthy of serious quantitative assessment but do not, to my mind, qualitatively contradict the theory.

Because of the many difficulties that are inevitably faced when building any new dynamical theory, my colleagues and I are carrying out our simulations in a conservative fashion, since we want to develop network designs that deal with a wide range of phenomena. While these simulations are going on, we do not "avoid settling on a specific, testable form of the network." Formulas like equation (23), Part II, and those in Appendix A are already as testable as other models in the literature. Parametric analyses and predictions are regularly contributed (e.g., Carpenter & Grossberg 1981, 1983; Cohen & Grossberg 1983b; and Grossberg 1983b.) Paradigms are suggested that may yield counterintuitive results. Nonetheless, the synthesis of a dynamical theory for explaining global visual phenomena necessarily proceeds on a slower time scale. To appreciate how rapidly progress is being made, one needs only recall the years of brilliant contributions that were needed for Bohr's model of the hydrogen atom to generate Schrödinger's model of the hydrogen atom (Hermann 1971; Rozental 1967).

Poltrock & Shaw question "whether any perceptual experiences are excluded" by my theory. Let me remind them that my major conclusions have been derived using networks constructed from a single physical process – the gated dipole field and its adaptively modifiable input filters. If this process is capable of even raising such a question, then it is worthy of serious investigation by a large number of gifted people. My colleagues and I at the Center for Adaptive Systems will heartily encourage such investigations on both the experimental and theoretical levels.

## References

- Amari, S. (1977) Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics* 27:77-87. [DSL]
- (1982) Competitive and cooperative aspects in dynamics of neural excitation and self-organization. In: *Competition and cooperation in neural nets* (Lecture Notes in Biomathematics, No. 45), ed. S. Amari & M. A. Arbib, pp. 1-28. Berlin: Springer Verlag. [GJD, rSG]
- Amari, S. I. & Arbib, M. A. (1977) Competition and cooperation in neural nets. In: Systems neuroscience, ed. J. Metzler. New York: Academic Press. [taSG]
- Arend, L. E. (1973) Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading. *Psychological Review*, 80:374–95. [LEA]
- Arend, L. E., Buehler, J. N. & Lockhead, G. R. (1971) Difference information in brightness perception. *Perception and Psychophysics* 9:367-70. [LEA]
- Arend, L. E. Jr., Lange, R. V. & Sandick, B. L. (1981) Nonlocal determination of brightness in spatially periodic patterns. *Perception and Psychophysics* 29:310-16. [taSG]
- Attneave, F. (1954) Some informational aspects of visual perception. Psychological Review 61:183-93. [DHF, taSG]
- Barlow, H. B. (1972) Optic nerve impulses and Weber's Law. In: Sensory coding, ed. W. R. Uttal. Boston: Little, Brown. [rSG]
- Barlow, H. B. & Levick, W. R. (1965) The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology* 178:447-504. [rSG]
- Baylor, D. A. & Hodgkin, A. L. (1974) Changes in time scale and sensitivity in turtle photoreceptors. *Journal of Physiology* 242:729-58. [taSC]
- Baylor, D. A., Hodgkin, A. L. & Lamb, T. D. (1974a) The electrical response of turtle cones to flashes and steps of light. *Journal of Physiology* 242:685-727. [taSG]
- (1974b) Reconstruction of the electrical responses of turtle cones to flashes and steps of light. Journal of Physiology 242:759-91. [taSG]
- Beck, J. (1972) Surface color perception. Ithaca, N.Y.: Cornell University Press. [taSG]
- Bergström, S. S. (1966) A paradox in the perception of luminance gradients, I. Scandinavian Journal of Psychology 7:209-224. [rSG]
  - (1967a) A paradox in the perception of luminance gradients, II.
  - Scandinavian Journal of Psychology 8:25-32. [rSG] (1967b) A paradox in the perception of luminance gradients, III.
  - Scandinavian Journal of Psychology 8:33-37. [rSG] (1973) A note on the neural unit model for contrast phenomena. Vision
- Research 13:2087-2092. [rSG] Blake, R. & Fox, R. (1973) The psychophysical inquiry into binocular
- Blake, R. & Fox, R. (1973) The psychophysical inquiry into Dinocula summation. Perception and Psychophysics 14:161-85. [HB]
- Blake, R., Sloane, M. & Fox, R. (1981) Further developments in binocular summation. Perception and Psychophysics 30:266-76. [taSG]

## References/Grossberg: Quantized geometry of visual space

Blakemore, C., Carpenter, R. H. & Georgeson, M. A. (1970) Lateral inhibition between orientation detectors in the human visual system. *Nature* 228:37-39. [taSG]

Blank, A. A. (1978) Metric geometry in human binocular perception: Theory and fact: In: Formal theories of visual perception, eds. E. L. J.

Leeuwenberg & H. F. J. M. Buffart. New York: Wiley. [taSG, MW] Boynton, R. M. (1968) The psychophysics of vision. In: Contemporary theory and research in visual perception, ed. R. N. Haber. New York: Holt, Rinchart, and Winston. [rSG]

- Bridgeman, B. (1971) Metacontrast and lateral inhibition. Psychological Review 78:528-39. [BB]
- (1977) A correlational model applied to metacontrast: Reply to Weisstein, Ozog, and Szoc. Bulletin of the Psychonomic Society 10:85-88. [BB]
   (1978) Distributed sensory coding applied to simulations of iconic storage
- and metacontrast. Bulletin of Mathematical Biology 40:605-623. [BB] Buffart, H. (1978) Brightness and contrast. In: Formal theories of visual perception, eds. E. L. J. Leeuwenberg & H. F. J. M. Buffart. New

York: Wiley. [HB]
(1981) A theory of cyclopean perception. Nijmegen: University. [HB]
(1982) Brightness estimation: A transducer function. In: Psychophysical judgment and the process of perception, eds. H.-G. Geissler, H. F. J. M. Buffart, P. Petzoldt & Y. M. Zabrodin. Amsterdam: North-Holland Publishing Co. [HB]

Buffart, H., Leeuwenberg, E. & Restle, F. (1981) Coding theory of visual pattern completion. Journal of Experimental Psychology 7:241-74. [HB]

Caelli, T. M. (1982) Visual perception: Theory and practice. Oxford: Pergamon Press. [TC]

Caelli, T., Hoffman, W. C. & Lindman, H. (1978) Apparent motion: Selfexcited oscillation induced by retarded neuronal flows. In: Formal theories of visual perception, ed. E. L. J. Leeuwenberg & H. F. J. M. Buffart, New York: Wiley. [taSG]

