Sperling, George. Mathematical models of binocular vision. In S. Grossberg (Ed.), Mathematical Psychology and Psychophysiology. Providence, Rhode Island: Society of Industrial and Applied Mathematics Association Proceedings, 1981, 13, 281-300.

SIAM
AMS
proceedings

volume 13

Mathematical Psychology and Psychophysiology

PROCEEDINGS OF THE SYMPOSIUM IN APPLIED MATHEMATICS
OF THE AMERICAN MATHEMATICAL SOCIETY
AND THE SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS

HELD IN PHILADELPHIA
APRIL 15–16, 1980

EDITED BY

STEPHEN GROSSBERG

AMERICAN MATHEMATICAL SOCIETY

PROVIDENCE · RHODE ISLAND

Codistributed by LAWRENCE ERLBAUM ASSOCIATES, INC., PUBLISHERS
HILLSDALE, NEW JERSEY and LONDON, ENGLAND

SIAM-AMS Proceedings
Volume 13
1981

Mathematical Models of Binocular Vision[1]

GEORGE SPERLING

Overview. When I accepted the invitation to participate in this symposium, I promised to discuss some representative models of perception. It soon became apparent that the mathematical models currently being proposed are so numerous and so complex that it would require volumes to do justice to even one subspeciality, such as color vision or binocular vision. For example, simple linear matrix operations sufficed for the theory of color mixture for 100 years (Wyszecki and Stiles [**1967**]); today's more comprehensive theories of color perception invoke additional nonlinear operations (e.g. Pugh and Mollon [**1979**]).[2] There are new developments in theories such as factor analysis, multidimensional scaling, and cluster analysis which have been used to describe the mapping of physical stimuli (such as simple color patches but also much more complex stimuli) into psychologically significant space (Shepard [**1980**]). Measurement theory has evolved as a branch of mathematics to describe the mapping of physical stimulus dimensions (most often intensity) into psychological dimensions (Krantz, Luce, Suppes, and Tversky [**1971**], Roberts [**1979**]).

In sensory psychology, particularly in the study of vision and of hearing, linear equations have been used with considerable success to describe the receptors, i.e., the optical properties of the lens of the eye (Fry [**1955**], Krauskopf [**1962**]) and the transducer properties of the outer, middle, and inner ear (basilar membrane, Allen [**1977a**], [**1977b**]). Sine wave stimulus patterns and linear theory are widely used to describe the first order psychological properties of sensory systems (Licklider [**1961**], Graham [**1981**]). But in these domains, familiar engineering methods have been all but exhausted; current theories deal with complex, nonlinear sensory mechanisms (Schroeder [**1975**], Victor, Shapley, and Knight [**1977**]). In fact, process models (such as models of the presumed

1980 *Mathematics Subject Classification.* Primary 92A25; Secondary 58C28.

[1]For reprints write to Professor G. Sperling, Department of Psychology, New York University, 6 Washington Place, New York, N. Y. 10003.

[2]It is impractical to give here complete reference lists for the various subjects mentioned in the text; only a few representative sources are cited.

neural processes that underly perception) can quickly become mathematically intractable; only a few workers have reduced the corresponding sets of nonlinear differential equations to something understandable (Sperling and Sondhi [**1968**], Grossberg [**1972**], [**1973**], Luce and Green [**1972**], [**1974**]).

Visual illusions that result from viewing "impossible" figures are better understood via topological analysis of the stimuli (Cowan [**1974**]); errors in the perception of a sequence in a string of characters are better understood by means of combinatorial algebraics (Sperling and Melchner [**1976**]); Riemannian geometry is used to describe the binocular perception of space (Luneberg [**1947**]); and Markov, random walk, and diffusion models occur in the study of reaction time, decision making, and other processes (Norman [**1972**], Link and Heath [**1975**]).

In the study of selective attention (e.g., to one part of the visual field or to one kind of character symbol in a visual search task) new concepts such as the attention operating characteristics have developed (Sperling and Melchner [**1978**], Navon and Gopher [**1979**]) that are intimately related to the receiver operating characteristics of signal detection theory (Green and Swets [**1966**], Egan [**1975**]), as well as to important concepts of economics. Relevant mathematical specialties are utility theory and linear programming.

In a study of visual motion perception in ambiguous displays, functional equations was the mathematical tool used to develop a process model (Burt and Sperling [**1981**]). 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 (Sperling [**1978**]), 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 perceptual problems I will be concerned with in this paper involve the dynamics of resolving perceptual ambiguities in general; examples are given from binocular vision and depth perception. The relevant areas of mathematics are nonlinear differential equations and catastrophe (or potential) theory. I have chosen these examples because the possibility of further mathematical development is most appealing.

Path-dependence in binocular vision: The phenomena. We begin with the general problem of path-dependent perceptual states: how do we understand the cases where our perception of the present stimulus depends on our perception of the immediately preceding stimuli?[3]

Multistability of vergence. A classical example of this kind of phenomenon is described by Helmholtz [**1924**, p. 58]. Helmholtz viewed a binocular stereogram, that is, two identical or nearly identical images, so that each eye saw only its

[3]Much of the subsequent discussion is based on Sperling [**1970**], which the reader should consult for details, additional references, and interaction models.

corresponding image (Figure 1). First, he achieved vergence-fusion of the sterogram, that is, he arranged for the left and right half-images to fall on corresponding retinal points, and they were perceived as a single, unitary image (Figure 1(b)). At this point, the stereo images were arrayed so that the lines of sight of his two eyes were parallel. Then, Helmholtz began to increase slowly the separation in physical space between the two half-images so that, in order to maintain fusion, his eyes were forced to diverge. He found that by pulling the images apart in this way he was able to cause the lines of sight of his eyes to diverge by as much as eight degrees to follow the images. When the divergence angle required to maintain fusion increased beyond eight degrees, Helmholtz lost fusion; that is, his eyes returned to a parallel position and he saw two partially overlapping stereo half-fields (Figure 1(c)). To restore fusion, Helmholtz had to bring the stereo half-images much closer together than the point at which fusion broke.

