BINOCULAR VISION: A PHYSICAL AND A NEURAL THEORY

GEORGE SPERLING, *Bell Telephone Laboratories, Murray Hill, New Jersey* and *New York University*

Abstract. The classical model of energy levels consists of a bumpy surface upon which a marble rolls freely. Dips in the surface represent stable states in which the marble can become entrapped. This physical theory is applied to three major processes of binocular vision: accommodation (a), horizontal vergence (v), and fusion (u). A simple model illustrates how the same external stimulus may produce different vergence states, depending on the preceding stimuli, and it shows how extreme values of vergence can be achieved. The model applies directly to vertical vergence and to torsional eye rotations. A similar model fits the accommodation and fusion systems and can be extended to account for the v-a-u interactions.

The neural structures underlying the systems differ. The neural theory of fusion proposes two neural binocular fields (NBFs): a primary NBF for fine details and for fine depth discrimination in the stimulus (its outflow corresponds to the cyclopean binocular view); and a secondary NBF for coarse details and large depth signs (its outflow directs fine analysis to relevant portions of the visual field, it contributes nothing except coarse depth signs directly to the cyclopean view). The fusion model illustrates how the same retinal stimulus may produce different stable perceptual states, depending on recent stimuli, and it accounts for fusion and rivalry within the same system and the same neurons. Appendix A offers a general definition of image blur, and Appendix B a quantitative analysis of the multistable phenomena in rivalry and fusion.

Because the two eyes provide two different views of the world, a person with normal binocular vision can determine the relative depth of unfamiliar objects in his field of view when he and the objects are motionless. This ability is called stereoscopic depth perception—or, often, just binocular depth perception, to emphasize its dependence on both eyes. The evolutionary value of binocular depth perception is obvious, and some interesting and complicated adapta-

Received for publication May 5, 1970. The author expresses his appreciation to Dr. Mohon Sondhi for his help with Appendixes A and B, to Dr. John Krauskopf for his stimulating criticisms, and to Dr. C. F. Strohmeyer III and the late Dr. Lloyd Marlowe for their helpful comments on the manuscript.

Sperling, G. (1970). Binocular vision: A physical and a neural theory. *American Journal of Psychology, 83,* 461-534.

tions have evolved in our visual systems to facilitate it. This article explores the intricate interrelationships between the three most important mechanisms for binocular vision: the mechanisms for vergence, for accommodation, and for fusion.

The most important subjective phenomenon of binocular vision is binocular *fusion*. When corresponding points of the retinas of the two eyes are stimulated with light from the same (or nearly the same) object, only one object is seen. We say the two retinal images are fused. Ordinarily, the experience of one object is so compelling that fusion requires no further definition.[1]

Binocular fusion is possible only when stimuli fall on corresponding—or nearly corresponding—points of the retinas. To clarify what we mean by corresponding points, we suppose the eyes are looking at a point at infinity. (For practical purposes, this means a point on an object several hundred yards or more away.) Then, the lines of sight of the left and of the right eye are parallel, and identical images fall on each retina. For the moment, we define corresponding points as the corresponding retinal points under these retinal images.[2] To view a near object the eyes must rotate inward (converge) in order for the object to stimulate these corresponding points. The *vergence* movements of the eyes are the second remarkable adaptation of binocular vision. When the eyes are verged on a particular object, only objects in a limited range of depth planes can be fused; objects outside this range are unfused.

[1] For the purist, binocular fusion can be defined objectively in terms of shape (or pattern) invariance. The principle is that small lateral translations (in opposite directions) of the stimuli to the two eyes leave the shape of a fused pattern unchanged. Additionally, it also must be shown that both eyes' stimuli contribute to the perceived pattern. These are sufficient (but not necessary) conditions. Specifically, a pattern of illumination to the left retina $[l(x, y)]$ and to the right retina $[l'(x, y)]$ are binocularly fused if there exists a Δx, $|\Delta x| > 2$ min., such that when the new pattern $[l(x + \Delta x, y)$ and $l'(x - \Delta x, y)]$ replaces the original pattern on the retinas, the observer reports the shape of the new pattern is the same as that of the old. Its depth in space may or may not have changed. To ensure that both l and l' contribute to the shape of the pattern, there must be points on the left stimulus [e.g., bright spots of light $\delta(x_1, y_1)$, $\delta(x_2, y_2)$ . . .] and on the right stimulus $[\delta'(x_1', y_1')$, $\delta'(x_2', y_2')$ . . .] such that the two sets of points $[\delta(x_i, y_i)$ and $\delta(x_i', y_i')]$ are all different and the observer reports seeing points from both sets in any neighborhood where fusion is being tested. The motivation for this definition derives from the neural model of fusion.

[2] This actually is a definition of 'corresponding coordinates.' Subsequently, 'corresponding points' are defined as those whose projections intersect in the middle layer of the neural binocular field. In the neighborhood of fixation, corresponding coordinates and corresponding points coincide exactly.

The third major phenomenon of binocular vision is *accommodation*, whereby the lens of the eye varies its focusing power so that the depth plane of the object being viewed is in focus on the retina. Though accommodation is also characteristic of monocular vision, it is important in binocular vision because it is closely linked to vergence. That is, accommodating for a near object causes convergence, and convergence causes accommodation.

A priori, neither fusion nor vergence nor accommodation would be necessary for binocular depth perception. However, once a mechanism of fusion has developed, the evolution of accurate vergence and accommodation would enable the fusion mechanism to evolve further so as to substantially increase its power of depth discrimination. In fact, once all three mechanisms exist, there are such great advantages to binocular vision to be gained from their mutual interactions that these interactions are virtually predetermined. This paper presents two kinds of models, physical and neural, to account for the fundamental phenomena of vergence, accommodation, and fusion, and for the interactions between the three mechanisms.

Multiple stable states and path dependence. Our starting point is an easily observed phenomenon: the existence of *multiple stable states*. For example, Helmholtz, writing in the last century, observed that "the eyes also may be made to diverge by viewing stereoscopic pictures and gradually separating them farther and farther apart, all the time trying to fuse them into a single image. I am able in this way to produce a divergence of the lines of fixation of my eyes amounting to as much as eight degrees."[3]

To restate Helmholtz's procedure in more formal terms, he presented a picture (L) to the left eye and an identical copy (L') to the right eye by means of an optical device, such as a Wheatstone stereoscope. By moving L slowly to the left and L' to the right, he was able to cause his eyes to follow the movement of L and L', and thereby to diverge. But if L and L' had first been presented at the extreme separation of 8 deg., his eyes could not have diverged upon them; his eyes would have remained approximately in their initial position. This demonstrates multiple stable states of vergence. When L and L' are 'separated' by 8 deg.—that is, when the ver-

[3] H. von Helmholtz, *Treatise on Physiological Optics*, 3d ed., J. P. C. Southall (trans.), 3, 1924 (reprinted 1962), 58.

gence angle required for fusion is 8 deg.—then the eyes may sometimes verge correctly and at other times be in their neutral position. There are two stable states of vergence for the same external stimulus. The particular vergence state attained by the eyes depends not only on the present stimulus but also on the sequence of previous stimuli.

As we shall see, Helmholtz's demonstration involves more complications than are apparent at first. First, the amount of divergence that can be attained depends on the choice of stimulus pictures (L and L'); for example, a greater amount of divergence is possible when pictures contain many high-contrast details. Second, with practice, some voluntary divergence is possible. Whether or not voluntary divergence suffices to produce fusion depends on the amount of divergence that is required and on the particular pictures (L and L') being viewed. Third, fusion of L and L' can occur even when the eyes are not fully verged on L and L'. The first two complications, and of course the basic observation, are shown below to be immediate consequences of the physical model of vergence. The third complication follows from the joint action of the models of vergence and fusion.

It often is convenient to represent stimuli or responses as points in an n-dimensional space. In the example above, the vergence response can be represented in one-dimensional space as a point along a line, where the distance of the point from the origin represents the vergence angle of the eyes. The sequence of responses (to a sequence of stimuli) can be thought of as a path from point to point in this space. The conventional way of saying that a response depends not only on the present stimulus but also on the sequence of past stimuli is to say that the response is *path dependent.* That is, how the eyes respond to a particular stimulus (whether they verge or not) depends on their current vergence state as well as on the current stimulus. Because the current vergence state depends on previous stimuli, the response depends on both current and past stimuli.

In all the examples in this paper, path dependence and multiple stable states refer to the same phenomenon. The only way to achieve multiple stable states (different responses to same stimulus) is by taking different stimulus and response paths prior to presentation of the same final stimulus; the only way to demonstrate path dependence is to achieve different stable states to the same final stimulus (by virtue of having taken different paths).

Thus, multiple stable states and path dependence are logically equivalent.

Multiple stable states also occur in the fusion process. For example, suppose that vergence of the eyes is artificially maintained on a particular depth plane by optically canceling vergence motions.[4] Fusion is possible not only of objects in that particular depth plane but also of objects in other depth planes. The fusion or nonfusion of a particular object in a particular plane exhibits the same kind of multiple stable states as the vergence motions of the eye. That is, whether the object is seen as fused or double depends on whether it was seen as fused in its previous position, and so on. Finally, by methods to be described later, we may also observe multiple stable states of accommodation (focusing of the eye).