Carpenter, G. A. & Grossberg, S. (1981) Adaptation and transmitter gating in vertebrate photoreceptors. Journal of Theoretical Neurobiology 1:1-42. [tarSC]

(1983) Dynamic models of neural systems: Propagated signals, photoreceptor transduction, and circadian rhythms. In: Oscillations in mathematical biology, ed. J. P. E. Hodgson. New York: Springer-Verlag. [rSG]

- Cogan, A. L., Silverman, G. & Sekuler, R. (1982) Binocular summation in detection of contrast flashes. *Perception and Psychophysics* 31:330-38. [taSG]
- Cohen, M. A. & Grossberg, S. (1983a) Some global properties of binocular resonances: Disparity scaling, filling-in, and figure-ground synthesis. In: Figural synthesis, eds. T. Caelli & P. Dodwell. Hillsdale, N.J.: Erlbaum. [tarSG]

 (1983b) The dynamics of brightness perception. In preparation. [rSG]
 (1983c) Absolute stability of global pattern formation and parallel memory storage in competitive neural networks. *Transactions IEEE*, in press. [rSG]

- Coren, S. (1969) Brightness contrast as a function of figure-ground relations. Journal of Experimental Psychology 80:517-24. [SG]
   (1972) Subjective contours and apparent depth. Psychological Review
- 79:359-67. [taSG]

Coren, S., Porac, C. & Ward, L. M. (1979) Sensation and perception. New York: Academic Press. [SC]

- Cornsweet, T. N. (1970) Visual perception. New York: Academic Press. [HB, SC, taSG]
- Crick, F. H. C., Marr, D. & Poggio, T. (1980) An information processing approach to understanding the visual cortex. In: The Cerebral Cortex, Neurosciences Research Program. [WELG]

Curtis, D. W. & Rule, S. J. (1978) Binocular processing of brightness information: A vector-sum model. Journal of Experimental Psychology: Human Perception and Performance 4:132-43. [HB]

- Dalenoort, C. J. (1982a) In search of the conditions for the genesis of cell assemblies: A study in self-organization. Journal of Social and Biological Structures 5:161-87. [GJD]
- . (1982b) Modelling cognitive processes in self-organizing neural networks, an exercise in scientific reduction. In: *Biomathematics in 1960*, eds. L. M. Ricciardi & A. C. Scott, Amsterdam: North-Holland, pp. 133-44. [GJD]

Day, R. H. (1972) Visual spatial illusions: A general explanation. Science 175:1335-40. [taSG]

- de Lange, H. (1957) Attenuation characteristics and phase-shift characteristics of the human fovea-cortex systems in relation to flicker-fusion phenomena. Delft: Technical University. [HB]
- Deregowski, J. B. (1973) Illusion and culture. In: Illusions in nature and art, eds. R. L. Gregory & G. H. Gombrich, pp. 161-192. New York: Scribner's. [SEP]

- Dev, P. (1975) Perception of depth surfaces in random-dot stereograms: A neural model. International Journal of Man-Machine Studies 7:511-28. [tarSG]
- de Weert, Ch. M. M. & Levelt, W. J. M. (1974) Binocular brightness combinations: Additive and nonadditive aspects. Perception and Psychophysics 15:551-62. [HB]
- Diner, D. (1978) Hysteresis in human binocular fusion: A second look. Ph.D. thesis California Institute of Technology, Pasadena. [BJ]
- Dodwell, P. C. (1975) Pattern and object perception. In: Handbook of Perception, Vol. 5: Seeing, eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [taSG]

Eijkman, E. G. J., Jongsma, H. J. & Vincent, J. (1981) Two-dimensional filtering, oriented line detectors, and figural aspects as determinants of visual illusions. *Perception and Psychophysics* 29:352-58. [taSC]

Ellias, S. A. & Grossberg, S. (1975) Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. *Biological Cybernetics* 20:69-98. [tarSC, DSL]

Emmert, E. (1881) Grössenverhaltnisse der Nachbilder. Klinische Monatsblatt der Augenheilk unde 19:443-50. [taSG]

Engel, G. R. (1967) The visual processes underlying binocular brightness summation. Vision Research 7:753-67. [HB]

- (1969) The autocorrelation function and binocular brightness mixing. Vision Research 9:1111-30. [HB]
- Enroth-Cugell, C. & Robson, J. G. (1966) The contrast sensitivity of retinal ganglion cells of the cat. Journal of Physiology 187:517-52. [taSG]

Fender, D. & Julesz, B. (1967) Extension of Panum's fusional area in binocularly stabilized vision. Journal of the Optical Society of America 57:819-30. [HB, taSG, BJ]

Festinger, L., Coren, S. & Rivers, G. (1970) The effect of attention on brightness contrast and assimilation. American Journal of Psychology 83:189-207. [SC]

- Foley, J. M. (1968) Depth, size and distance in stereoscopic vision. Perception and Psychophysics 3:265-74. [JMF]
- (1976) Binocular depth mixture. Vision Research 16:1263-67. [JMF]
  (1980) Binocular distance perception. Psychological Review
  87:411-34. [JMF, taSG]
- Foster, D. H. (1978) Visual apparent motion and the calculus of variations. In: Formal Theories of Visual Perception, eds. E. L. J. Leeuwenberg & H. F. J. M. Buffart, pp. 67-82. New York: Wiley. [DHF]
- (1980) A spatial perturbation technique for the investigation of discrete internal representations of visual patterns. *Biological Cyberbetics* 38:159-69. [DHF]
- Fox, R. & McIntyre, C. (1967) Suppression during binocular fusion of complex targets. Psychonomic Science 8:143-44. [HB]
- Freeman, W. J. (1973) Cinematic display of spatial structure of EEG and averaged evoked potentials (AEPs) of olfactory bulb and cortex. *Electroencephal. Clin. Neurophysiol.* 37:199. [WJF]
- (1975) Mass action in the nervous system. New York: Academic Press. [WJF]
- (1979a) EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb. *Biological Cybernetics* 35:221-34. [WJF]
- (1979b) Nonlinear dynamics of paleocortex manifested in the olfactory EEG. Biological Cybernetics 35:21-37. [WJF, rSC]

(1979c) Nonlinear gain mediating cortical stimulus response relations. Biological Cybernetics 33:237-47. [WJF]

- (1981) A physiological hypothesis of perception. Perspectives in Biology and Medicine 24:561-92. [WJF]
- Freeman, W. J. & Schneider, W. (1982) Changes in spatial patterns of rabbit olfactory EEG with conditioning to odors. *Psychophysiology* 19:44-56. [W]F]
- Frisby, J. P. (1979) Seeing. Oxford: Oxford University Press. [WELG]
- Frisby, J. P. & Julesz, B. (1975) Depth reduction effects in random line stereograms. *Perception* 4:151–58. [BJ]
- Gerrits, H. J. M., de Haan, B. & Vendrick, A. J. H. (1966) Experiments with retinal stabilized images: Relations between the observations and neural data. Vision Research 6:427-40. [rSG]
- Gerrits, H. J. M. & Timmerman, J. G. M. E. N. (1969) The filling-in process in patients with retinal scotomata. Vision Research 9:439–42. [rSG]