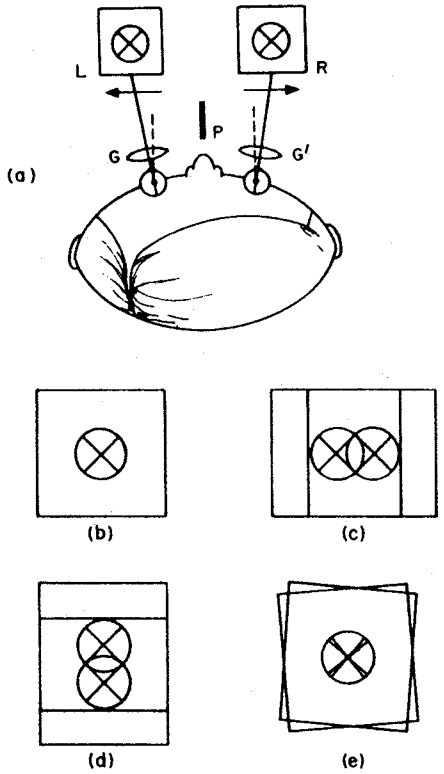

FIGURE 1.(a) A sterogram showing the left L and right R half-images viewed by the left and right eyes, respectively. The lenses G, G' enable the observer to focus at the short viewing distance. The partition P isolates the views of the two eyes. The arrows indicate the direction in which the stimuli are moved in Helmholtz's horizontal vergence demonstration. The broken lines indicate parallel lines of sight; the solid lines indicate the actual lines of sight when the observer is able to verge accurately at the angle of divergence shown. (b) Representation of the cyclopean perception when the eyes are correctly verged. The perception after vergence fails for (c) horizontal vergence, (d) vertical vergence, and (e) torsional vergence.

This example illustrates path-dependence for a particular stereogram separation, say, eight degrees. If the eyes were diverged prior to the eight degree stereogram–perhaps because the previous stereogram had a seven degree separation and they were able to fuse it–the eyes remained fused. If the immediately preceding stereogram had a zero or a 12 degree separation, the eyes would not fuse. Whether the eyes were fused or not depended on their history. Thus, we have two stable states, fused or not fused, in response to the same present stimulus, and the particular state that is achieved depends on the sequence of the preceding stimuli.

This example exploits the most familiar of the eyes' vergence systems–the system that controls the convergence and divergence of the lines of sight of the two eyes. Many people have voluntary control over lateral vergence, particularly convergence, and that control can considerably complicate actual demonstrations and theory. However, the demonstration works equally well with vertical vergence which, so far, has not been voluntarily controlled. Helmholtz found that when one stereo half-image was raised and the other lowered, the eyes would diverge vertically–one up, the other down–to maintain fusion. The angular range of vertical vergence was about the same vertically as the range of horizontal divergence. Similarly, when the stereo half-images are rotated around their centers in opposite directions, the eyes will follow these rotations with corresponding torsional rotations around their axes.

Multistability of accommodation (focusing). The accommodation (focusing) system of the eyes provides perhaps the simplest demonstration of path-dependence.[4] Take a piece of paper that has some printing on it and make a hole about $\frac{1}{2}$ to 1 cm diameter in the center. Close one eye and bring the hole as close to the other as possible, all the while maintaining the paper in sharp focus. Direct the line of sight through the middle of the hole at a distant object while maintaining the paper in focus. The hole should appear to be filled with an extremely blurred image while the surrounding paper is sharp (Figure 2). Now move the paper away while maintaining the line of sight. The eye will focus on the distant object, which now appears clearly. Without changing the line of sight, reintroduce the paper to the position it occupied just previously. It should be possible to now look through the hole and to see the distant object in focus and the surrounding paper blurred. (Younger subjects have a much better demonstration because of their greater range of accommodation. Once the accommodative range becomes too small, voluntary–rather than a stimulus–control of accommodation can predominate.) However, by adjusting the parameters of the demonstration–aperture size, distance, contrast, etc.–bistability can usually be restored. The demonstrations show path-dependence in accommodation. Precisely the same visual stimulus (distant view through a near aperture) can lead to two states of accommodation: near focus or far focus, depending on

[4]The demonstration of path-dependence in accommodation was reported by Sperling [1970].

the previous stimuli, i.e., on where the eye happened to be focused before looking through the hole.

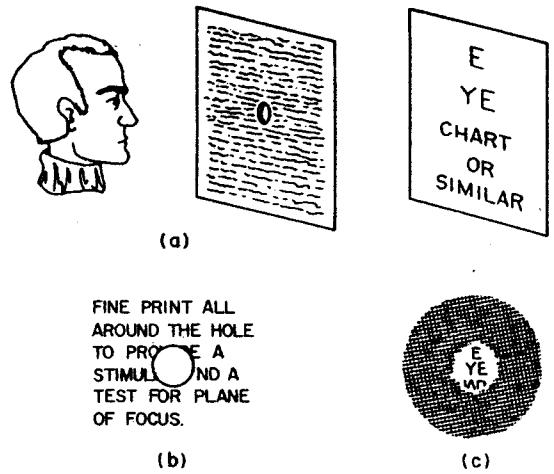

FIGURE 2. (a) Arrangement for observing two states of accommodation (focusing of the eye). The observer looks through the aperture with one eye at a distant object. (b) The view, accommodated on the aperature. (c) The view, accommodated on the distant object.

Multistability of fusion. Perhaps the most interesting instance of path-dependence in this class of visual phenomena occurs in perceptual fusion. To demonstrate it, the eyes view a stereogram through a complex optical system that cancels the effect of eye movements; that is, the images are optically stabilized on their respective retinas irrespective of where the eyes happen to point. Now, let the images be moved apart on the retinas just as, in the first example, they were moved apart in physical space. Eye movements cannot compensate for the image slip; nevertheless, perceptual fusion can be maintained for an appreciable image separation before fusion suddenly "breaks" and the two half-images appear to be partially overlapping instead of a fused unitary image (Diner **[1978]**, Fender and Julesz **[1967]**). Again, we have two perceptual states–fused and unfused–in response to the same physical stimulus on the retina.

Path-dependence, multiple stable states, hysteresis. Insofar as we consider only deterministic theories, path-dependence and multiple stable states are equivalent. Different states can be reached only via different paths, and path-dependence implies that the same stimulus will lead to different responses (multiple stable states) depending on the path taken. Hysteresis refers to a particular kind of path-dependence, namely, the tendency of a state to perseverate even after inducing conditions have changed to be more appropriate for some other state. This is, in fact, the kind of path-dependence that is most frequently observed, and the kind that has been the subject of all the examples.