The multiple stable states of vergence, fusion, and accommodation follow from the same elementary principles, are representable by the same kind of physical model, and exhibit analogous correlated phenomena. Although the physical models of vergence, fusion, and accommodation are formally the same, the underlying mechanisms are different. The last main section of this article contains a speculative suggestion toward a neural model of these mechanisms.

THE PHYSICAL MODEL—WITHOUT INTERACTIONS

The classical model of energy levels consists of a bumpy surface, the *energy surface,* upon which a marble rolls freely. Dips in the surface are called *energy wells*; they represent stable states in which the marble can be trapped. Here, this classical model is applied to the three major processes of binocular vision: vergence (v), accommodation (a), and fusion (u). In these applications, the lateral position of the marble represents the instantaneous value of the depth plane of v or a or u. For example, the basic surface governing horizontal vergence is bowl-shaped, the shape being determined by internal factors. The marble rolls toward the center of the bowl, representing the tendency of the eyes to verge to a neutral position in the absence of a stimulus to vergence. The external stimulus is assumed to add perturbations to the basic surface, thereby creating new stable states.

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

The physical model of binocular vision is thus based on a concept that, because of the physical analogy, is called energy (e). For example, in the case of vergence, the visual system is assumed to compute the net vergence energy of various possible states of vergence and—within certain limits—to achieve the state which has the lowest net vergence energy (e_v).

Let v denote the vergence angle between the eyes, a their accommodation, and u the plane of fusion. Let l and l' denote the stimuli on the left and right retinas respectively. There are assumed to be two component sources of energy, which add to produce net energy: *displacement energy* (g) and *image-disparity energy* (h). Displacement energy depends only on the current states of v, a, and u. It represents the amount of energy needed to displace v, a, and u from their natural resting values to their current values. Image-disparity energy depends only on l and l'. It represents the energy that a visual stimulus can contribute to displacing v, a, and u.

The physical model is a particular embodiment of a set of equations that expresses the general principles outlined above. The equations deal with the quantities e, g, and h, which represent respectively the *total*, the *internal*, and the *external* components of energy. Here the subscripts v, a, and u indicate the system whose energy is under consideration; and the variables v, a, and u represent the current values of vergence, accommodation, and fusion. Commensurate units for v, a, and u themselves are defined in the following sections. The equations contain all the interactions of the model and are presented here to provide an overview. They are not necessary for understanding the principles of the model, however, and readers who are not interested in mathematics should simply skip to the next section, on the model of vergence.

$$e_v(v, \cdots,) = g_v(v - a) + h_v(l, l') \qquad [1]$$

$$e_a(a, \cdots,) = g_a(a - v) + h_a(l, l') \qquad [2]$$

$$e_u(u, z, \cdots,) = g_u(u - v, z) + h_u(l, l', z), \qquad z \subset R \qquad [3]$$

Equation 1 represents vergence; Equation 2 represents accommodation; and Equation 3 represents a set of equations for fusion, one for each region (z) of the retina (R). Each equation may be thought of as describing an energy surface upon which rolls a marble, whose projection on the v, a, and u axes respectively represents the value of v, a, and u.

Equations 1–3 do not give information about the dynamics of v, a, u, because time does not enter into them. Dynamic specification of v, a, and u requires another set of equations. Let x_1, x_2, x_3 represent v, a, and u respectively. The dynamics of x_i are defined by the first-order differential equations:

$$k_{1i}\frac{d^2x_i}{dt^2} + k_{2i}\frac{dx_i}{dt} = \frac{-\partial e(x_i)}{\partial x_i} \qquad (i = 1, 2, 3). \qquad [4]$$

The right-hand term of each Equation in 4 represents force acting on the marble; the coefficients k_{1i} and k_{2i} respectively reflect the marble's mass and the friction of the medium.

Although only the spatial derivative of energy (force) enters into Equations 4, the ensuing description of the systems is in terms of energy because it gives better intuitive insight. We may now begin to deal with the explicit definition of the quantities in Equations 1–3, the interpretation of the resulting equations, and their comparison with data.

Model of Vergence

Units of vergence. The horizontal vergence of the eyes is best expressed as the *angle* between their lines of sight. The angular unit of vergence is particularly convenient because it fits in with the spherical coordinate system used to describe the position of stimulus points in space and with the dioptric system used to measure accommodation. However, binocular vision is concerned mainly with depth, so it is useful to keep in mind that there is a one-to-one correspondence between the angle of vergence and the distance from the eyes to the verged-upon point. (We confine ourselves here to points directly in front of the observer, so-called symmetrical vergence.) The angular measure of vergence can thus be considered to be a transformed measure of depth, and in the discussion that follows, the term vergence *position* is used to evoke the idea of the distance at which the eyes are verged. The sign convention followed is that *con*vergence angles (lines of sight intersect in front of the observer) are positive and correspond to positive distances; *di*vergence angles (lines of sight intersect behind the observer) are negative and correspond to negative distances (i.e., behind the observer). In the case of divergence, of course, 'distance' does not have the usual physical interpretation, and we have no need of the distance transformation.

In this section on vergence, we assume that accommodation of the eye's lens is fixed, at a value corresponding to accurate focus of objects at the given viewing distance (e.g., 1 m.). To a first approximation, accommodation does remain fixed if stimuli are confined to a particular viewing distance (e.g., 1 m.). To maintain accommodation more accurately at a given nominal value would require a negative-feedback optical device.[5]

Vergence-displacement energy (g_v). For any given amount of accommodation of the eye lens, there is a natural vergence angle (v_0) of the eyes. That is, the natural vergence angle is the angle formed by the lines of sight of the eyes in the absence of any stimulus to vergence. (For example, a stimulus seen only in one eye is a stimulus to accommodation but not directly to vergence.)

When the vergence angle of the eyes is any value (v), vergence displacement ($\triangle v$) is defined as the angular deviation of the eyes from the natural vergence angle: $\triangle v = v - v_0$. Vergence-displacement energy is defined as a function of $\triangle v$: as $g_v(\triangle v)$, where $g_v(0) = 0$; that is, vergence-displacement energy is zero when the eyes are in the natural vergence angle. Any deviation ($\triangle v$) of the eyes from the natural vergence angle is assumed to have a larger vergence-displacement energy, according to a U-shaped function such as that illustrated in Figure 1.

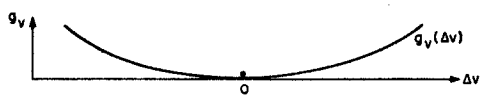

FIG. 1. Vergence-displacement energy, $g_v(\triangle v)$, as a function of vergence displacement, $\triangle v$. The marble representing vergence is at its equilibrium position, $\triangle v = 0$. Positive values of $\triangle v$ indicate convergence (relative to the equilibrium position); negative values indicate divergence. In this example, $g_v(\triangle v) = \log_e (1 + z) - z - .307$ where $z = [1 + (\triangle v)^2]^{1/2}$. This surface is the gravimetric analog of Hook's law: the horizontal restoring force (f) on the marble due to gravity is $f = -k\triangle v$.

The physical analogy here is a simplified version of the complete analogy, which is described below after all interactions have been considered. The curve $g_v(\triangle v)$ of Figure 1 represents a physical

[5] Whenever accommodation strays from its nominal value, a negative-feedback optical device so alters the viewing distance of the objects that the eyes' normal attempt to maintain accurate accommodation then restores accommodation to the nominal value.

surface, a 'bowl.' A small marble rolls on this surface, and the projection of the marble's position on the horizontal axis represents the vergence angle of the eyes. The rule is that the marble tends to roll downhill, due to gravity. The force on the marble is proportional to the slope of the bowl ($-\partial e/\partial v$, Equations 1 and 4);[6] its motion depends on its mass and the friction it encounters (k_{1i} and k_{2i} respectively in Equation 4).

The g_v surface of Figure 1 represents the internal factors controlling vergence in the absence of an external stimulus to vergence. On this surface the marble rolls to a stable equilibrium position, to $\triangle v = 0$, the center of the g_v ($\triangle v$) curve. In order for the model to correspond to the eyes, the coefficient of mass of the marble is chosen sufficiently smaller than the coefficient of friction so that the marble rolls to the bottom without overshooting. In a real bowl, for example, this situation is achieved by filling the bowl with fluid. To move the marble from $\triangle v = 0$ requires 'work,' or 'energy.' The source of energy is image-disparity energy, described below.

Vergence image-disparity energy (h_v). Vergence image-disparity energy (h_v) represents the extent to which two retinal images agree. It is defined as follows. Let spherical coordinate systems for each retina be defined so that $x = y = 0$ is the fixation point (center of the visual field, line of sight) and so that corresponding points of the two retinas are assigned equal values of x and y. Let the illuminance distributions (images) on the left and right retinas respectively be $l(x, y)$ and $l'(x, y)$. The image-disparity energy $\triangle h_v$ in a small neighborhood ($\triangle x \triangle y$) of a point (x, y) is defined as

$$\triangle h_v(x, y) = |l(x, y) - l'(x, y)| \, \triangle x \triangle y. \qquad [5]$$

That is, $\triangle h$ is proportional to the absolute value of the amount by which l and l' differ.