Gerrits, H. J. M. & Vendrik, A. J. H. (1970a) Artificial movements of a stabilized image. Vision Research 10:1443-56. [HB]

(1970b) Simultaneous contrast, filling-in process and information processing in man's visual system. *Experimental Brain Research* 11:411-30. [HB, rSG]

<sup>(1972)</sup> Eye movements necessary for continuous perception during stabilization of retinal images. *Bibliotheca Ophthalmologica* 82:339-47. [HB]

- (1974) The influence of stimulus movements on perception in parafoveal stabilized vision. Vision Research 14:175-80. [HB]
- Gibson, J. J. (1950) Perception of the visual world. Boston: Houghton Mifflin. [taSG]
- Gilchrist, A. L. (1977) Perceived lightness depends on perceived spatial arrangement. Science 195:185–87. [SEP]
- (1979) The perception of surface blacks and whites. Scientific American 240:112-24. [taSG, SEP]
- Glass, L. (1970) Effect of blurring on perception of a simple geometric pattern. Nature 228:1341-42. [WELG]
- Glass, L. & Switkes, E. (1976) Pattern recognition in humans: Correlations which cannot be perceived. *Perception* 5:67-72. [taSG]
- Gogel, W. C. (1956) The tendency to see objects as equidistant and its reverse relations to lateral separation. *Psychological Monograph* 70 (whole no. 411). [taSC, DSL]
- (1965) Equidistance tendency and its consequences. *Psychological Bulletin* 64:153-63. [taSC, DSL]
- (1970) The adjacency principle and three-dimensional visual illusions. Psychonomic Monograph Supplement 3 (whole no. 45), pp. 153-169. [taSG, DSL]
- Gonzales-Estrada, M. T. & Freeman, W. J. (1980) Effects of carnosine on olfactory bulb EEG, evoked potentials and DC potentials. Brain Research 202:373-86. [WJF]
- Graham, N. (1981) The visual system does a crude Fourier analysis of patterns. In: Mathematical psychology and psychophysiology, ed. S. Grossberg. Providence, R.I.: American Mathematical Society. [taSG]
- Graham, N. & Nachmias, J. (1971) Detection of grating patterns containing two spatial frequencies: A test of single-channel and multiple channel models. Vision Research 11:251-59. [taSG]
- Graham, N., Robson, J. G. & Nachmias, J. (1978) Grating summation in fovea and periphery. Vision Research 16:816-25. [taSG]
- Gregory, R. L. (1966) Eye and brain. New York: McGraw-Hill. [taSG]
- Grimson, W. E. L. (1981) A computer implementation of a theory of human stereo vision. Philosophical Transactions of the Royal Society of London B 292:217-53. [WELG]
- (1982a) A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society of London B* 298:395-427. [WELG, rSG]
- (1982b) From images to surfaces: a computational study of the human early visual system. Cambridge, Mass.: MIT Press. [rSG, WELG, KAS]
- (1963) Surface consistency contraints in vision. In: Computer Graphics and Image Processing (in press). [WELG]
- Grossberg, S. (1968) Some physiological and biochemical consequences of psychological postulates. Proceedings of the National Academy of Sciences 60:758-65. [taSG]
- (1969a) On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics* 1:319-50. [rSG]
- (1969b) On the serial learning of lists. Mathematical Biosciences 4:201-53. [rSG]
- (1970a) Neural pattern discrimination. Journal of Theoretical Biology 27:291-337. [tarSG]
- (1970b) Some networks that can learn, remember, and reproduce any number of complicated space-time patterns, II. Studies in Applied Mathematics 49:135-66. [rSG]
- (1971a) On the dynamics of operant conditioning. Journal of Theoretical Biology 33:225-55. [rSG]
- (1971b) Pavlovian pattern learning by nonlinear neural networks. Proceedings of the National Academy of Sciences 68:828-31. [rSG]
- (1972a) A neural theory of punishment and avoidance, I. Qualitative theory. Mathematical Biosciences 15:39-67. [rSG]
   (1972b) A neural theory of punishment and avoidance, II. Quantitative
- theory. Mathematical Biosciences 15:253-85. [rSG]
- (1972c) Pattern learning by functional-differential neural networks with arbitrary path weights. In: *Delay and functional-differential equations* and their applications, ed. K. Schmitt, New York: Academic Press. [rSG]
- (1972d) Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik* 10:49–57. [tarSG]
- (1973) Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics* 52:217-57. [tarSC, DSL]
- (1974) Classical and instrumental learning by neural networks. In: Progress in theoretical biology, Vol. 3, eds. R. Rosen & F. Snell. New York: Academic Press. [rSG]
- (1975) A neural model of attention, reinforcement, and discrimination learning. International Review of Neurobiology 18:263–327. [rSG]

## References/Grossberg: Quantized geometry of visual space

(1976a) Adaptive pattern classification and universal recoding, 1: Parallel development and coding of neural feature detectors. *Biological Cybernetics* 23:121-34. [rSG, DSL]

(1976b) Adaptive pattern classification and universal recoding, II: Feedback expectation, olfaction, and illusions. *Biological Cybernetics* 23:187-202. [tarSG]

(1976c) On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics* 21:145-59. [DSL]

- (1978a) Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? Journal of Mathematical Psychology 17:199-219. [taSG]
- (1978b) Communication, memory, and development. In: Progress in theoretical biology, Vol. 5, eds. R. Rosen & F. Snell, New York: Academic Press. [taSG]
- (1978c) Competition, decision, and consensus. Journal of Mathematical Analysis and Applications 66:470-93. [tarSG]
- (1978d) Decisions, patterns, and oscillations in the dynamics of competitive systems with applications to Volterra-Lotka systems. *Journal of Theoretical Biology* pp. 101-30. [taSG]
- (1978e) A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In: *Progress in theoretical biology*, Vol. 5, eds. R. Rosen & F. Snell. New York: Academic Press. [tarSG]
- (1980a) Biological competition: Decision rules, pattern formation, and oscillations. Proceedings of the National Academy of Sciences 77:2338-42. [tarSG]
- (1980b) How does a brain build a cognitive code? *Psychological Review* 87:1-51. [HB, tarSG]