Path-dependence: Theory. To begin to describe these phenomena mathematically, we start with one particular example–vergence. Let the vergence angle that would be required to place the two half-images of a stereogram onto exactly

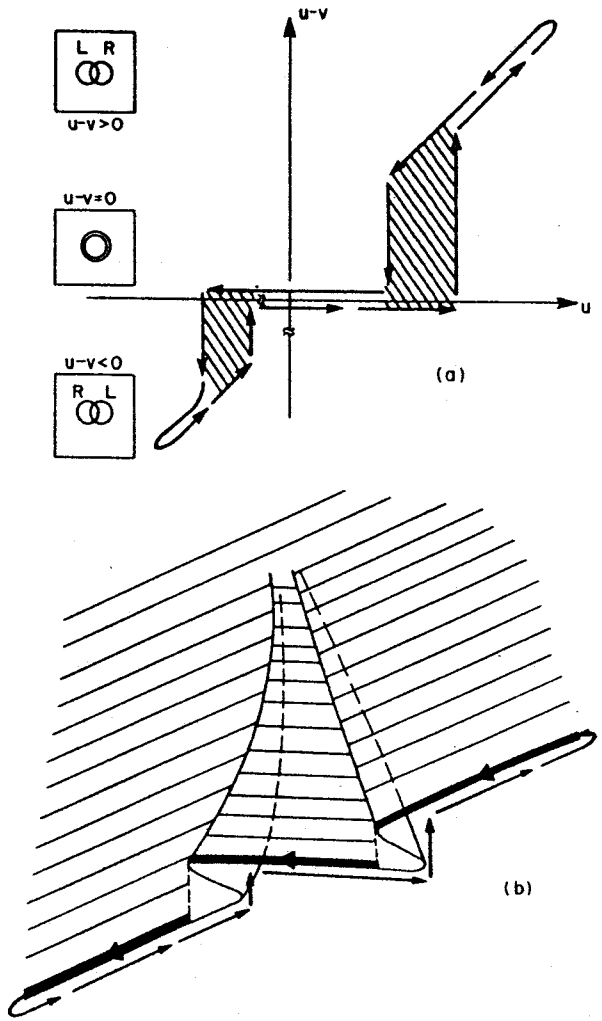

FIGURE 3. Graphical representation of Helmholtz's vergence experiment. (a) Abscissa u represents the stimulus disparity between images to the left and right eyes; v is the eyes' vergence angle, and the ordinate is $u - v$ the perceived disparity, which is shown, as perceived in the inserts at the extreme left. In the inserts, L and R indicate left-eye and right-eye stimuli, respectively. The arrows indicate the sequence of successive events (stimuli presented and the resulting perceived disparity) in an experiment. Right part of graph represents divergence, the left part convergence; the break in the axes indicates that the breakaway vergence angles are different in the two directions. The shaded area represents hysteresis–the difference between perceived disparity depending upon the path taken. (b) A catastrophy theory representation of the experiment in (a). The arrows represent the path of an experiment; heavy arrows, the stimulus converging, and light arrows, the stimulus diverging. See text for details.

corresponding retinal points be u. Let the actual vergence angle between the eyes be v. Then the (perceived) disparity is $u - v$. Helmholtz's experiment can be represented in a graph of $u - v$ versus v, as in Figure 3. We start with the eyes parallel ($v = 0$) and the stimulus positioned on corresponding retinal points

($u - v = 0$), represented at the point (0, 0) in Figure 3(a). As u is increased by pulling the stereo half-images apart (vertically or horizontally), the eyes follow at first: thus the perceived disparity $u - v$ is 0. When the critical angle is reached, the eyes fall back to their parallel position and the perceived disparity equals the actual disparity $u - v$ between the two stereo half-images. For all larger values of u, the perceived disparity $u - v$ remains equal to u since the eyes remain parallel ($v = 0$). When the angle between the stereo half-images is again reduced, another critical angle, smaller than the breakaway angle, is reached where vergence-fusion is again restored. These events are represented by following the arrows around one half of Figure 3(a). As the separation between the stereo half-images is reduced further and becomes negative, a similar sequence of events is produced in the other half of the graph. Thus, if the initial separation (right side of Figure 3(a)) represents divergence, the left side represents convergence. If the right side represents displacement of the left image above the right image, the other side represents the opposite displacement.

Catastrophe theory description. Figure 3(b) represents a catastrophe theory (Thom [**1974**], Susman and Zahler [**1978**]) description of the events of Figure 3(a). The path of Figure 3(a) is traced on the surface of Figure 3(b). When a fold in the surface is reached, a catastrophe, that is, a jump to another level occurs. In this representation, the third dimension (depth into the page) is not essential. It could have an interpretation as the level of illumination upon or as the contrast of the stimulus, with the unfolded rear section of sheet representing subthreshold amounts of illumination or contrast.

Potential theory description. Elegant though catastrophe theory may be, this particular representation of a path-dependent phenomenon within that framework is merely descriptive; it does not yield any insight into the nature of the underlying processes. A theory for the class of phenomena in question is needed. The theory proposed here is a kind of potential theory in which an electrical potential field governs the movement of a charged particle. I will use a gravimetric analog of potential theory, in which a marble rolls on a bumpy surface–the energy surface–under the influence of gravity, because this analogy is particularly concrete and intuitively accessible.[5]

Two kinds of factors designated as internal and external factors control the shape of the energy surface. For example, the basic surface governing horizontal vergence is bowl-shaped, the shape being determined by internal factors. The marble tends to roll to the center of the bowl, representing the tendency of the eyes to verge to a neutral position in the absence of an external stimulus. Specifically, let the displacement vergence energy surface $g(v)$ be a concave-up function of v. Let the movement of the marble along this surface be governed by

[5]In a potential field $e'(x)$, the force acting on a particle is proportional to e'. An energy field $e(x)$ is the integral of a potential field so that the force is proportional to $e' = \partial e(x)/\partial x$. In the gravimetric analog, force $f = -ke'/(1 + e'^2)$, an irrelevant complication, which becomes negligible as $e' \to 0$. For appropriately scaled units, the gravimetric and potential models are equivalent.

a simple differential equation:

$$\frac{d^2v}{dt^2} + k\frac{dv}{dt} = \frac{-dg(v)}{dv}, \qquad k > 0. \tag{1}$$

For concreteness, think of the marble as rolling in a smooth bowl filled with salad oil. The dynamic properties of motion are given by equation (1), but it is evident that so long as there is nonzero friction in the medium, the marble must eventually come to rest at the equilibrium point, the bottom of the bowl. The horizontal projection of the marble's position represents the instantaneous vergence position of the eyes. For convenience, let the value of v at the minimum of $g(v)$ be zero; this point represents the neutral position of eyes in the absence of any visual stimulus, i.e., in the dark. The motion of the marble when it is placed up on the side of the bowl and released represents the vergence movements of the eyes when they are verged on some particular stimulus, say, a book, and the light is suddenly turned off. In the dark, the eyes return to their neutral position (Figure 4).