The image-disparity energy (h_v) is given by summing $\triangle h_v(x, y)$ over the whole retina (R). In summing image disparity over the retina, central areas of the retina are given greater weight than peripheral areas. This selective weighting of the central area is ac-

[6] Let $e' = \partial e(x)/\partial x$. In the system of Equations 1–4, the horizontal force (f) acting to change the vergence position (v) is defined by $f = -ke'$. In the gravimetric analog, $f_g = -k \sin e' \cos e' = -ke'/(1 + e'^2)$, which is an irrelevant complication. The gravimetric analog becomes equivalent to Equations 1–4 as $e'^2 \to 0$. Thus, by scaling units appropriately, the gravimetric analog satisfies the equations.

complished by multiplying $\triangle h_v(x, y)$ by a weighting function $[w(x, y); 0 \leq w(x, y) \leq 1]$ before summing. Equation 6 then gives vergence image-disparity energy:

$$h_v = \sum_R w(x, y) \triangle h(x, y)$$

$$\approx \iint_R |l(x, y) - l'(x, y)| \, w(x, y) \, dx \, dy. \qquad [6]$$

Equation 6 represents a point-by-point comparison of the left and right eye's retinal images (l and l'). When l is identical with l', h_v is zero. When l is not identical with l', h_v is greater than zero.

In practice, Equation 6 is somewhat inaccurate because it gives too much weight to large areas where l and l' differ slightly, and too little weight to small areas where l and l' differ sharply (boundary conflicts). This discrepancy arises because in the nervous system, each visual image is transformed before the two images are compared in the brain. The difficulty is overcome by not using $l(x, y)$ directly in Equation 6 but $T{:}l(x, y)$ where T represents the various retinal (and subsequent) neural transformations on l, such as adaptation and boundary enhancement. Another problem with Equation 6 is that the attention-weighting function $[w(x, y)]$ is not specified. The elucidation of neural transformations and of attention are interesting problems in visual psychology, but we can bypass them. The elementary definition of h_v by Equation 6 suffices for the present, providing that $w(x, y)$ is taken as any reasonable function that decreases as x and y increase.

Calculation of l. Given any external luminance distribution (L), we wish to calculate the retinal illuminance distribution (l). In general, L and L' (to left and right eyes respectively) could occur at any arbitrary distance from the observer and at any direction relative to his head; his eyes could be pointed in some other direction and be verged at another distance. Our interest here is in the effect of vergence upon retinal images, and we can avoid the hideous complications of this general formulation by restricting ourselves to symmetrical vergence; that is, to the case of an observer looking at stimuli directly in front of him. For this purpose it is convenient to define coordinates of the external stimuli (L and L') as those of the retinal stimuli they would produce if the observer's eyes were pointed straight ahead with a vergence angle of $v = 0$ (i.e., were verged at infinity.) Spherical coordinates are used, although so long as we restrict consideration to objects directly in

front of the observer and to a small neighborhood of fixation, there is no important difference between Cartesian and spherical coordinates. The external horizontal coordinate (x) is given in angular units; $x = 0$ represents the projection at infinity of the midsaggital plane (the plane through the nose and perpendicular to the line connecting the centers of the two eyes). Positive values (angles) of x are to the right. The external vertical coordinate (y) is defined with positive values upward.

When $v = 0$, the retinal coordinates (x_r and y_r) and the external coordinates (x and y) coincide exactly, by definition. When the vergence angle is v, external and retinal coordinates are related by $x = x_r + v/2$ (left retina) and $x = x_r - v/2$ (right retina). For both eyes, it is assumed that $y = y_r$. For example, when the vergence angle is 2 deg., the left eye is rotated 1 deg. to the right (toward the nose) and the right eye is rotated 1 deg. to the left. The center of the left retina then is pointed at $x = v/2 = +1$ deg.

If we neglect the blur of the retinal image caused by the imperfect optics of the eye, then the retinal illuminance distributions (l and l') are exactly proportional to the external luminance distributions: $l(x_r, y) = kL(x, y)$ and $l'(x_r, y) = kL'(x, y)$, where k expresses the conversion factor from external luminance to retinal illuminance. Substituting the appropriate external coordinates for the left and right retinal coordinates gives simple expressions to relate l and L in terms of the eyes' vergence angle (v):

$$l(x - v/2, y) = kL(x, y) \quad \text{and} \quad l'(x + v/2, y) = kL'(x, y).$$

The expressions for l and l' contain the relevant information about the external stimulus, so that L and L' are not needed in subsequent calculations.

Calculation of $h_v(\Delta v)$. Let the external stimuli be $L(x, y)$ and $L'(x, y)$. Let the eyes be in their natural vergence position (v_0) for the state of accommodation, so that the retinal stimuli are $l(x - v_0/2, y)$ and $l'(x + v_0/2, y)$. The vergence image disparity (h_v) for this pair of retinal images is given by substituting l and l' into Equation 6. Now, let the eyes alter their vergence by Δv, to $v = v_0 + \Delta v$. The vergence image disparity still is given by substituting l and l' in Equation 6; now it is a function of Δv, namely,

$$h_v(\Delta v) = \iint_R |l[x - (v_0 + \Delta v)/2, y] - l'[x + (v_0 + \Delta v)/2, y]| \, w(x, y) \, dx \, dy. \qquad [7]$$

Equation 7 is at the very heart of the theory so it is important that the reader understand it completely. To help intuitive understanding of Equation 7, the concept of a cyclopean image is introduced here. Our two eyes receive different retinal images (l and l') of the world; yet, we are unaware of having two separate images and perceive things as single, like the one-eyed Cyclops. For the moment we define the cyclopean image ($^*l$) simply as the sum of the two separate images:

$$^*l(x, y) = l(x, y) + l'(x, y).$$

(Later, in the section on the neural model and its predictions for binocular rivalry, we shall see that the subjective cyclopean view is a much more complicated combination of l and l' than their sum.)

When the external stimuli (L and L') remain fixed, then the retinal images (l and l', and hence $^*l$) depend on the vergence angle of the eyes. Figure 2c illustrates a pair of external stimuli (L and L'); the cyclopean images that would result for two different vergence positions are shown in Figure 2ab.

The critical term of Equation 7 is $|l - l'|$, which is like a cyclopean image except that l and l' are subtracted instead of being summed. Since image-disparity energy (h_v) depends on the difference between l and l', it is more nearly zero the more nearly identical l and l' are. For the vergence position v_0 of the eyes illustrated in Figure 2a, $h_v(v_0)$ is at a relative minimum because the two rectangles are optimally superposed. The image-disparity energy $h_v(v_0)$ is not zero, however, because neither the rectangles nor the disks superpose exactly, and thus they fail to cancel completely. When the vergence angle of the eyes is v_1, the two disks are superposed but the rectangles fail to superpose. Image-disparity energy calculated at $\triangle v_1$ again is at a relative minimum, but $h_v(\triangle v_1) > h_v(0)$ because of the large contribution of the now unsuperposed rectangles.

Figure 2 also illustrates that changing vergence of the eyes is equivalent to keeping the eyes stationary but shifting the positions of the physical stimuli (L and L'). The image-disparity energy function $h_v(\triangle v)$ is an index of the similarity of L and L' as a function of their relative position. However, Equation 7 is written in terms of the retinal stimuli (l and l') in order to emphasize that the eyes need not actually verge to compute $h_v(\triangle v)$. A single pair of retinal images (l and l'), obtained in a single glimpse, contains

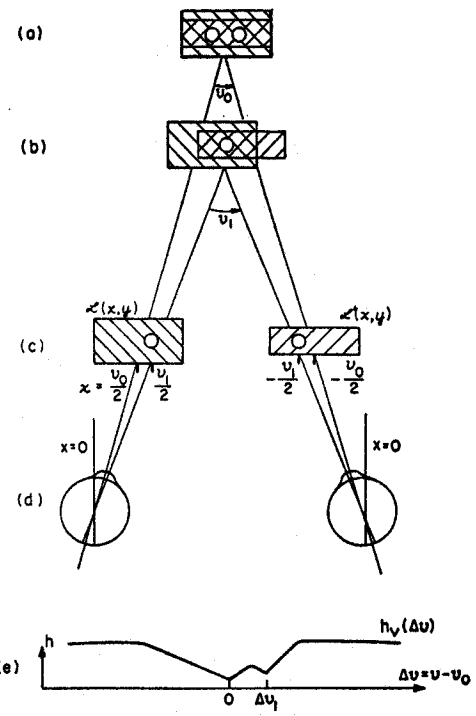

FIG. 2. Calculation of image-disparity energy, $h_v(\triangle v)$. (d) The eyes are in the natural vergence position, v_0, induced by accommodation on the retinal stimuli, $L(x, y)$ and $L'(x, y)$, illustrated in (c); the vertical position of L and L' is arbitrary. (a) Cyclopean image illustrating $l(x - v_0/2,y) + l'(x + v_0/2,y)$, from which $h_v(0)$ is calculated. (b) Cyclopean image illustrating $l[x - (v_0 + \triangle v_1)/2,y] + l'[x + (v_0 + \triangle v_1)/2,y]$, from which $h_v(\triangle v_1)$ is calculated. (e) The $h_v(\triangle v)$ for the image pair (c), as given by Equation 7. The abscissa, $\triangle v$, is in angular units of vergence. Clockwise rotations of the left eye and counterclockwise rotations of the right eye are indicated by positive values of $\triangle v$, which represent convergence.

all the information needed to compute the entire function $h_v(\triangle v)$. In other words, a single glimpse suffices to 'tell' the visual system the cyclopean image it should expect to see in any vergence position, and from this information it can compute $h_v(\triangle v)$. Just how much of this potentially available information actually is used by the visual system is considered below. Here, for convenience, the model is stated as though the entire function $h_v(\triangle v)$ were known; in fact, knowledge of just two values of $h_v(\triangle v)$ at any one time would be a sufficient assumption.