(1981) Adaptive resonance in development, perception, and cognition. In: Mathematical Psychology and Psychophysiology, ed. S. Grossberg. Providence, R.I.: American Mathematical Society. [tarSG]

- (1982a) Associative and competitive principles of learning and development: The temporal unfolding and stability of STM and LTM patterns. In: Competition and cooperation in neural networks, eds. S. I. Amari & M. Arbib. New York: Springer-Verlag. [rSG]
- (1982b) A psychophysiological theory of reinforcement, drive, motivation, and habit. Journal of Theoretical Neurobiology 1:286-369. [rSG]
- (1982c) The processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review* 89:529-72. [tarSG]
- (1982d) Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory. In: *Cognition and brain activity*, eds. J. Cohen, R. Karrer & P. Tueting. New York: New York Academy of Sciences. [tarSG]

(1982e) Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control. Boston: Reidel Press. [tarSG]

- (1983a) The adaptive self-organization of serial order in behavior: Speech and motor control. In: Perception of speech and visual form: Theoretical issues, models, and research, eds. E. C. Schwab & H. C. Nusbaum, New York: Academic Press. [rSG]
- (1983b) Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory. In: Brain and information: Evoked potential correlates, eds. R. Karrer, J. Cohen & P. Tueting. New York: New York Academy of Sciences.
- Grossberg, S. & Kuperstein, M. (1983) Adaptive dynamics of the saccadic eye movement system. In preparation. [rSG]
- Grossberg, S. & Levine, D. S. (1975) Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. *Journal of Theoretical Biology* 53:341-80. [tarSG, DSL]
- Grossberg, S. & Pepe, J. (1970) Schizophrenia: Possible dependence of associational span, bowing, and primacy vs. recency on spiking threshold. *Behavioral Science* 15:359-62. [rSG]
- (1971) Spiking threshold and overarousal effects in serial learning. Journal of Statistical Physics 3:95-125. [rSG]

Grünau, M. W. von (1979) The involvement of illusory contours in stroboscopic motion. Perception and Psychophysics 25:205-08. [taSG]

- Hagen, M. A. & Teghtsoonian, M. (1951) The effects of binocular and motiongenerated information on the perception of depth and height. *Perception* and Psychophysics 30:257-65. [taSG]
- Hamada, J. (1976) A mathematical model for brightness and contour perception. Hokkaido Report of Psychology HRP-11-76-17. [taSG]
  - (1980) Antagonistic and non-antagonistic processes in the lightness perception. Proceedings of XXII International Congress of Psychology, Leipzig, July 6-12. [tarSG]
- Hebb, D. O. (1949) The organization of behavior. New York: Wiley. [GJD]

- Hecht, S. (1934) Vision. II. The nature of the photoreceptor process. In: A handbook of general experimental psychology, ed. C. Murchison. Worcester, Mass.: Clark University Press. [TI]
- Helmholtz, H. L. F. von (1962) Treatise on physiological optics. J. P. C. Southall, trans. New York: Dover. [taSG]
- Hepler, N. (1968) Color: A motion-contingent after-effect. Science 162:376-77. [rSG]
- Hering, E. (1964) Outlines of a theory of the light sense. Cambridge, Mass.: Harvard University Press. [taSC]
- Hermann, A. (1971) The genesis of quantum theory (1899-1913). C. W. Nash, trans. Cambridge, Mass.: MIT Press. [rSG]
- Hildreth, E. C. (1980) Implementation of a theory of edge detection. MIT Artificial Intelligence Laboratory Technical Report TR-579. [WELG]
- Hochberg, J. (1964) Contralateral suppressive fields of binocular combination. Psychonomic Science 1:157-58. [HB]
- Hochberg, J. & Beck, J. (1954) Apparent spatial arrangement and perceived brightness. American Journal of Psychology 47:263-66. [SEP]
- Holway, A. F. & Boring, E. G. (1941) Determinants of apparent visual size with distance variant. American Journal of Psychology 54:21-37. [taSG]
- Horn, B. K. P. (1974) Determining lightness from an image. Computer Graphics and Image Processing 3:277-99. [rSG, WELG]
   (1977) Understanding image intensities. Artificial Intelligence 8:201-31. [WELG]
- Hubel, D. H. & Wiesel, T. N. (1977) Functional architecture of macaque monkey visual cortex. Proceedings of the Royal Society of London (B) 198:1-59. [taSG]
- Hurvich, L. M. & Jameson, D. (1955) Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision. Journal of the Optical Society of America 45:602-16. [rSG]
- Indow, T. (1979) Alleys in visual space. Journal of Mathematical Psychology 19:221-58. [TI]
- (1983) An approach to geometry of visual space with no a priori mapping functions. Journal of Mathematical Psychology (in press). [TI]
- Johansson, C. (1978) About the geometry underlying spontaneous visual decoding of the optical message. In: Formal theories of visual perception, eds. E. L. J. Leeuwenberg and H. F. J. M. Buffart. New York: Wiley. [taSG]
- Julesz, B. (1960) Binocular depth perception of computer-generated patterns. Bell System Technical Journal 39:1125-62. [B]
- (1962) Towards the automation of binocular depth perception (AUTOMAP). Proceedings of the IFIP Congress 62, 27 Aug-1 Sep 1962, pp. 439-444. Amsterdam: North Holland Publishing Co. [BJ]
- (1964) Binocular depth perception without familiarity cues. Science 145:356-62. [BJ]
- (1971a) Binocular depth perception in man—a cooperative model of stereopsis. In: Pattern recognition in biological and technical systems, eds. O.-J. Grusser & R. Klinke, pp. 300–315. Proceedings of the German Cybernetic Society, Berlin, April 6–9, 1970. Berlin-Heidelberg: Springer-Verlag. [BJ]
- (1971b) Foundations of cyclopean perception. Chicago: University of Chicago Press. [HB, taSG, BJ]
- (1974) Cooperative phenomena in binocular depth perception. American Scientist 62:32-43. Reprinted in: Current trends in psychology: Readings from American Scientist ed. I. L. Janis. Los Altos, Calif.: W. Kaufmann. [B]]
- (1978a) Global stereopsis: Cooperative phenomena in stereoscopic depth perception. In: *Handbook of sensory physiology*, Vol. 8. *Perception*, eds. R. Held, H. W. Leibowitz & H.-L Teuber, pp. 215–256. Berlin-Heidelberg-New York: Springer-Verlag. [BJ]
- (1978b) Perceptual limits of texture descrimination and their implications to figure-ground separation. In: Formal theories of visual perception, eds.
   E. L. J. Leeuwenberg & H. F. J. M. Buffart. New York: Wiley. [taSG]
- Julesz, B. & Chang, J. J. (1976) Interaction between pools of binocular disparity detectors tuned to different disparities. *Biological Cybernetics* 22:107-19. [BJ]
- Just, M. A. & Carpenter, P. A. (1976) Eye fixations and cognitive processes. Cognitive Psychology 8:441-80. [DHF]
- Kaczmarek, L. K. & Babloyantz, A. (1977) Spatiotemporal patterns in epileptic seizures. Biological Cybernetics 26:199–208. [taSG]
- Kaufman, L. (1974) Sight and mind: An introduction to visual perception. New York: Oxford University Press. [taSG]
- Kaufman, L., Bacon, J. & Barroso, F. (1973) Stereopsis without image segregation. Vision Research 13:137-47. [HB, taSG]
- Klatt, D. H. (1980) Speech perception: A model of acoustic-phonetic analysis and lexical access. In: *Perception and production of fluent speech*. ed. R. A. Cole. Hillsdale, N.J.: Erlbaum. [taSG]
- König, A. & Brodhun, E. (1889) Experimentelle Untersuchungen Über die