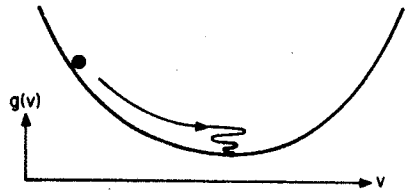

FIGURE 4. Vergence displacement energy $g(v)$ as a function of vergence angle v. The eye's vergence position is represented by the projection onto the abscissa of a marble rolling on the surface. This surface governs the eyes' return to their neutral vergence positon when they are somehow displaced from it; a typical path is indicated.

To show how external factors–the external stimuli impinging on the eyes–can introduce new equilibrium states, we need some definitions. Let the stimulus to the left retina be described as a luminance distribution $l_L(x, y)$ which is a function of x and y, the horizontal and vertical coordinates expressed in units of visual angle. Let the stimulus to the right eye be $l_R(x, y)$. Define $h(v)$, the vergence image-disparity energy as a function of vergence angle v,

$$h(v) = -\int\int_{x,y}\left[l_L\left(x - \frac{v}{2}, y\right) - l_R\left(x + \frac{v}{2}, y\right)\right]^2 dx\, dy. \tag{2}$$

Except for the minus sign, the vergence image-disparity is essentially the unnormalized cross-correlation between the images to the left and right eyes; h has a minus sign so that it has a minimum when the correlation has a maximum (Figure 5). Vergence image-disparity energy $h(v)$ expresses how well the images on the two retinas would match if the eyes were to assume a vergence angle of exactly v.

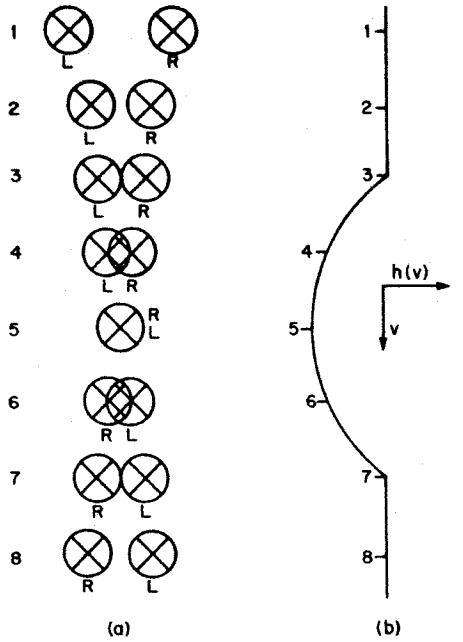

FIGURE 5. Illustration of the computation of vergence image-disparity energy. (a) Luminance distributions l_L and l_R on the left and right retinas as a function of vergence shift v for the stimuli illustrated in row 4. (b) Vergence image-disparity energy $h(v)$ for the stimulus in row 4. The numbers on the graph of $h(v)$ indicate the corresponding row in (a). The axes are placed at 0, 0; that is, v is zero in row 4. The eyes would have to diverge slightly from their initial position at 5 to achieve vergence-fusion of stimulus 4, as indicated by the corresponding minimum in $h(v)$. The figure can be rotated 90 degrees to place the left side on the bottom; in this case, it represents vertical vergence.

The internal and external factors controlling vergence are added to obtain the net vergence energy:

$$e(v; l_L, l_R) = g(v) + h(v; l_L, l_R). \tag{3}$$

Figure 6(a) shows an example of a unistable vergence stimulus. That is, a stereogram is viewed in which the separation of the two half-images is so slight that the eyes inevitably fuse on it. The vergence energy surface corresponding to it has a single minimum.

Now let the stereo half-images be pulled apart. This has the effect of moving the minimum of g and the minimum of h further apart. Sooner or later a point is reached where the two minima are so far apart that when they are added to form $e(v)$, it has two minima (Figure 6(b)). These minima correspond to the two stable equilibrium points for such stereograms. The eyes achieve the off-center equilibrium point when the stereo half-images are drawn apart slowly, all the while maintaining vergence-fusion. This corresponds to drawing the minima of $e(v)$ apart slowly, all the while keeping the marble in the bottom of the moving minimum. If the marble were initially at $v = 0$ when the energy surface was suddenly formed to have two minima, it would remain in the minimum at zero. Thus, there are two equilibrium points for the marble, depending on the sequence of events leading up to the two-minimum surface.

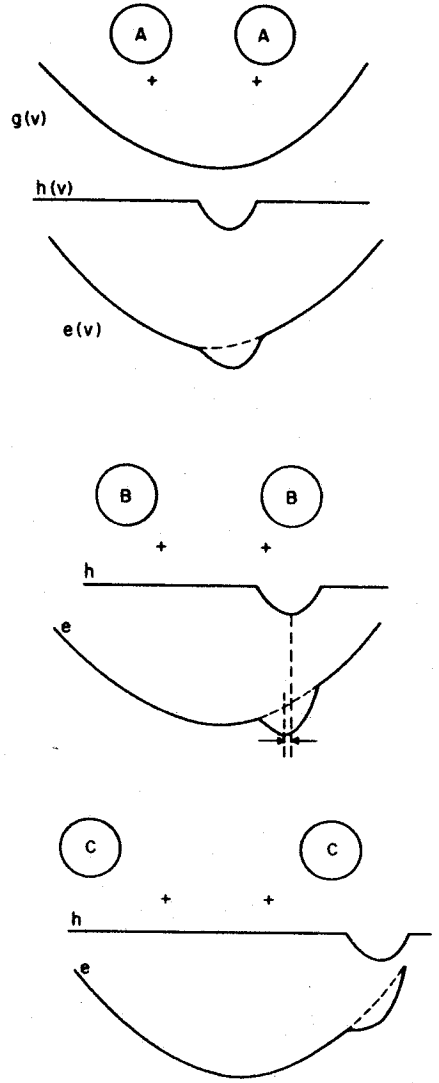

FIGURE 6. Potential theory model. (a) Two halves of a stereogram (A, A), fixation points are indicated by +. Vergence displacement energy $g(v)$ adds to vergence image-disparity energy $h(v)$ to produce (net) vergence energy $e(v)$, which has a single minimum slightly displaced in the direction of divergence. (b) Two halves of a stereogram requiring a greater divergence than (a) to achieve fusion. The corresponding vergence energy surface has two minima. The arrows indicate "vergence disparity": the displacement of the minimum of $e(v)$ [which determines the actual vergence position] away from the minimum of $h(v)$ [which minimizes L/R image disparity]. (c) A stereogram which is at the limit of vergence fusion. The $e(v)$ surface has only one minimum corresponding to the neutral vergence position.

Figure 6(c) illustrates what happens when the separation between the stereo half-images is increased further. The minimum produced by $h(v)$ eventually becomes too shallow to hold the marble; beyond this critical separation there is again only one stable equilibrium point.