Figure 2e illustrates $h_v(\triangle v)$ as a function of $\triangle v$. Two interesting properties of $h_v(\triangle v)$ are evident: first, that $h_v(\triangle v)$ consists of the sum of functions, each of which represents registration of some part of the images and, therefore, each of which is symmetrical about its minimum (if the effect of w is negligible); and, second, that at any relative minimum of $h_v(\triangle_v)$, its derivative $dh_v(\triangle v)/d(\triangle v)$ is discontinuous. The discontinuity of the derivative results from the absolute-value operation in Equation 3. As a practical matter, such discontinuities do not exist in real life because image blur and 'noise' (uncertainty) have the effect of rounding the minimum.[7]

Examples of the model for vergence. A simpler illustration of binocular stimuli (L and L') and their corresponding $h_v(\triangle v)$ function is given in Figure 3. In this example, the retinal images (l and l') would become identical if the eyes diverged slightly. Thus in

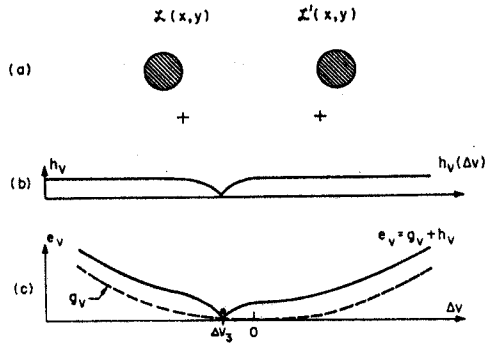

FIG. 3. A visual display with a single stable vergence state. (a) The left eye sees $L(x, y)$, the right eye $L'(x, y)$; the crosses in L and L' indicate the point $x = y = 0$; the disks are symmetrically displaced from $x = y = 0$ (nasally on the retinas, temporally in the external world). (b) Vergence image-disparity energy, $h_v(\triangle v)$, as a function of vergence displacement, $\triangle v$, for the display illustrated in (a). (c) Vergence energy, $e_v(\triangle v)$, as a function of $\triangle v$. The point $\triangle v_s$ indicates the projection on the horizontal axis of the minimum of the energy well in e_v. The marble representing vergence is in the energy well, its only stable position.

[7] The discontinuity of its derivative at zero gives the absolute-value function some rather strange properties that probably do not accurately reflect nature (which abhors discontinuous derivatives). Taking into account the effect of the neural transformation on l and l' before $h_v(\triangle v)$ is computed improves matters (see Appendix A).

Figure 3b, $h_v(\Delta v_3) = 0$. Figure 3c illustrates vergence energy $e_v(\Delta v)$ defined as the sum of $g_v(\Delta v)$ and $h_v(\Delta v)$.

The binocular stimuli (L and L') of Figure 3a yield an $e_v(\Delta v)$ function that has only one relative minimum. For this function, there is one and only one stable position of the marble, namely, at the bottom of the $e_v(\Delta v)$ function. This means that no matter what the initial vergence position of the eyes may be, they ultimately stop at the vergence position Δv_3.

Consider now what happens when L and L' of Figure 3a are separated horizontally; that is, when the left eye's stimulus is moved to the left and the right eye's stimulus to the right. Vergence becomes increasingly difficult. Figure 4 illustrates that a marble initially at rest (at $\Delta v = 0$) does not roll into the energy well

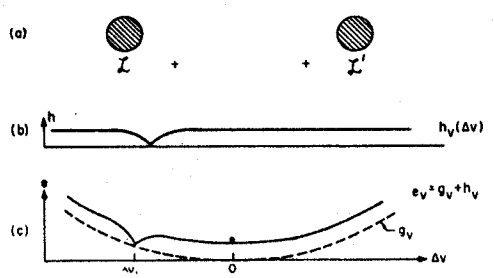

FIG. 4. A visual display with two stable vergence states. The marble representing vergence is in its neutral position, $\Delta v = 0$. Vergence on the disks cannot be established if the disks are turned on when the eyes are at $\Delta v = 0$. Once vergence is established (e.g., by slow lateral separation of the disks from the displacements shown in Figure 3), vergence can be maintained at Δv_4.

at Δv_4 but remains at $\Delta v = 0$. However, if the left and right eyes are initially verged on their stimulus at a displacement of about Δv_3 (Figure 3), and then the two stimuli are slowly separated, the marble remains in the energy well. The eyes remain verged. This example shows that the stimuli illustrated in Figure 4 permit two stable states of vergence.

When L and L' are separated by the amount illustrated in Figure 5, the energy well in the $e_v(\Delta v)$ curve has flattened out to such an extent that it no longer can hold the marble, which then rolls to the position of zero vergence displacement. The binocular stimulus pair of Figure 5 thus yields only one stable vergence position for the eyes.

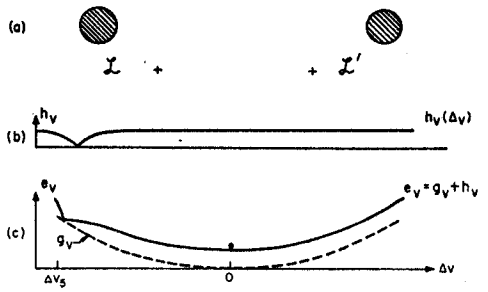

FIG. 5. A display for which vergence cannot be established. The disks are displaced laterally even further than those illustrated in Figure 4a. The marble representing vergence is in its only stable position, $\Delta v = 0$.

Suppose that L and L' initially are separated by the amount illustrated in Figure 5. Then, to establish vergence they would have to be brought closer together than the separation Δv_4 (Figure 4), where vergence would be maintained but not established; they would have to be brought to a displacement of about Δv_3 (Figure 3). A similar effect may be observed when illumination is interrupted. Suppose vergence somehow has been established at Δv_4 (Figure 4). Then, turning the illumination off for a few tenths of a second removes the energy well caused by h_v and allows the marble to move under the influence of g_v toward $\Delta v = 0$. It then requires a slow separation of the stimuli to reestablish vergence at Δv_4.

Slow separation of stimuli is not the only way to establish vergence at extreme divergence values. A succession of small, quick displacements (with longer time intervals between them) also succeeds. The amplitude of the quick displacement must be such that the marble remains inside the energy well. Then, during the interval between displacements, it rolls down into the new well. The process may be repeated until the critical divergence illustrated in Figure 5 is reached.

'Vergence disparity.' When the minima of $h_v(\Delta v)$ are rounded at their bottom [i.e., $dh_v(\Delta v)/d(\Delta v)$ is continuous at a minimum], then the minima of $e_v(\Delta v)$ do not correspond exactly to the minima of h_v but are shifted toward the center of e_v, toward $\Delta v = 0$. The difference in location of minima is negligible when they lie near $\Delta v = 0$, but it becomes significant as Δv increases, because the e_v function becomes steeper.

In the model, the marble rests only at minima of the e_v function.

When a minimum of e_v does not coincide exactly with a minimum of $h_v(\triangle v)$, it means that the eyes verge insufficiently to carry their retinal stimuli to exactly corresponding points. This phenomenon of human vision is known as vergence disparity and has been measured under a variety of conditions.[8] The psychophysical measurements of vergence disparity are qualitatively in agreement with those predicted by the model. Further comparisons between the model and experiment are deferred until the full interactions between vergence and accommodation have been considered.

Model of Vertical Vergence and of Torsional Rotation

The vergence model can be applied to two other kinds of eye movements. Vertical vergence occurs when the stimulus to one eye is displaced vertically. The eye can move vertically to again center itself on the displaced stimulus, even while the other eye remains relatively stationary. Similarly, when the stimulus to one eye is rotated around the line of sight, the eye can rotate to realign itself with the stimulus. These movements occur only within certain bounds, about 6 deg. for vertical vergence and about 7 deg. for torsional rotation (see n. 3, pp. 58–62).

The angle between the vertical elevation of the two eyes is exactly analogous to the horizontal vergence angle, and the same model applies to control of both kinds of angle. In the case of torsion, the torsional angle of the eye is the analog of the horizontal vergence angle, and the model applies independently to each eye. In fact, in composing demonstrations of multistability phenomena, it often is more convenient to use vertical vergence or torsional rotation, because these movements are less under voluntary control than is horizontal vergence.