psychophysische Fundamentalformel in Bezug auf den Gesichtssinn. Sitzungsberichte der preussischen Akademie der Wissenschaften, Berlin 27:641–44. [HB]

- Koffka, K. (1935) Principles of gestalt psychology. New York: Harcourt, Brace. [taSG]
- Kulikowski, J. J. (1978) Limit of single vision in stereopsis depends on contour sharpness. Nature 275:126–27. [taSG]
- Laming, D. R. J. (1973) Mathematical psychology. London: Academic Press. [DL]
- Land, E. H. (1977) The retinex theory of color vision. Scientific American 237:108-28. [LEA, taSG]
- Land, E. H. & McCann, J. J. (1971) Lightness and retinex theory. Journal of the Optical Society of America 61:1-11. [LEA, WELG]

x

- Lanze, M., Weisstein, N. & Harris, J. R. (1982) Perceived depth vs. structural relevance in the object-superiority effect. Perception and Psychophysics 31:376-82. [taSG]
- Leake, B. & Annines, P. (1976) Effects of connectivity on the activity of neural not models. Journal of Theoretical Biology 58:337-63. [TC]
- Leeuwenberg, E. (1982) The perception of assimilation and brightness contrast. Perception and Psychophysics 32:345-52. [HB]
- Legge, G. E. & Foley, J. M. (1980) Contrast masking in human vision. Journal of the Optical Society of America 70:1458-71. [JMF]
- Legge, G. E. & Rubin C. S. (1981) Binocular interactions in suprathreshold contrast perception. Perception and Psychophysics 30:49-61. [raSG]
- LeGrand, Y. (1957) Light, colour, and vision. New York: Dover Press. [rSG]
- Leshowitz, B., Taub, H. B. & Raab, D. H. (1968) Visual detection of signals in the presence of continuous and pulsed backgrounds. *Perception and Psychophysics* 4:207-13. [DL]
- Lettvin, J. Y. (1981) "Filling out the forms": An appreciation of Hubel and Wiesel. Science 214:518-20. [taSG]
- Levelt, W. J. M. (1965) On binocular rivalry. Soesterberg, The Netherlands: Institute for Perception, RVO-TNO. [HB, tarSG]
- Levine, D. S. & Grossberg, S. (1976) Visual illusions in neural networks: Line neutralization, tilt aftereffect, and angle expansion. Journal of Theoretical Biology 61:477-504. [tarSG, DSL]
- Logan, B. F., Jr. (1977) Information in the zero-crossings of bandpass signals. Bell System Technical Journal 56:487-510. [WELG]
- Luneberg, R. K. (1947) Mathematical analysis of binocular vision. Princeton, N.J.: Princeton University Press. [taSG, MW]
- (1950) The metric of binocular visual space. Journal of the Optical Society of America 60:637-42. [TI]
- McCourt, M. E. (1982) A spatial frequency dependent grating-induction effect. Vision Research 22:119-34. [JMF]
- Marr, D. (1974) The computation of lightness by the primate retina. Vision Research 14:1377. [rSG]
- (1976) Early processing of visual information. Philosophical Transactions of the Royal Society of London (B) 275:483-524. [WELG, KAS]
- (1977) Artificial Intelligence—a personal view. Artificial Intelligence, 9:37-48. [HB, WELG, KAS]
- (1978) Representing visual information. Lectures on Mathematics in the Life Sciences 10:101-80. [KAS]
- (1982) Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman. [BB, WELG]
- Marr, D. & Hildreth, E. (1980) Theory of edge detection. Proceedings of the Royal Society of London (B) 207:187-217. [taSG, WELG, KAS]
- Marr, D. & Poggio, T. (1976) Cooperative computation of stereo disparity. Science 194:283-87. [tarSG]
  - (1977) From understanding computation to understanding neural circuitry. Neurosciences Research Progress Bulletin 15:470-88. [KAS]
- (1979) A computational theory of human stereo vision. Proceedings of the Royal Society of London (B) 204:301-28. [JMF, taSG, WELG, KAS]
- Maudarbocus, A. Y. & Ruddock, K. H. (1973) Non-linearity of visual signals in relation to shape-sensitive adaptation processes. Vision Research 13:1713-37. [DHF]
- Mayhew, J. E. W. & Frisby, J. P. (1981) Psychophysical and computational studies towards a theory of human stereopsis. Artificial Intelligence 17:349-85. [WELG]
- Miller, R. F. (1979) The neuronal basis of ganglion-cell receptive-field organization and the physiology of amacrine cells. In: *The neuroscience fourth study program*, ed. F. O. Schmitt. Cambridge, Mass.: MIT Press. [rSG]
- Minor, A. V., Flerova, G. I. & Byzov, A. L. (1969) Integral evoked potentials of single neurons in the frog olfactory bulb (in Russian). *Neurophysiologica* 1:269-78. [WJF]
- Mori, T. (1982) Apparent motion path composed of a serial concatenation of translations and rotations. *Biological Cybernetics* 44:31-34. [DHF]
- Nachmias, J. & Kocher, E. C. (1970) Visual detection and discrimination of

luminance increments. Journal of the Optical Society of America 60:382-89. [DL]

Newell, A. (1980) Harpy, production systems, and human cognition. In: Perception and production of fluent speech, ed. R. Cole. Hillsdale, N.J.: Erlbaum. [taSG]