This qualitative model of multistable path-dependent states is in agreement with the major experimental phenomena (multiple stable states, hysteresis, etc.) described in the examples and summarized in Figure 3. The minima of the energy function $\partial e/\partial v = 0$ correspond to the points on the catastrophe surface in Figure 3(b). The advantage of the potential theory representation is that it provides a mechanism for the system dynamics, not merely a representation of the equilibrium states, which is all that is provided by a catastrophe theory representation.

Vergence disparity. The potential theory representation also provides an explanation for some second order phenomena, such as "vergence disparity". When the eyes are induced to verge at a position other than the neutral point, they do not move away far enough from the neutral point to produce perfect left/right image registration. There is a disparity between the optimum and the actual vergence position (Ogle [1950], Fincham and Walton [1957]). In the theory, this "vergence disparity" is represented by the fact that the minimum of $e(v)$ does not lie exactly over the minimum of $h(v)$; because of the concaveness of $g(v)$, the minimum of $e(v)$ is displaced towards the intrinsic neutral point (minimum of $g(v)$). A closely related observation is that detailed, high contrast, large area stimuli are most effective in inducing the eyes to depart from their neutral position and in minimizing "vergence disparity". According to the computation of $h(v)$ in the theory, such stimuli have larger, steeper minima and therefore these minima are less deviated by the addition of $g(v)$ to produce $e(v)$ (reducing "vergence disparity") and their steeper sides preserve minima further from the neutral point.

Optimization theory. The vergence theory can also be thought of as an error theory. There is an error function $h(v)$ that represents the error in image registration as a function of vergence angle v, and an "error" function $g(v)$ that represents the displacement v of the eyes from their preferred position as an error. The equilibrium vergence position is the one that minimizes the summed errors.

A theory for accommodation can be generated according to principles that are quite analogous to those of the theory for vergence. There is an error function [an accommodation displacement function $g(v)$] that represents the deviation v of accommodation from its preferred position as an error. And, there is an error function [image defocus function, $h(v)$] that represents image blur. The equilibrium value of accommodation minimizes the sum of these two "errors". The main phenomena and conclusions about vergence can be transposed directly to accommodation.

When we consider accommodation and vergence together, we observe that where the eyes are induced to accommodate determines where they prefer to verge and vice versa. The interaction between these two systems produces very interesting phenomena and theories to account for them (Sperling [1970]). Complex though such theories may be, they are elementary in another sense.

The eyes can be verged at only one angle at one time. They can be focused only at one distance. To reproduce the path-dependent phenomena of vergence and accommodation, a neural model need compute only one value, $\partial h(v_i)/\partial v$ [at $v = v_i$, the eyes' current (instantaneous) vergence or accommodation position] and add this value to a reference signal $\partial g(v)/\partial v$ [at $v = v_i$] to produce the motor control signal. But in a neural system for perceptual fusion, fusion image disparity must be calculated for a range of possible z-depth values (z_{min}, z_{max}) because for any fixed vergence position, perceptual fusion is possible over a range of depths. Furthermore, at each visual direction x, y, fusion at a different depth z is possible. This leads to an interaction between all points in the x, y plane, instead of an interaction between just two points as in the interaction of vergence and accommodation.

In fact, in one visual direction, only one fusion plane is observable.[6] This is a manifestation of a very general property of perceptual systems; they cannot simultaneously perform two antagonistic responses. Therefore, there is no reason to preserve information that would ultimately lead to antagonistic responses. Some of these kinds of decisions occur early in perceptual processing.

Two illustrative examples are figure/ground interaction (perceiving one part of a scene as figure precludes simultaneously perceiving it as ground) and binocular rivalry (seeing the image from one eye precludes simultaneously seeing a contradictory image from the other eye). Here, we will develop a particular theory of the neural processes that subserve binocular fusion, but the principles are widely applicable.

Neural theory.

Neural binocular field. The neural binocular field (NBF) is a representation inside the organism of the depth and space relations outside, literally, a reflection of the world in the brain of the beholder. The basic theory (Figure 7) usually is attributed to Kepler (Boring [1933], Kaufman [1965]). The x, y dimensions of the NBF represent, approximately, the dimensions perpendicular to the observer's visual axes; the z dimension represents depth in space. Signals from the left and right retinas enter the NBF at an "angle" so that as z increases, these signals are represented at successively greater x translations relative to each other. The middle z-plane of the field represents the horopter, the surface in space that projects to precisely corresponding retinal points.[7]

---

[6] Sperling [1970]. The possible counterexample to the rule of one depth to one point is a partially transparent screen that partially obscures an object behind it. It appears, however, that each surface (screen and object) is represented by a patchwork of many small areas in the visual field; the two kinds of patches do not intersect.

[7] In fact, in the vertical direction, the horopter is not even approximately perpendicular to the visual axes (Helmholtz [1924]). Points along the midline in space (horizon) project to the horizontal retinal midline, but points along the vertical midline in space actually project to a retinal line that intersects the vertical retinal midline at an angle of several degrees. However, these practical complications in the precise shape of the horopter are irrelevant for the general principles being considered here, particularly insofar as we restrict our attention to a horizontal meridian.

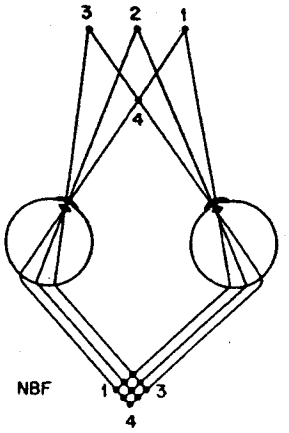

FIGURE 7. Keplerian projection theory. Objects in space, 1, 2, 3, 4, project through the retinas onto a neural binocular field, NBF, that mirrors the external spatial relationships. Intersections of signals from the two eyes (indicated by dots) represent possible matches. Three objects produce nine possible matches in the NBF; only three of these are "correct", and the remainder are "ghosts". For example, it is locally ambiguous whether the correspondence at point 4 in the NBF represents a real object 4 or a ghost match of 3 with 1.

Given a Keplerian NBF, the neural problem of stereoscopic depth perception reduces to three subproblems: (1) detecting correspondences (intersections) in the NBF; (2) determining the "distance" of such correspondences from the midline of the NBF; and (3) utilizing this depth information, i.e., using depth information by concatenating it with pattern information about an object to reach a determination of object shape or identity, or using depth information directly to control vergence movements of the eyes and other motor responses. The present summary focuses on the first of these problems– which is challenging enough. The other problems are treated in Sperling [**1970**] and elsewhere.