Model of Accommodation

Just as the current state of accommodation determines the natural vergence position, the value of vergence determines the natural accommodation position. To displace accommodation from its nat-

[8] Among others, see K. N. Ogle, *Researches in Binocular Vision*, 1950; E. F. Fincham and J. Walton, The reciprocal actions of accommodation and convergence, *J. Physiol. (London)*, 137, 1957, 488–508.

ural position—from its value in the absence of any stimulus to accommodation—requires an amount of energy described by the accommodation-displacement energy function (g_a); this energy is supplied by a stimulus to accommodation via an image-defocus function (h_a); and the actual state of accommodation is predicted from the accommodation energy function (e_a): $e_a = g_a + h_a$. The model for accommodation is thus formally identical to the model for vergence and differs only in the details of calculating the functions. These details offer some special problems, which are considered below.

Units of accommodation. The conventional measurement unit of accommodation is the *diopter*. When the eye is focused on an object $1/n$ meters distant from it, it is said to be accommodated to n diopters. The scales of diopters of accommodation and of degrees of vergence are related almost linearly. For objects in front of the observer, the deviation from linearity is of the third order and does not exceed about 1% at 8 diopters of accommodation, which is the practical limit of the adult range. Therefore with negligible error, we can use the same scale for accommodation as for vergence. There is, in fact, an advantage in using the diopter scale: it is independent of the interpupillary distance (p) of the observer. For an average value of p (6.0 cm.), 1 diopter = 3.5 deg.

Accommodation-displacement energy (g_a). Theoretically, accommodation-displacement energy (g_a) may be regarded as a function of only $\triangle a$, the displacement of accommodation from the value (a_0) determined by vergence. In practice, however, the value of accommodation is strongly influenced by the mechanical properties of the lens.[9] The situation may be conceived of as one in which neural signals attempt to control accommodation according to a function $[g_a(\triangle a)]$ that depends only on $\triangle a$; but the transduction of these signals into accommodative power also depends on a. The nonlinear (a-dependent) transduction of attempted accommodation into achieved accommodation is an important and unfortunate consequence of the hardening of the lens caused by aging.[10] It is worth noting that rotation of the eyeball in vergence also is limited by

[9] For a summary of the evidence for distinguishing between the muscular force exerted on the lens and its mechanical response, see M. Alpern, W. M. Kincaid, and M. J. Lubeck, Vergence and accommodation: III, Proposed definitions of the AC/A ratios, *Amer. J. Ophthal.*, 48, 1959, 141–148.

[10] See Fincham and Walton (n. 8 above).

mechanical factors; in vergence, however, the neural control operates so far within the mechanical limits that the mechanical factors can be neglected.

Image-defocus energy (h_a). Image-defocus energy (h_a) represents the blurredness of a retinal image. Although we do not know what particular h_a is computed by the nervous system, we shall define an h_a with these three properties: it is a function that *could* be computed by neurons; it is monotonically related to blur; and it is mathematically tractable. Given $l(x, y)$ and $l'(x, y)$, the illuminance distributions on the retinas, an h_a satisfying the three conditions is defined by

$$-h_a = \iint_R [\nabla^{\nu} l(x, y)]^2 w(x, y)\, dx\, dy + \iint_R [\nabla^2 l'(x, y)]^2 w(x, y)\, dx\, dy. \quad [8]$$

In words, $-h_a$ is the square of the magnitude of the Laplacian of $l(x,y)$, weighted by $w(x,y)$, integrated over the retina, and summed for the two eyes.

The algebraic value of h_a as defined by Equation 8 increases when $l(x,y)$ is blurred and decreases when $l(x,y)$ is sharpened. Specifically, if $l(x, y)$ represents any luminance distribution on the retina and if $l_b(x, y)$ represents this distribution after it has been blurred, it is shown in Appendix A that

$$\iint_R [\nabla^{\eta} l(x, y)]^2\, dx\, dy \geq \iint_R [\nabla^n l_b(x, y)]^2\, dx\, dy, \quad [9]$$

where ∇^n represents any order of differentiation or integration applied to l.

The reason for choosing the Laplacian—rather than $l^2(x, y)$ directly, for example—is that it represents a good approximation to the spatial analysis performed by the retinal receptive fields (see Appendix A). The reason for using a square inside the integral of Equation 8 instead of the absolute value (as in the definition of h_v) is for the mathematical elegance of the theorems which follow therefrom. In fact, the same general properties hold for the absolute value as for the square (Appendix A).

Calculating $\partial e_a/\partial a$; How the visual system estimates h_a. To compute h_a as a function of $\triangle a$, the accommodation displacement, it

need only be noted that l is implicitly a function of a [i.e., $l = l(x, y, a)$]. Substitution in Equation 8 gives $h_a(\triangle a)$. This, however, leads to a problem.

In the model, $\partial e_a/\partial a$ controls the behavior of the marble. To know $\partial e_a/\partial a$ it is necessary to know $\partial h_a/\partial a$, and to estimate $\partial h_a/\partial a$ requires knowledge of at least two values of h_a. The problem is one of determining how the nervous system computes h_a for more than one value of a when the lens can assume only one value of a at any one time, for any one wavelength of light at any one point. These restrictions suggest three possible solutions: that the visual system obtains two (or more) values of h_a by determining a at two or more times, by determining a at two or more wavelengths, or by determing a at two or more points. Let us examine each of these solutions in turn.

The temporal solution is a trial-and-error solution. When the eye is confronted with a new stimulus, it would alter its present value of accommodation (a_1) to a new value (a_2, a trial value). The new value (a_2) might be an improvement or not, but no matter which, it would provide a second value of h_a, from which $\partial h_a/\partial a$ may be estimated by

$$\partial h_a/\partial a = [h_a(a_2) - h_a(a_1)]/(a_2 - a_1).$$

In such temporal trial and error, the initial change of accommodation in response to a new stimulus would be in error about half the time; subsequent changes of accommodation would be in the right direction. However, the first measurable accommodation response to a step or to a pulse displacement in the accommodative distance of a stimulus is generally in the right direction.[11] We must infer that at least two values of h_a are computed before accommodation has changed.

The chromatic aberration of the lens, which introduces a $\triangle v$ of about 1.3 diopters ($\triangle v \approx 4.5$ deg.) between the long and short wavelengths, provides a second possible way of realizing simultaneously two values of a on the retina. This cannot be the only mechanism, because some subjects make initially correct accommodation responses even in monochromatic illumination.[12]

[11] F. W. Campbell and G. Westheimer, Dynamics of accommodation responses of the human eye, *J. Physiol.*, 151, 1960, 285–295; E. F. Fincham, The accommodation reflex and its stimulus, *Brit. J. Ophthal.*, 35, 1951, 381–393.

[12] Fincham (n. 11).

The alternative to time and wavelength as means of obtaining different values of a is spatial variation of a. A mechanism for spatial variation of a is known to exist. In the neighborhood of the optic axis, the retina is not exactly perpendicular to the optic axis; thus, some parts of the retina are optically closer to the lens than others.[13] Comparison of two regions in different focus would provide an estimate of $\partial e_a/\partial a$ to even the stationary eye. By elimination, then, spatial variation of a is the means by which the relevant position of $e_a(a)$ is computed by the visual system. This complication requires: redefining $l(x, y, a)$ as $l[x, y, a(x, y)]$; calculating the predicted difference $\triangle h/\triangle a$ (rather than $\partial h/\partial a$) from Equation 8; and using two different weighting functions for the two terms of the difference [i.e., $\triangle h = h(w_1, l, l') - h(w_2, l, l')$]. These complications, though real, are not of great theoretical importance for the understanding of accommodation.

Whether h_a is known simultaneously for only two values of $\triangle a$ or for infinitely many values makes little difference. Only two values are needed to estimate $\partial e_a/\partial a$, and this estimate suffices to control the dynamics of a. However, when temporal information is used to estimate $\partial e_a/\partial a$ (and it occasionally may be), the situation can be vastly more complicated. For example, if it is assumed that there is a fixed time interval ($\triangle t$) between the two instants of sampling h_a, then the estimate depends on $\triangle a/\triangle t$, the average velocity of a during $\triangle t$. This leads to an additional term $[k_3(\partial e_a/\partial a)(-\partial a/\partial t]$ on the right-hand side of Equation 4 and converts a highly stable system into a highly unstable one, depending on how large k_3 is. Such a system would tend to have big, slow drifts in accommodation because when da/dt is small, the stabilizing force $\partial e_a/\partial a$ is small. Even small amounts of extraneous 'noise' would perturb a far from the minimum of e_a. In fact, this kind of 'd.c. drift' is a distinguishing characteristic of accommodation.

Demonstration of the multistability of accommodation. If the reader is young enough to possess a lens still capable of a few diopters of accommodation, he can easily demonstrate to himself two different stable states of accommodation in response to a single external stimulus configuration. First, cover one eye and establish the near point for accommodation by moving a pencil toward the open eye, while focusing on its point. (The near point is the small-

[13] G. L. Walls, *The Vertebrate Eye and Its Adaptive Radiation*, 1942, 7 and 255.

est distance at which a sharp image is possible.) Second, fixate a textured object (say, a picture) a few meters away. Third, while maintaining focus on the picture, move the pencil to the near point. Move it to various positions in front of the picture so that it obscures the point of fixation. Maintain focus on the picture. Fourth, focus on the pencil keeping the picture in the background. Then, alternate the third and fourth steps until you find a particular line of sight for the eye and a particular position of the pencil such that they on some occasions lead to accommodation of the picture and on other occasions to accommodation of the pencil. It is easy to focus either on the picture or on the pencil, but these changes of focus usually are consequences of changes in line of sight. The purpose of alternating the third and fourth steps is to find a single external stimulus configuration (same line of sight) that can yield either state of accommodation. The external stimulus (picture plus pencil) thus induces two different states of accommodation (accommodation on picture, accommodation on pencil); there are two stable response states to one stimulus (multistability).