- O'Brien, V. (1958) Contour perception, illusion and reality. Journal of the Optical Society of America 48:112-19. [SC, taSG]
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957) The measurement of meaning. Urbana: University of Illinois. [taSG]
- Poggio, T. (1980) Neurons sensitive to random-dot stereograms in areas 17 and 18 of rhesus monkey. Society for Neuroscience Abstracts (November) 6. [BJ]
- Poggio, T. (1982) Trigger features or Fourier analysis in early vision: A new point of view. In: The recognition of pattern and form, lecture notes in biomathematics, ed. D. Albrecht. New York: Springer, 44:88-99. [WELG]
- Pollen, D. A. & Ronner, S. F. (1981) Phase relationships between adjacent simple cells in the visual cortex. Science 212:1409-11. [rSG]
  (1982) Spatial computation performed by simple and complex cells in the
- visual cortex of the cat. Vision Research 22:101-18. [rSG] Pulliam, K. (1981) Spatial frequency analysis of three-dimensional vision.
- Proceedings of the Society of Photo-Optical Instrumentation Engineers 303:71-77. [JMF, rSG]
- Raaijmakers, J. G. W. & Shiffrin, R. M. (1981) Search of associative memory. Psychological Review 88:93-134. [taSG]
- Rall, W. (1977) Core conductor theory and cable properties of neurons. In: Handbook of physiology: The nervous system, vol. I, Part I, ed. E. R. Kandel, pp. 39-97. Bethesda, Md.: American Physiological Society. [rSG]
- Rashevsky, N. (1968) Mathematical biophysics. Chicago: University of Chicago Press. [TI]
- Ratliff, F. (1965) Mach bands: Quantitative studies on neural networks in the retina. New York: Holden-Day. [BB, taSG]
- Rauschecker, J. P. J., Campbell, F. W. & Atkinson, J. (1973) Colour opponent neurones in the human visual system. *Nature* 245:42-45. [taSG]
- Restle, F. (1971) Mathematical models in psychology. Baltimore: Penguin Books. [taSC]
- Richards, W. (1975) Visual space perception. In: Handbook of perception, Vol. 5: Seeing, eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [taSC]
- Richards, W. & Marr, D. (1981) Computational algorithms for visual processing. M.I.T. Artificial Intelligence Lab. [taSG]
- Richards, W. & Miller, J. F., Jr. (1971) The corridor illusion. Perception and Psychophysics 9:421-23. [taSG]
- Richter, J. & Ullman, S. (1982) A model for the temporal organization of Xand Y-type receptive fields in the primate retina. *Biological Cybernetics* 43:127-45. [rSG, WELG]
- Robson, J. C. (1975) Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. In: *Handbook of perception* (Vol. 5), eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [taSG]
- Robson, J. G. & Graham, N. (1981) Probability summation and regional variation in contrast sensitivity across the visual field. Vision Research 21:409-18. [taSG]
- Rock, I. (1977) In defense of unconscious inference. In: Stability and constancy in visual perception, ed. W. Epstein. New York: Wiley. [HB]
- Rodieck, R. W. & Stone, J. (1965) Analysis of receptive fields of cat retinal ganglion cells. Journal of Neurophysiology 28:833-49. [taSG]
- Rozental, S., ed. (1967) Niels Bohr. New York: Wiley. [rSG] Rushton, W. A. (1965) Visual adaptation. The Ferrier Lecture, 1962.
- Proceedings of the Royal Society of London (B) 162:20-46. [WJF] Sakata, H. (1981) Mechanism of Craik-O'Brien effect. Vision Research
- 21:693–99. [rSG] Schriever, W. (1925) Experimentelle studien über stereokopische sehen. Zeitschrift fuer Psychologie 96:113–70. [SEP]
- Schrödinger, E. Müller-Pouillets Lehrbuch der Physik 11. Auflage, Zweiter Band. Braunschweig. [HB]
- Schwartz, E. L. (1980) Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vision Research 20:645-69. [taSG]
- Sekuler, R. (1975) Visual motion perception. In: Handbook of perception, Vol. 5, eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [rSG]
- Shepard, R. N. (1980) Multidimensional scaling, tree-fitting, and clustering. Science 210:390-98. [taSG]
- Shepard, R. N. & Chipman, S. (1970) Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology 1:1-17. [BB]

## References/Grossberg: Quantized geometry of visual space

Shepard, R. N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. Science 171:701-03. [DHF]