The most obvious complication in detecting correspondences in the NBF is that depth information is by its very nature *local* (that is, it is concerned with computing the depth plane of very small areas, x, y), and the information from any small area is inherently ambiguous. For example, if the neural inputs to the NBF carry information simply about whether a small area is lighter or darker than its immediate surround, then the "true" matches will represent a small fraction of all the matches, most of which will be accidental false matches–or ghost matches as they are called. If the inputs to the NBF carry highly specific information, then we are not likely to find the specific information coded in a small area, (so that area's depth cannot be computed). Furthermore, to convey depth in small areas with specialized codes would require an enormous proliferation of inputs so that there would be a reasonable chance of finding not only a highly localized input but also an appropriately-specialized one. The solution to this problem is to use information from the larger context to assist in the local decision–a solution with very interesting mathematical properties.

There is an even more basic principle the visual system seems to have adopted *en route* to its solution of the ghost matching problem. At any given x, y

coordinate (representing one visual direction in the cyclopean field of view) only one surface is perceived and only one depth magnitude is computed. What kind of neural organization in the NBF can accomplish this?

The mathematical neuron. In order that this section be self-sufficient, we begin with the description of a neuron, the basic element of the nervous system, with apologies to the readers for whom this is repetitious. A neuron receives excitatory inputs $[x_i(t)]$ and inhibitory inputs $[z_j(t)]$ from many other neurons, $[i]$, $[j]$. It combines all its inputs (all of which are non-negative) to produce a single non-negative output $y(t)$, which in turn becomes an excitatory and/or inhibitory input to other neurons. The exceptions are the first-order neurons, which receive direct sensory input, and the last neurons in the chain, which produce outputs to muscles, glands, etc.

From a functional point of view, a neuron is a mass of electrically conducting jelly surrounded by a thin membrane which is characterized by high resistance R and a capacitance C. An excitatory input to the neuron consists, essentially, of injections of electrical charge (in the form of charged ions) into the neuron, i.e., an excitatory input is a current $x(t)$.

In modeling a neuron, the excitatory input currents are assumed to add linearly. The accumulated charge decays (exponentially) through the RC membrane, and the output is proportional to the resulting voltage. Thus, to a first approximation, a real neuron is modeled by an RC integrator; its response to an excitatory impulse $\delta(t)$ at time $t = 0$ is simply $e^{-t/(RC)}$.

The most common response of neurons to inhibitory inputs is a proportional increase in membrane conductance $(1/R)$. Thus, inhibition is not a subtractive process–the opposite of excitation; inhibition is modeled as a *shunting* process (Furman [**1965**]) that divides excitation. A nonlinear differential equation to describe these relations is

$$\frac{d}{dt}y(t) + \frac{y(t)}{CR(t)} = X(t) \tag{4a}$$

where

$$R^{-1}(t) = R_0^{-1} + Z(t). \tag{4b}$$

Capital letters X, Z are used to designate the sums over all excitatory and all inhibitory inputs, respectively. This system was proposed and investigated by Sperling and Sondhi [**1968**].

Closer examination of typical real neurons implies complications to the seemingly simple equations (4). Some of these can be incorporated gracefully; others produce intractable complications. Saturation of excitatory inputs is incorporated into (4a) by replacing the right-hand term $x(t)$ with $x(t)(B - Y(t))$. [The resulting systems have been elegantly dealt with by Grossberg [**1973**]. The feedforward shunting inhibition proposed by Sperling and Sondhi [**1968**] yielded a formulation that is virtually equivalent to Grossberg's.] Other complications are: the equilibrium voltage to which inhibition drives the neuron interior generally is somewhat lower than the resting voltage for zero excitatory inputs.

Many neurons have a threshold $\varepsilon > 0$ below which there is no output. This is modeled by replacing $y(t)$ by $y'(t) = \max(0, y(t) - \varepsilon)$. The conducting interior of the neuron has significant resistance, so that–depending on the shape of the neuron–the exact placement of excitatory and inhibitory inputs can have important effects. This effect has been dealt with by computer simulations (Rall [1964]).

Functional model. The major psychophysical findings for which we seek a neural theory are: (1) multiple stable states/path-dependence in perceptual fusion (i.e., in fusion with stabilized retinal images); (2) only one perceivable depth at any one time in any one visual direction; (3) context dependency of fusion–a solution for the ghost-ambiguity problem. We propose operations that are carried out on the information in the NBF, and then a neural model that could accomplish these operations. Uniqueness is not claimed either for the operations or the model.

The functional model begins with the assumption that there is a structure like the NBF (Figure 7) receiving spatially labelled inputs, and that there are elements capable of detecting binocular correspondences. The outflow from the NBF is generated by the correspondence-detecting elements, and this outflow determines object perception, stereoscopic depth perception, vergence, etc. The binocular correspondence-detecting elements interact among themselves by competition and by cooperation to determine the outflow.

Competition, cooperation/competition. First consider a thin column in the NBF, occupying a small area *dxdy* and extending in the z direction. The column is assumed to function so as to never allow more than one z-level to be active at one time ("monoactivity", Sperling [1970]). This monoactivity function simply associates with each small visual area one and only one perceived depth. In the neural model, monoactivity is achieved by a purely competitive interaction. The second function is cooperative/competitive. Activity at one level z in one column cooperates with (does not inhibit) activity in adjacent columns at the same level z and inhibits (competes with) everything else. The restriction of activity to one level in a z-depth column is obviously a trivial but perfectly satisfactory way to account for the restriction of perceived fusion to a single depth plane at any x, y point in space.

The cooperative/competitive principle offers a solution to the ghost-ambiguity problem since, on the one hand, it restricts activity to one level of the NBF at any one point, and on the other, it encourages promising areas of activity to expand and to dominate weaker areas. The advantage of a cooperative/competitive principle for the solution to the binocular matching problem has been widely appreciated since the principle was first proposed by Sperling [1970]. The principle has subsequently been adopted as the basis of their binocular models by Julesz [1971], Dev [1975], Nelson [1975], and Marr and Poggio [1976].

Monoactive neural model: A case of pure competition. A simple neural network to achieve competitive selection of a single neuron in a column is illustrated in

Figure 8. Each neuron i receives an external input $x_i(t)$, produces an output $y_i(t)$, and receives its net inhibitory input Z_i from every other neuron j in the column $Z_i = k \sum_{j \neq i} y_j(t)$.