In this demonstration, vergence of the covered eye changes with accommodation. The correlated vergence change is irrelevant to the demonstration; it means only that the demonstration may not reveal the dynamic parameters of accommodation when the covered eye exerts 'drag' on the changing value of accommodation.

Model of Fusion

When the eyes are at a given vergence angle (v), this vergence position defines a surface called the horopter. The defining characteristic of the horopter is that whenever a point lies on the horopter, it stimulates corresponding points of the two retinas.[14] In a small neighborhood of fixation, the horopter may be considered a plane, which is called the vergence plane, or plane of fixation. The horopter corresponding to v defines the optimum depth of an object for fusion of the object. When an object lies slightly in front of or behind the horopter, fusion is more difficult. When such objects are

[14] It is assumed here that the nonius method is used for the empirical determination of the horopter, as suggested by Ogle (n. 8). Also see K. N. Ogle, The problem of the horopter, in H. Davson (ed.), *The Eye*, 4, 1962, 325–348. By this method, there is no significant difference in the neighborhood of fixation between corresponding points (horopter) and corresponding coordinates (as defined early in this paper).

fused, they are seen in depth relative to objects on the horopter. This depth sensation is here called *fusion depth.*[15]

The model of fusion is formally equivalent to the models for vergence and for accommodation. First, in the fusion model the fusion distance (distance from eyes to plane of fusion) corresponds precisely to the vergence distance in the vergence model (distance from eyes to plane of vergence). Second, the perceived fusion depth in the fusion model corresponds to the achieved vergence angle in the vergence model. Third, the calculation of fusion disparity from the retinal stimulus (l and l') is the same as the calculation of vergence disparity, except that fusion is considered separately for each small retinal neighborhood (in vergence, the relevant region is the whole retina). Fourth, in the fusion model there is an internal bias to fuse objects in the plane of vergence (the corresponding bias in the vergence model is the bias to verge on the plane of accommodation). Fifth, the external stimulus energies and internal bias energies add to form the surface governing the fusion-depth plane just as the corresponding vergence energies add to form the e_v surface that determines the vergence plane.

Before developing the model for fusion in detail, it must be noted that when an object lies too far behind or in front of the horopter, fusion is impossible. Nevertheless, the object still may seem to be located at the appropriate depth plane. This depth sensation is here called *postfusion depth;* it corresponds to what is conventionally called 'qualitative stereopsis.' In the physical model of fusion phenomena, the perception of fusion and the perceived fusion depth are two different manifestations of the same underlying variable: the plane of fusion. By taking fusion depth and postfusion depth to be two different regions of a single continuum of perceived depth, the physical model for fusion depth (which extends over about $\pm\frac{1}{3}$ deg., or $\pm.03$ diopters, in the fovea) may be extended to include postfusion depth (about ± 2 deg. in central vision) without introducing any new principles. The complete, complex relations between perceived fusion, fusion depth, and postfusion depth are described in the neural model. Here we consider only the most basic phenomena and how the model applies to them.

[15] 'Fusion depth' corresponds approximately to what has been called 'patent stereopsis'; see K. N. Ogle, On the limits of stereoscopic vision, *J. exp. Psychol.*, 44, 1952, 253–259. Because of binocular rivalry, however, the range of patent stereopsis may vastly exceed the range of fusion depth—see the sections on binocular combinations.

Units of fusion displacement ($\triangle u$). The unit of vergence was defined as the angle between the lines of sight of the eyes. The vergence angle is also an unconventional measure of distance from which conventional distance can be calculated. The distance (d) from the midpoint between the observer's eyes to an object located directly in front of him is given by

$$d = \left(2 \tan \frac{u}{2}\right)^{-1} p \xrightarrow[u \to 0]{} p/u, \qquad [10]$$

where p is the interocular distance and u is the vergence angle of the object. Although d, the linear distance from the observer to the plane of fusion, is the conventional unit of distance, it is more convenient to designate the plane of fusion by u, its *angular distance* as defined by Equation 10. This scale of distance for u is commensurate with the scales for v and a, and it shares their advantages of being directly proportional to distance on the retina (see Figure 2d). When u is small, the tangent function can be removed from Equation 10, giving directly the reciprocal 'diopter' relation between d and u. In considering fusion, the distance of the horopter is v, the distance of the plane of fusion is u, and the distance from plane of fusion to the horopter is $\triangle u = u - v$. This follows the earlier sign convention for $\triangle v$; points beyond the horopter are represented by $\triangle u < 0$.

According to the model, whether a pair of retinal images (l and l'; an object) is seen as fused or unfused is determined by the fusion energy function $e_u(\triangle u)$ corresponding to that object, in the same general way that the vergence energy function $e_v(\triangle v)$ determined whether the eyes verged or did not verge on the object. The $\triangle u$ value of an e_u minimum determines the perceived fusion depth. Two components sum to give fusion energy: fusion-displacement energy (g_u) and image-disparity energy (h_u).

Fusion-displacement energy (g_u). Fusion-displacement energy (g_u) is given by a U-shaped function [$g_u(\triangle u)$], which governs whether or not fusion occurs. The function does not extend more than about $\frac{1}{8}$ deg. to either side, representing the fact that fusion is impossible for objects displaced more than about $\frac{1}{8}$ deg. from the horopter. The function $g_u(\triangle u)$ may also be interpreted as governing the depth sensation that results from successful fusion. In this case, it is useful to think of it as having an extension to $\pm$ 2 deg. as fusion depth ($\pm$ $\frac{1}{8}$ deg.) and postfusion depth ($\pm$ 2 deg.) form

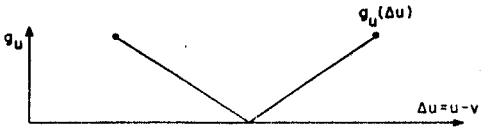

Fig. 6. Fusion-displacement energy, g_u, as a function of fusion displacement, Δu. The function $g_u(\Delta u)$ is defined only for $|\Delta u| < \frac{1}{8}$ deg. in central vision. The same function extended to $|\Delta u| < 2$ deg. describes the corresponding continuum of perceived depth (fusion depth plus postfusion depth).

a single continuum. The sharp minimum of $g_u(\Delta u)$ represents the fact that unfused objects (objects seen by only one eye) tend to be seen in the plane of the horopter; the sides of $g_u(\Delta u)$ are drawn straight (see Figure 6) mainly to help the reader distinguish g_u from g_v on graphs where both functions occur.

Fusion image-disparity energy (h_u). The contribution of h_u to e_u is defined as was the contribution of h_v to e_v, with one difference: whereas h_v was calculated by summing disparity over the whole retina, h_u is calculated in each small neighborhood (Δz) of the retina and summed over only those neighborhoods which cover the object. When several objects are present simultaneously in the visual field, $h_u(u, z)$ is calculated separately for each of them.

Examples of the model for fusion. The study of fusion in isolation requires the control of Δu, which in turn requires control of vergence movements. Some voluntary control of vergence is possible with the aid of a strong vergence stimulus (e.g., a picture frame within which other stimuli are presented). A better method is that of Fender and Julesz (n. 4), who studied the fusion properties of stimuli whose retinal position was stabilized by optical means (i.e., stimuli whose retinal position was independent of vergence). In general, they demonstrated that fusion had the same characteristics as vergence, although they did not analyze their data in this way. Their results can be summarized efficiently in the terminology of the model.

Fender and Julesz observed that a stabilized image pair could always be seen as a single object if the minimum of their $h_u(\Delta u)$ function occurred for $\Delta u < \frac{1}{2}$ deg. (compare with Figure 3) and could conditionally be seen as a single object (by slow separation of the images) for Δu in the range $\frac{1}{2}$ deg. $< \Delta u < 1$ deg. (compare with Figure 4). Their results show, as the model predicts, that the

range of the fusion depends on the retinal stimuli (l and l') used—particularly on the steepness of the sides of the minimum of the $h_u(\triangle u)$ function—since their most complex textured stimulus, which would give the deepest, steepest h_u minimum, also gave the greatest range of single vision. (The quantitative discrepancy between the range of single vision observed by Fender and Julesz and the range of fusion predicted by the model is explained in the section on predictions of the neural model, below.)

One interesting incidental observation of Fender and Julesz is the existence of various stable *'secondary' fusion points*. This too is predicted by the model. The secondary fusion points occur because a complex retinal stimulus (l and l') may yield an $h_u(\triangle u)$ function that has several smaller relative minima in addition to its major minimum. The minor minima result from chance superposition of unrelated but similar parts of the complex images. Such minor minima provide alternative fusion modes. In normal vision, saccadic eye movements produce a 'noise' perturbation of vergence ($\triangle v$). In normal vision, therefore, fluctuations in vergence from one eye movement to the next make it unlikely that the point of fusion could long remain in a small, shallow relative minimum, especially when it is near a major minimum of the $e_u(\triangle u)$ function. In stabilized vision, the eye movements are optically canceled and therefore secondary fusion points are more easily observed.