- Shepherd, G. M. (1972) Synaptic organization of the mammalian olfactory bulb. *Physiological Review* 52:864-917. [WJF]
- Shipley, T. (1965) Visual contours in homogeneous space. Science 150:348-50. [taSG]
- Singer, W. (1982) The role of attention in developmental plasticity. Human Neurobiology 1:41-43. [taSG]
- Smith, A. T. & Over, R. (1979) Motion aftereffect with subjective contours. Perception and Psychophysics 25:95-98. [taSG]
- Sperling, G. (1970) Binocular vision: A physical and a neural theory. American Journal of Psychology 83:461-534. [GJD, taSG, BJ]
- (1981) Mathematical models of binocular vision. In: Mathematical psychology and psychophysiology, ed. S. Grossberg. Providence, R.I.: American Mathematical Society. [taSG]
- Sperling, G. & Sondhi, M. M. (1968) Model for visual luminance discrimination and flicker detection. Journal of the Optical Society of America 58:1133-45. [taSG, DSL]
- Stevens, S. S. (1959) The quantification of sensation. Daedalus 88:606-21. [taSG]
- Stromeyer, C. F. III & Mansfield, R. J. W. (1970) Colored after-effects produced with moving edges. *Perception and Psychophysics* 7:108-14. [rSG]
- Swets, J. A. (1961) Is there a sensory threshold? Science 134:168-77. [DL]
- Tschermak-Seysenegg, A. von (1952) Introduction to physiological optics. P. Boeder, trans. Springfield, Ill.: C. C. Thomas. [taSG]
- Tynan, P. & Sekuler, R. (1975) Moving visual phantom: A new contour completion effect. Science 188:951-52. [JMF, taSG]
- Uttal, W. (1973) The psychobiology of sensory coding. New York: Harper and Row. [BB]
- van den Brink G. & Keemink, C. J. (1976) Luminance gradients and edge effects. Vision Research 16:155-59. [HB]
- van Nes, F. L. (1968) Experimental studies in spatio-temporal contrast transfer by the human eye. Utrecht: University. [HB]
- van Nes, F. L. & Bouman, M. A. (1965) The effects of wavelength and luminance on visual modulation transfer. Excerpta Medica International Congress Series 125:183-92. [HB]
- van Tuijl, H. & Leeuwenberg, E. (1979) Neon color spreading and structural information measures. *Perception and Psychophysics* 25:269-84. [HB]
- Von Békésy, G. (1968) Mach- and Hering-type lateral inhibition in vision. Vision Research 8:1483–99. [rSG]
- Wallach, H. & Adams, P. A. (1954) Binocular rivalry of achromatic colors. American Journal of Psychology 67:513-16. [taSG]
- Watson, A. S. (1978) A Riemann geometric explanation of the visual illusions and figural after-effects. In: Formal theories of visual perception, eds. E. C. J. Leeuwenberg & H. F. T. M. Buffart. New York: Wiley. [taSG]
- Weisstein, N. (1980) The joy of Fourier analysis. In: Visual coding and adaptability, ed. C. S. Harris. Hillsdale, N.J.: Erlbaum. [rSG]
- Weisstein, N. & Harris, C. S. (1980) Masking and the unmasking of distributed representations in the visual system. In: Visual coding and adaptability, ed. C. S. Harris. Hillsdale, N.J.: Erlbaum. [rSG]
- Weisstein, N., Harris, C. S., Berbaum, K., Tangney, J. & Williams. A. (1977) Contrast reduction by small localized stimuli: Extensive spatial spread of above-threshold orientation-selective masking. Vision Research 17:341-50. [rSC]
- Weisstein, N. & Maguire, W. (1978) Computing the next step: Psychophysical measures of representation and interpretation. In: Computer vision systems, eds. E. Riseman & A. Hanson. New York: Academic Press. [rSG]
- Weisstein, N., Maguire, W. & Berbaum, K. (1976) Visual phantoms produced by moving subjective contours generate a motion aftereffect. Bulletin of the Psychonomic Society 8:240 (abstract). [rSG]
- (1977) A phantom-motion aftereffect. Science 198:955-98. [JMF, taSC] Weisstein, N., Maguire, W. & Williams, M. C. (1978) Moving phantom
- contours and the phantom-motion aftereffect vary with perceived depth. Bulletin of the Psychonomic Society 12:248 (abstract). [rSG]
- Weisstein, N., Matthews, M. & Berbaum, K. (1974) Illusory contours can mask real contours. Bulletin of the Psychonomic Society 4:266 (abstract). [rSG]
- Werblin, F. S. (1971) Adaptation in a vertebrate retina: Intracellular recordings in Necturus. Journal of Neurophysiology 34:228-41. [tarSG]
- Werner, H. (1937) Dynamics in binocular depth perception. Psychological Monograph (whole no. 218). [taSG]
- Wilson, H. R. (1980) A transducer function for threshold and suprathreshold human vision. Biological Cybernetics 38:171-78. [JMF]
- Wilson, H. R. & Bergen, J. R. (1979) A four-mechanism model for spatial vision. Vision Research 19:19-32. [tarSG]
- Wilson, H. R. & Cowan, J. D. (1972) Excitatory and inhibitory interactions in

## References/Grossberg: Quantized geometry of visual space

luminance increments. Journal of the Optical Society of America 60:382-89. [DL]

Newell, A. (1980) Harpy, production systems, and human cognition. In: Perception and production of fluent speech, ed. R. Cole. Hillsdale, N.J.: Erlbaum. [taSG]

- O'Brien, V. (1958) Contour perception, illusion and reality. Journal of the Optical Society of America 48:112-19. [SC, taSG]
- Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957) The measurement of meaning. Urbana: University of Illinois. [taSG]
- Poggio, T. (1980) Neurons sensitive to random-dot stereograms in areas 17 and 18 of rhesus monkey. Society for Neuroscience Abstracts (November) 6. [BJ]
- Poggio, T. (1982) Trigger features or Fourier analysis in early vision: A new point of view. In: *The recognition of pattern and form, lecture notes in biomathematics*, ed. D. Albrecht. New York: Springer, 44:88-99. [WELG]
- Pollen, D. A. & Ronner, S. F. (1981) Phase relationships between adjacent simple cells in the visual cortex. Science 212:1409-11. [rSG]
   (1982) Spatial computation performed by simple and complex cells in the
- visual cortex of the cat. Vision Research 22:101-18. [rSG]
   Pulliam, K. (1981) Spatial frequency analysis of three-dimensional vision. Proceedings of the Society of Photo-Optical Instrumentation Engineers 303:71-77. [JMF, rSG]
- Raaijmakers, J. C. W. & Shiffrin, R. M. (1981) Search of associative memory. Psychological Review 88:93-134. [taSC]
- Rall, W. (1977) Core conductor theory and cable properties of neurons. In: Handbook of physiology: The nervous system, vol. I, Part I, ed. E. R. Kandel, pp. 39-97. Bethesda, Md.: American Physiological Society. [rSG]
- Rashevsky, N. (1968) Mathematical biophysics. Chicago: University of Chicago Press. [TT]
- Ratliff, F. (1965) Mach bands: Quantitative studies on neural networks in the retina. New York: Holden-Day. [BB, taSG]
- Rauschecker, J. P. J., Campbell, F. W. & Atkinson, J. (1973) Colour opponent neurones in the human visual system. *Nature* 245:42-45. [taSG]
- Restle, F. (1971) Mathematical models in psychology. Baltimore: Penguin Books. [taSC]
- Richards, W. (1975) Visual space perception. In: Handbook of perception, Vol. 5: Seeing, eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [taSC]
- Richards, W. & Marr, D. (1981) Computational algorithms for visual processing. M.I.T. Artificial Intelligence Lab. [taSG]
- Richards, W. & Miller, J. F., Jr. (1971) The corridor illusion. Perception and Psychophysics 9:421-23. [taSG]
- Richter, J. & Ullman, S. (1982) A model for the temporal organization of Xand Y-type receptive fields in the primate retina. *Biological Cybernetics* 43:127-45. [rSG, WELG]
- Robson, J. C. (1975) Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. In: *Handbook of perception* (Vol. 5), eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [taSG]
- Robson, J. G. & Graham, N. (1981) Probability summation and regional variation in contrast sensitivity across the visual field. Vision Research 21:409-18. [taSG]
- Rock, I. (1977) In defense of unconscious inference. In: Stability and constancy in visual perception, ed. W. Epstein. New York: Wiley. [HB]
- Rodieck, R. W. & Stone, J. (1965) Analysis of receptive fields of cat retinal ganglion cells. Journal of Neurophysiology 28:833-49. [taSG]
- Rozental, S., ed. (1967) Niels Bohr. New York: Wiley. [rSG]
- Rushton, W. A. (1965) Visual adaptation. The Ferrier Lecture, 1962. Proceedings of the Royal Society of London (B) 162:20-46. [WJF]
- Sakata, H. (1981) Mechanism of Craik-O'Brien effect. Vision Research 21:693-99. [rSG]
- Schriever, W. (1925) Experimentelle studien über stereokopische sehen. Zeitschrift fuer Psychologie 96:113-70. [SEP]
- Schrödinger, E. Müller-Pouillets Lehrbuch der Physik 11. Auflage, Zweiter Band. Braunschweig. [HB]
- Schwartz, E. L. (1980) Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. Vision Research 20:645-69. [taSG]
- Sekuler, R. (1975) Visual motion perception. In: Handbook of perception, Vol. 5, eds. E. C. Carterette & M. P. Friedman. New York: Academic Press. [rSG]
- Shepard, R. N. (1980) Multidimensional scaling, tree-fitting, and clustering. Science 210:390-98. [taSG]
- Shepard, R. N. & Chipman, S. (1970) Second-order isomorphism of internal representations: Shapes of states. Cognitive Psychology 1:1-17. [BB]