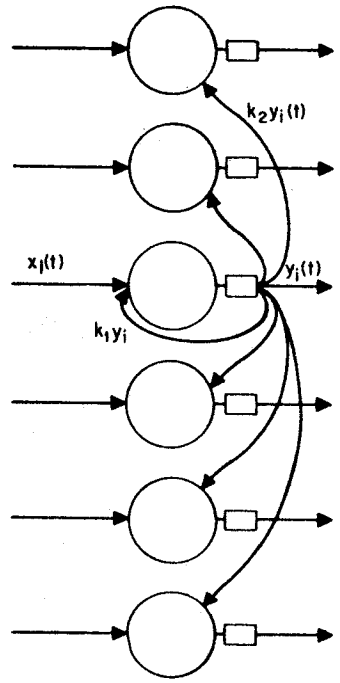

FIGURE 8. A monoactive column of model neurons. Output connections are shown for one neuron, i. It receives an external input $x_i(t)$ that sums with an output-produced feedback excitatory input $k_1 y_i(t)$. It sends shunting inhibitory signals $k_2 y_i(t)$ to all other cells. The small rectangular boxes in the input path represent thresholding operations: when the input is y, $y > 0$, the output is $\max(y - \varepsilon, 0)$.

A neural network with merely recurrent inhibition would be a very tame "filter" until three additions were made: (1) Positive feedback: each neuron feeds back a portion of its output as an excitatory input; (2) Threshold: each neural output y is subjected to a thresholding operation such that $y_t = \max(y - \varepsilon, 0)$; (3) Saturation: each neuron has a maximum output, B. The effect of positive self-feedback is to destabilize the system greatly and to increase the intensity of interactions. If the strength of inhibitory connections k_y is sufficient so that an active neuron with output B can reduce all other outputs to below their threshold, then it alone will have an output and all other neurons will be silent. This situation obtains until another neuron captures the system. The system thus appears to have the desired property of *monoactivity*–only one active neuron at any one time (Sperling [1970, Appendix B]).

Although the above argument may sound believable, it was (Grossberg [1973], [1978]) who first provided an adequate mathematical analysis of competitive systems of this kind (excitatory self-feedback, widespread inhibitory feedback).

Let us re-examine the assumptions. As has been pointed out earlier (in the section entitled *The mathematical neuron*) output saturation is an inherent part of Grossberg's theory so it does not require a special assumption. In order for the system to have the multistable monoactive property, Grossberg found that the form of the inhibitory feedback return function was critical; that is, the amount of signal fed back had to be an *S*-shaped function of input intensity. [Sperling's output function is zero for inputs below threshold and rises linearly to *B* when inputs are above the threshold; thus, threshold and saturation create an angular member of the *S*-function class.]

The point of this discussion is that monoactivity of a column of model neurons can be achieved under a wide variety of conditions and assumptions. Sufficient conditions are: inhibitory feedback onto the other neurons within the monoactive column, excitatory self-feedback, and an *S*-shaped feedback return function (or output threshold and saturation).

Monoactive systems perform an exceedingly universal and useful function: selecting one from many inputs. For example, in motivational systems, when an animal is hungry, thirsty, and sleepy, and can satisfy these drives at three different locations, it does not walk to the mean location. It selects these activities one at a time. When there are several objects of possible interest in the visual field, the eyes do not point at the blank space between them. They fixate the objects successively. The same is true of vergence movements when objects at different depth planes are present in the field of view.

In the case of binocular fusion, the theory proposes that there is a monoactive system at each x, y column of the NBF. This is the correlate in the NBF of the restriction that depth can be perceived at only one depth at any one visual direction. Another possible example is the ambiguous figure: perceiving one perceptual organization precludes simultaneously perceiving the other possible organizations. In general, perceptual and motor functions that involve categorization are likely to be accomplished by means of a monoactive network.

Double-coupled neural model: cooperation/competition. The monoactive neural model essentially made single decisions: which neuron will be dominant at any one time. It had only one kind of interaction and may be considered to be single-coupled. The second neural model uses context to resolve ambiguity. Thus, many neurons are active at every time and the model deals with determining the *combination* of active neurons. It is double-coupled in the sense that it exhibits one kind of interaction with members of its column and another kind between columns.

The basic building block is a monoactive column of many strongly competitive neurons. Such a column represents the z dimension, depth, in the NBF. Many of these columns are densely packed together in the x, y plane, perpendicular to the z axis. The position in the *column* of the one active neuron *column* represents a decision about depth locally. How does this decision affect the decision arrived at in the adjacent columns? The proposal put forward by

Sperling [1970] is that a neuron at level z in one column facilitates neurons at and near to level z in adjacent columns, by inhibiting neurons elsewhere in the adjacent columns. In effect, the neuron attempts to influence adjacent columns to respond more-or-less in the same way as its own column. It makes inhibitory feedback connections to neurons at different levels in adjacent columns as in its own column. Lateral excitatory connections between same-level neighbors strengthen the cooperativity beyond that achieved by inhibiting-the-inhibition. But it is necessary that lateral interactions between neighbors be different from vertical interactions within a column. Otherwise, the whole NBF would function like a single column. The lateral connections to adjacent columns are not so strong that they override strong input signals. When inputs are weak or ambiguous, the effect of cooperation/competition interaction is to build "planes of influence" at the z level of the model neurons that happen to have the strongest input signals. When there is a depth-ambiguous area between two unambiguous areas, the boundary between the two planes of influence can be quite unstable. [In stereograms constructed to satisfy these conditions, it is sometimes possible actually to see the rapid sweep of one depth plane over another (Sperling [1970]).]

The interaction between adjacent model neurons that represent different depth planes can result in far-reaching effects that travel like a wave. In other words, local cooperation/competition can result in global interactions. This is one of the features that makes cooperation/competition models so interesting and attractive to theorists. While systems like the one proposed here have been simulated, I do not know of a comprehensive mathematical treatment; it seems to be a worthwhile area of investigation.

The cooperation/competition mechanism that serves binocular vision quite probably serves other perceptual functions that involve local/global interactions, such as figure/ground relations and depth perception. Interpreting a portion of a scene as figure causes the area interpreted as figure to spread and to extend its boundaries to the limits of the area interpreted as ground. Similarly, seeing an ambiguous motion stimulus (composed of many points) moving in one direction in one part of a stimulus causes the motion interpretation to be applied to all other parts.

Summary and conclusion. Simple demonstrations of path-dependent perceptual states led to simple, useful descriptive models (potential theory) and to formidable process models (cooperation/competition networks) that still present substantial mathematical challenges.

References

Allen, J. B., 1977a. *Two-dimensional cochlear fluid model: New results*, J. Acoust. Soc. Amer. **61**, 110–119.

_____, 1977b. *Cochlear micromechanics–a mechanism for transforming mechanical to neural tuning within the cochlea*, J. Acoust. Soc. Amer. **62**, 930 – 939.

Boring, E. G., 1933. *The physical dimensions of consciousness*, Century Co., New York.