Panum's fusional area. In the case of vergence movements, 'vergence disparity' occurred because the minima of the $e_v(\triangle v)$ function did not coincide exactly with the minima of the $h_v(\triangle v)$ function. The same displacement phenomenon occurs in the fusion model when the minima of $e_u(\triangle u)$ function fail to coincide exactly with the minima of the $h_u(\triangle u)$ function. This means that when retinal fusion occurs for a stimulus that is not in the horopter, the apparent plane of the fused object is shifted slightly from the true stimulus plane and toward the horopter. The more important psychophysical correlate—insofar as the shift is detectable—would be blurring and then doubleness of vision of the object. Blurring might be observed even when the error in fusion depth itself is too slight to be measured.

The retinal range within which fusion is possible without noticeable blurring or double images is called Panum's fusional area. It traditionally has been measured with fine-line stimuli and, so measured, is remarkably small, about $\frac{1}{10}$ deg. (6 min. of arc) in foveal

vision.[16] We now examine the predictions of the model of the range of singleness of vision (Panum's area), of the range of fusion (which is larger than Panum's area), and of how these ranges depend on the nature of the stimuli used to measure them.

In the model, the correlate of doubleness of vision is the shift of a minimum of h_u caused by adding g_u to form e_u. (The analogous effect in the model of vergence produced 'vergence disparity.') We can make a quantitative statement. The displacement of the minimum of h_u is small when it is a sharp minimum (i.e., when the sides of h_u near its minimum are steep). This sharpness is determined by d^2h_u/du^2; the greater d^2h_u/du^2, the greater the range of singleness of vision. The range of fusion is determined by the maximum slope of h_u (max dh_u/du); the greater this quantity, the greater the range of fusion (see Figure 5).

We can say immediately that fine lines are poor stimuli for vergence and for fusion because the minimum of h_u is very shallow; large-area, fine-textured, high-contrast stimuli produce the deepest minima in h_u, and thereby the largest values of d^2h_u/du^2 and max dh_u/du. Therefore the physical model correctly predicts that large-area stimuli produce a greater range of singleness of vision and a greater range of fusion than do fine lines. It also predicts that the range of fusion (determined by max dh_u/du) may far exceed the range of singleness of vision (determined by d^2h_u/du^2). This prediction accounts in part for the paradox of the enormously large range of binocular fusion in casual observations (several degrees) and the tiny range of Panum's area ($\frac{1}{10}$ deg.). The difference between the fine-line stimuli used to measure Panum's area and the large-area stimuli encountered in normal vision is not the whole explanation, however, because binocular rivalry is involved. The problem is considered again in the section on predictions of the neural model.

THE PHYSICAL MODEL—WITH INTERACTIONS

Up to this point, the physical model has dealt with each system in isolation. The model for vergence considered a and u fixed; the model for accommodation considered v and u fixed; and the model for fusion that was just discussed considered v and a fixed. We are now ready to consider the interactions between the three mechanisms and to describe the operation of the complete model.

[16] Ogle (n. 8).

Interaction of Vergence and Accommodation

In Equations 1 and 2, both v and a occur in both terms on the right side of the equations: the displacement terms (g) contain a and v explicitly, and the image terms (h) contain a and v implicitly in l and l'.

The image-term interactions are relatively unimportant and may merely be mentioned. In Equation 2, vergence affects the position on the retina of the stimulus (l and l') to accommodation. Because the weighting function $[w(x, y)]$ is not a constant, retinal positions differ in the effectiveness with which they affect accommodation. Because vergence can shift l and l' relative to w, it can affect h_a slightly. Similarly, in Equation 8, the value of accommodation affects the shape of minima in the h_v function; blurred images have shallower minima than sharp images. In practice both of the effects of these image-term interactions are negligible compared to the effects of the displacement-term interactions.

The displacement-term interactions between vergence and accommodation are evident in Equations 1 and 2. In the accommodation-displacement term of Equation 2 $[g_a(\triangle a)]$, $\triangle a$ is defined in terms of both a and v, by $\triangle a = a - v$. Thus, the value of v determines the center of the accommodation-displacement function. Similarly, in Equation 1, the center of the vergence-displacement term g_v is determined by the value of a. These relations are incorporated into the model illustrated in Figure 7. This model consists of two templates (g_a and g_v), which slide sideways, parallel to each other, along rails to which they are attached. Two marbles (v and a, representing vergence and accommodation respectively) roll along the templates. A constant force (a spring, not gravity, in this model) drives each marble toward the 'bottom' of each template (i.e., toward the opposite rail). The v marble is held by a rod-and-spring device which is attached rigidly and perpendicularly to the center of the g_a template, so that when the marble moves, the g_a template slides horizontally by exactly the same amount. The a marble is similarly attached to the g_v template. Figure 7a shows the systems temporarily out of equilibrium, at the instant when a visual stimulus has been turned off; Figure 7b illustrates the equilibrium reached in the absence of a stimulus.

Figure 8ab illustrates the vergence and accommodation systems' response to a stimulus to convergence. In Figure 8a the stimulus has just been turned on and the systems are starting from their previous

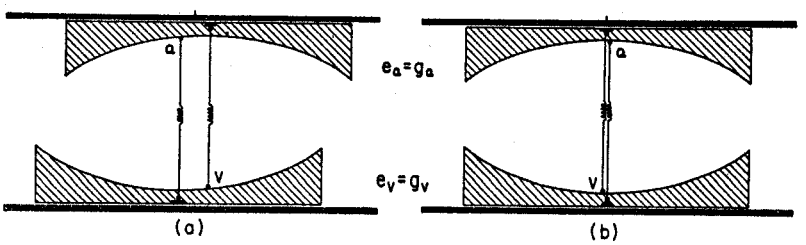

FIG. 7. Vergence-accommodation interaction in the dark. The marble representing vergence rolls on the g_v template, pushed toward the lower rail by a constant force from the spring; the g_a template, rigidly attached to the rod-and-spring device holding the v marble, follows the v marble exactly. The accommodation marble rolls on the g_a surface, pushed toward the upper rail by its spring; it carries the g_v template with it. The templates slide along the rails without friction. (a) The v-a systems temporarily out of equilibrium; (b) the final equilibrium. No visual stimulus to vergence or to accommodation is present here.

equilibrium position. Figure 8b illustrates the equilibrium position reached in response to this new stimulus. The image-displacement functions h_a and h_v are also illustrated in Figure 8. They remain fixed in the external world as the g_a and g_v templates slide. The e_a and e_v surfaces upon which the marbles roll are formed by adding the stationary hs to the sliding gs. The physical realization of the $g + h$ addition is a cumbersome, irrelevent mechanical problem; it is trivial in an electrostatic analogy.

Figure 8 also illustrates how accommodation helps vergence. If accommodation were fixed at the initial value, the vergence energy function e_v of Figure 8a would indicate a bistable vergence configuration. Starting from the position of Figure 8a, vergence on the stimulus would not occur (i.e., the marble would not leave its initial position). However, when accommodation is free to vary, it carries the vergence template (g_v) along with it until vergence also occurs.

Figure 8cd illustrates what happens when the stimuli to accomodation and vergence do not agree. This would happen, for example, if a person with normal vision, who had verged and accommodated on a stimulus, suddenly put on reading spectables with, say, a correction of +1 diopter ($\approx$3.5 deg.). Figure 8c illustrates the situation the instant the spectacles have been put on. His vergence is correct, but his accommodation is now too large by +1 diopter and, as Figure 8c illustrates, the a marble would start to roll into the e_a energy well. As it moves, it carries with it the g_v template. There