Shepard, R. N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. Science 171:701-03. [DHF]

- Shepherd, G. M. (1972) Synaptic organization of the mammalian olfactory bulb. Physiological Review 52:864-917. [WJF]
- Shipley, T. (1965) Visual contours in homogeneous space. Science 150:348-50. [taSG]
- Singer, W. (1982) The role of attention in developmental plasticity. Human Neurobiology 1:41-43. [taSG]
- Smith, A. T. & Over, R. (1979) Motion aftereffect with subjective contours. Perception and Psychophysics 25:95-98. [taSG]
- Sperling, G. (1970) Binocular vision: A physical and a neural theory. American Journal of Psychology 83:461-534. [GJD, taSG, BJ]
- (1981) Mathematical models of binocular vision. In: Mathematical psychology and psychophysiology, ed. S. Grossberg. Providence, R.I.: American Mathematical Society. [taSG]
- Sperling, C. & Sondhi, M. M. (1968) Model for visual luminance discrimination and flicker detection. Journal of the Optical Society of America 58:1133-45. [taSG, DSL]
- Stevens, S. S. (1959) The quantification of sensation. Daedalus 88:606-21. [taSG]
- Stromeyer, C. F. III & Mansfield, R. J. W. (1970) Colored after-effects produced with moving edges. *Perception and Psychophysics* 7:108-14. [rSG]
- Swets, J. A. (1961) Is there a sensory threshold? Science 134:168-77. [DL]
- Tschermak-Seysenegg, A. von (1952) Introduction to physiological optics. P. Boeder, trans. Springfield, Ill.: C. C. Thomas. [taSC]
- Tynan, P. & Sekuler, R. (1975) Moving visual phantom: A new contour completion effect. Science 188:951-52. [JMF, taSG]
- Uttal, W. (1973) The psychobiology of sensory coding. New York: Harper and Row. [BB]
- van den Brink G. & Keemink, C. J. (1976) Luminance gradients and edge effects. Vision Research 16:155-59. [HB]
- van Nes, F. L. (1968) Experimental studies in spatio-temporal contrast transfer by the human eye. Utrecht: University. [HB]
- van Nes, F. L. & Bouman, M. A. (1965) The effects of wavelength and luminance on visual modulation transfer. Excerpta Medica International Congress Series 125:183-92. [HB]
- van Tuijl, H. & Leeuwenberg, E. (1979) Neon color spreading and structural information measures. *Perception and Psychophysics* 25:269-84. [HB]
- Von Békésy, G. (1968) Mach- and Hering-type lateral inhibition in vision. Vision Research 8:1483-99. [rSG]
- Wallach, H. & Adams, P. A. (1954) Binocular rivalry of achromatic colors. American Journal of Psychology 67:513-16. [taSG]
- Watson, A. S. (1978) A Riemann geometric explanation of the visual illusions and figural after-effects. In: Formal theories of visual perception, eds. E.
- C. J. Leeuwenberg & H. F. T. M. Buffart. New York: Wiley. [taSG] Weisstein, N. (1980) The joy of Fourier analysis. In: Visual coding and adaptability, ed. C. S. Harris. Hillsdale, N.J.: Erlbaum. [rSG]
- Weisstein, N. & Harris, C. S. (1980) Masking and the unmarking of distributed representations in the visual system. In: Visual coding and adaptability, ed. C. S. Harris, Hillsdale, N.I.: Erlbaum. [rSG]
- Weisstein, N., Harris, C. S., Berbaum, K., Tangney, J. & Williams. A. (1977) Contrast reduction by small localized stimuli: Extensive spatial spread of above-threshold orientation-selective masking. Vision Research 17:341-50. [rSG]
- Weisstein, N. & Maguire, W. (1978) Computing the next step: Psychophysical measures of representation and interpretation. In: Computer vision systems, eds. E. Riseman & A. Hanson. New York: Academic Press. [rSG]
- Weisstein, N., Maguire, W. & Berbaum, K. (1976) Visual phantoms produced by moving subjective contours generate a motion aftereffect. Bulletin of the Psychonomic Society 8:240 (abstract). [rSG]
- (1977) A phantom-motion aftereffect. Science 198:955-98. [JMF, taSG]
- Weisstein, N., Maguire, W. & Williams, M. C. (1978) Moving phantom contours and the phantom-motion aftereffect vary with perceived depth. Bulletin of the Psychonomic Society 12:248 (abstract). [rSG]
- Weisstein, N., Matthews, M. & Berbaum, K. (1974) Illusory contours can mask real contours. Bulletin of the Psychonomic Society 4:266 (abstract). [rSG]
- Werblin, F. S. (1971) Adaptation in a vertebrate retina: Intracellular recordings in Necturus. Journal of Neurophysiology 34:228-41. [tarSG]
- Werner, H. (1937) Dynamics in binocular depth perception. Psychological Monograph (whole no. 218). [taSG]
- Wilson, H. R. (1980) A transducer function for threshold and suprathreshold human vision. Biological Cybernetics 38:171-78. [JMF]
- Wilson, H. R. & Bergen, J. R. (1979) A four-mechanism model for spatial vision. Vision Research 19:19-32. [tarSG]
- Wilson, H. R. & Cowan, J. D. (1972) Excitatory and inhibitory interactions in

## References/Grossberg: Quantized geometry of visual space

localized populations of model neurons. *Biophysical Journal* 12:1-24. [DBL]

- (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13:55-80. [DSL]
- Winston, P. H. (1979) MIT progress in understanding images. Proceedings: Image Understanding Workshop, Palo Alto, California, pp. 25-36. [WELC]
- Wyatt, H. J. & Daw, N. W. (1975) Directionally sensitive ganglion cells in the rabbit retina: Specificity for stimulus direction, size, and speed. *Journal of Neurophysiology* 38:613–26. [rSG]
- Zucker, S. W. (1980) Motion and the Mueller-Lyer illusion. McGill Department of Electrical Engineering, Technical Report 80-2R. [WELG]