Burt, P. and G. Sperling, 1981. *Time, distance, and feature trade-offs in visual apparent motion*, Psych. Rev. **88**, 171–195.

Cowan, T. M., 1974. *The theory of braids and the analysis of impossible figures*, J. Math. Psych. **11**, 190–212.

Dev, P., 1975. *Perception of depth surfaces in random-dot stereograms: a neural model*, Internat. J. Man-Machine Studies **7**, 511–528.

Diner, D. B., 1978. *Hysteresis in human binocular fusion: a second look*, doctoral thesis, California Institute of Technology.

Egan, J. P., 1975. *Signal detection theory and ROC analysis*, Academic Press, New York.

Fender, D. and B. Julesz, 1967. *Extension of Panum's fusional area in binocularly stabilized vision*, J. Optical Soc. Amer. **57**, 819–830.

Fincham, E. F. and J. Walton, 1957. *The reciprocal actions of accommodation and convergence*, J. Physiology (London) **137**, 488–508.

Fry, G. A., 1955. *Blur of the retinal image*, Ohio State University Press, Columbus, Ohio.

Furman, G. G., 1965. *Comparison of models for subtraction and shunting lateral-inhibition in receptor-neuron fields*, Kybernetika **2**, 257–274.

Graham, N., 1981. *The visual system does a crude Fourier analysis of patterns*, these PROCEEDINGS.

Green, D. M. and J. A. Swets, 1966. *Signal detection theory and psychophysics*, Wiley, New York.

Grossberg, S., 1972. *Pattern learning by functional-differential neural networks with arbitrary path weights*, Delay and Functional-Differential Equations and their Applications (K. Schmitt, ed.), Academic Press, New York, pp. 121–160.

_____, 1973. *Contour enhancement, short term memory, and constancies in reverberating neural networks*, Stud. Appl. Math. **52**, 217–257.

_____, 1978. *Competition, decision, and consensus*, J. Math. Anal. Appl. **66**, 470–493.

von Helmholtz, H. L. F., 1924. *Treatise on physiological optics*, 3rd ed. (J. P. C. Southall, trans.), Optical Soc. Amer., Rochester, N. Y., 1924; reprint, Dover, New York, 1962.

Julesz, B., 1971. *Foundations of cyclopean perception*, Univ. of Chicago Press, Chicago, Ill.

Kaufman, Lloyd, 1965. *Some new stereoscopic phenomena and their implications for the theory of stereopsis*, Amer. J. Psych. **78**, 1–20.

Krantz, D. H., R. D. Luce, P. Suppes and A. Tversky, 1971. *Additive and polynomial representation*, Foundations of Measurement, vol. 1, Academic Press, New York.

Krauskopf, J., 1962. *Light distribution in human retinal image*, J. Optical Soc. Amer. **52**, 1046–1050.

Licklider, J. C. R., 1961. *Basic correlates of the auditory stimulus*, Handbook of Experimental Psychology (S. S. Stevens, ed.), Wiley, New York, pp. 985–1039.

Link, S. W. and R. A. Heath, 1975. *A sequential theory of psychological discrimination*, Psychometrika **40**, 77–195.

Luce, R. D. and D. M. Green, 1972. *A neural timing theory for response times and the psychophysics of intensity*, Psych. Rev. **79**, 14–57.

_____, 1974. *Counting and timing mechanisms in auditory discrimination and reaction time*, Contemporary Developments in Mathematical Psychology, vol. 2, Measurement, Psychophysics, and Neural Information Processing (D. H. Krantz, R. C. Atkinson, R. D. Luce and P. Suppes, eds.), Freeman, San Francisco, Calif.

Luneburg, R. K., 1947. *Mathematical analysis of binocular vision*, Princeton Univ. Press, Princeton, N. J.

Marr, D. and T. Poggio, 1976. Science **194**, 283–287.

Navon, D. and D. Gopher, 1979. *On the economy of the human-processing system*, Psych. Rev. **86**, 214–255.

Nelson, J. I., 1975. *Globality and stereoscopic fusion in binocular vision*, J. Theoret. Biol. **49**, 1–88.

Norman, F., 1972. *Markov processes and learning models*, Academic Press, New York.

Ogle, K. N., 1950. *Researches in binocular vision*, W. B. Saunders, Philadephia, Pa.

Pugh, Jr., E. N. and J. D. Mollon, 1979. *A theory of the Π_1 and Π_3 color mechanisms of Stiles*, Vision Res. **19**, 293–312.

Rall, W., 1964. *Theoretical significance of dendritic trees for neuronal input-output relations*, Neural Theory and Modeling (R. F. Reiss, ed.), Stanford Univ. Press, Stanford, Ca.

Roberts, F. S., 1979. *Encyclopedia of mathematics and its applications*, vol. 7, *Measurement theory*, Addison-Wesley, Reading, Mass.

Schroeder, M. R., 1975. *Models of hearing*, Proc. IEEE, **63**, 1332–1350.

Shepard, Roger N., 1980. *Multidimensional scaling, tree-fitting, and clustering*, Science **210**, 390–398.

Sperling, G., 1970. *Binocular vision: a physical and a neural theory*, Amer. J. Psych. **83**, 461–534.

_____, 1978. *The goal of theory in experimental psychology*, unpublished technical memorandum, Bell Laboratories.

Sperling, G. and M. J. Melchner, 1976. *Estimating item and order information*, J. Math. Psych. **13**, 192–213.

_____, 1978. *Visual search, visual attention, and the attention operating characteristic*, Attention and Performance. VII (J. Requin, ed.), Erlbaum, Hillsdale, N. J., pp. 675–686.

Sperling, G. and M. M. Sondhi, 1968. *Model for visual luminance discrimination and flicker detection*, J. Optical Soc. Amer. **58**, 1133–1145.

Sussman, H. J. and R. S. Zahler, 1978. *Catastrophe theory as applied to the social and biological sciences: a critique*, Synthese **37**, 117–216.

Thom, T., 1974. *La théorie des catastrophes: état présent et perspectives*, Dynamical Systems (A. Manning, ed.), Lecture Notes in Math., vol. 468, Springer-Verlag, Warwick, pp. 366–372.

Victor, J., R. Shapley and B. Knight, 1977. *Nonlinear analysis of cat retinal ganglion cells in the frequency domain*, Proc. Nat. Acad. Sci. U.S.A. **74**, 3068–3072.

Wyszecki, G. and W. S. Stiles, 1967. *Color Science*, Wiley, New York.

DEPARTMENT OF PSYCHOLOGY, NEW YORK UNIVERSITY, NEW YORK, NEW YORK 10003