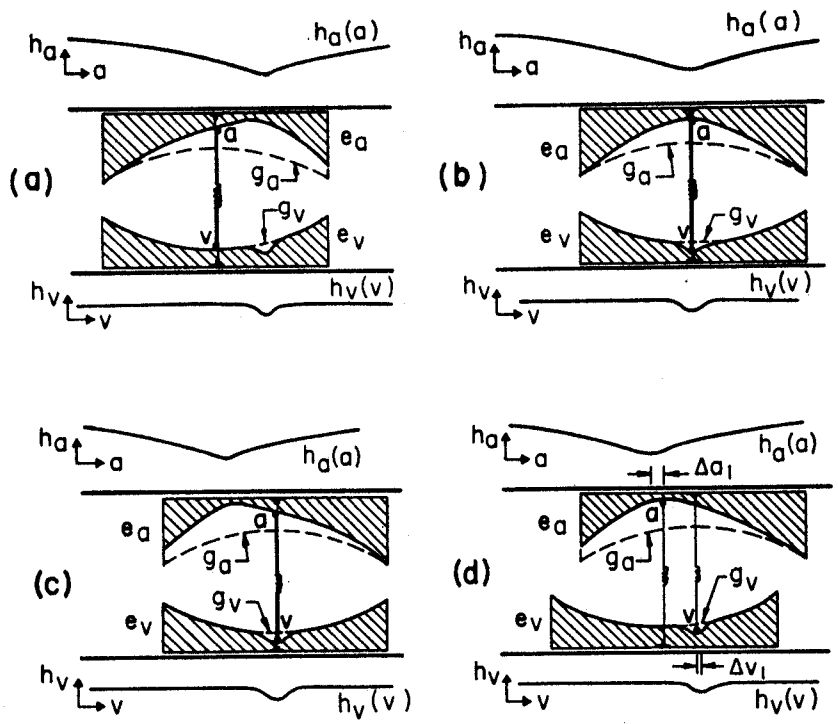

FIG. 8. Vergence-accommodation interaction. Vergence image-disparity energy, h_v, is shown below the vergence template's rail, and accommodation image-defocus energy, h_a, is shown above accommodation's rail. The image functions, h_a and h_v, add directly to the displacement functions, g_a and g_v, to form their respective energy functions, e_a and e_v (Equations 1 and 2). Here, g_a and e_a (but not h_a) are represented 'upside down.' The a and v marbles roll on e_a and e_v surfaces propelled outward by their springs; they carry the opposite templates with them, and in so doing cause the e_v and e_a surfaces to change shape continuously as g and h are added in different spatial relations. (a) The v-a systems (just previously accommodated and verged on a distant stimulus) the instant after a near stimulus has been turned on. (b) Equilibrium on the near stimulus. (c) Equilibrium on the near stimulus disturbed by sudden placement of a positive lens before each eye. (d) Equilibrium achieved in viewing the near stimulus through the positive lenses; 'accommodation disparity' is Δa_1, 'vergence disparity' is Δv_1.

is no initial 'resistance' of the g_v template to the movement because the vergence marble was resting at the bottom of an energy well where the slope was zero. At some point before accommodation reaches the bottom of the energy well in e_a, equilibrium is reached, as illustrated in Figure 8d.

The equilibrium position illustrated in Figure 8d has two interesting properties. First, the final equilibrium position may not be uniquely determined, because the marbles come to rest when they both lie on e_a and e_v slopes that are zero. This situation may occur for a significant range of possible values of a and v, with a correspondingly wide ambiguity of the 'equilibrium' values of a and v. Second, the equilibrium positions of a and v occur at the minima of e_a and e_v, and these minima are displaced from the corresponding minima of h_a and h_v. The displacement of the minima of e_v from those of h_v was previously noted to be the cause of 'vergence disparity.' A similar displacement of the minima of e_a from those of h_a is the cause of 'accommodation disparity.'

At the instant the +1 diopter of 'accommodation disparity' was introduced by the spectacles, the sum of the 'vergence disparity' plus 'accommodation disparity' was 1.0. As the overall system approaches the equilibrium position (Figure 8d), the total disparity is continuously reduced, but it can never quite reach zero; some 'vergence disparity' and some 'accommodation disparity' remain at the minima of h_a and h_v whenever $a \neq v$. The final equilibrium depends complexly on all of the relevant factors: the stimulus to accommodation, the stimulus to vergence, and the exact shape of the accommodation and vergence displacement functions.

Interaction of Fusion and Vergence

In the model, the effect of vergence on fusion is simple: vergence determines the horopter and hence the center of the g_u displacement function.[17] So far, however, the model does not contain a corresponding effect of fusion on vergence. The computation of image disparity, which is virtually identical for fusion and vergence, suggests a possible action of fusion on vergence.

The speed of vergence is restricted by the mechanical properties of the eyeball and its muscular apparatus. Fusion is not so restricted. Consequently, fusion of an object can occur before vergence upon it is complete—or, for that matter, before vergence has even begun. The model of vergence is based on $\partial e/\partial v$ (Equation 1), which is calculated in an infinitesimal neighborhood of v (Equation

[17] The effect of vergence on perceived depth is far more complex, and beyond the scope of this theory. See K. N. Ogle, Spatial localization through binocular vision, in H. Davson (ed.), *The Eye*, 4, 1962, 271–324; W. Richards and J. F. Miller, Convergence as a cue to depth, *Percept. & Psychophys.*, 5, 1969, 317–320.

6). The definition of e_v does not take any account of the information the fusion system may have about vergence possibilities up to ±2 deg. from v. Therefore, *a new vergence image-disparity energy function* (H_v) is proposed here, to incorporate fusion information.

In defining H_v, the h_v definition (Equation 6) is modified to give fused stimuli greater importance than unfused stimuli in determining vergence, by incorporating a fusion factor $[U(x, y, u)]$ into the defining integral:

$$H(v, U) = \iint_R [|l(x - v/2, y) - l'(x + v/2, y)| + U(x, y, v)]w(x, y)\, dx\, dy. \quad [11]$$

The fusion factor is $U = 0$ when fusion has occurred at the plane $v = u$ in a neighborhood of x, y; and $U > 0$ when fusion has not occurred. In general, $U(x, y, v)$ is assumed to vary with distance from the plane of fusion in roughly the same way that $g_u(\triangle u)$ varies with $\triangle u$ (Figure 6). The revised definition of H_v by Equation 11 changes nothing of fundamental importance in the treatment. It merely makes e_v a more accurate representation of vergence by incorporating an interaction of fusion with vergence.

Before concluding this section, it must be mentioned that horizontal vergence, and horizontal convergence in particular, can—to a much greater extent than vertical vergence, fusion, or accommodation—be controlled voluntarily with practice. Voluntary control is represented as a force, $\partial e_v/\partial v$; f_{vol} has been neglected in the theoretical treatment, but it must be considered, particularly in composing demonstrations that involve convergence.

Interaction of Accommodation and Fusion

The model contains an implicit image-term interaction of a and u; that is, failure to accommodate blurs an image and affects fusion of it. The model contains no provision for u to affect a, and there are no data to suggest that one is needed. Nor does the author know of any authentic three-term interactions (i.e., interactions not derivable from two-term interactions).

The Complete Physical Model: An Illustrative Example

Naturally occurring objects provide highly correlated stimuli to a, v, and u; they challenge neither the visual system nor the model.

However, images viewed in a stereoscope frequently not only challenge but defeat the visual system's attempts to cope with them. The model now undertakes to joust with one such stimulus.

Figure 9 illustrates a typical stereoscopic stimulus (L and L') and the intended depth illusion (seeing a raised disk above a rectangle). The illusion of stereoscopic depth is difficult to achieve when the rectangle and disk are separated by about 1 diopter ($\approx$3.5 deg.). Most observers readily can fuse that portion of the scene for which accommodation is appropriate—say, the rectangle —but they cannot at first fuse the disk. After some 'trying' (which consists of looking at various parts of the display), they suddenly are able to fuse the disk also. Once this fusion has occurred, they can look from the rectangle to the disk without disturbing fusion of either.

Figure 10 shows the model for the sequence of events just described. The complete model of Figure 10 differs from the model of Figure 8 only in that templates representing fusion have been added. The fusion templates are carried along with vergence, so that

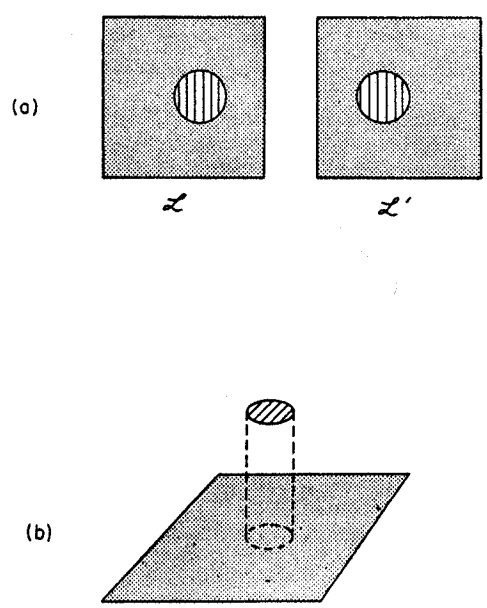

FIG. 9. A stimulus (a) to be viewed in a stereoscope; L and L' respectively are the images seen by the left and right eyes. (b) Perspective representation of percept induced by the stimulus in (a).

g_u always lies centered under g_v. The image-disparity functions $h_u(u, z)$ that add to g_u remain fixed, of course, and do not move. The two fusion templates illustrated represent two small areas of the retina, one in the central fovea, one in the periphery.

Figure 10a shows the starting position of the model when the picture is turned on. Immediately before, the observer has been looking at the frame of the stereoscope. Accommodation and vergence therefore are appropriate for the background rectangle. The disk, which falls on the fovea, is unfused and the eyes are not verged on it. The vergence image disparity (h_v) is indicated according to its original definition in Equation 6. The weighting function (w) assigns less weight to the peripheral stimulus (rectangle) than to the foveal stimulus (disk). The 'unweighting' of *peripheral* image disparity is confined to vergence and does not occur in the h_u fusion image-disparity function.

When confronted with unfused foveal images and fused peripheral images, the eyes usually turn to fixate the fused image. Figure 10b illustrates this situation, with the fovea pointed to one side of the rectangle. There now are excellent stimuli to accommodation, vergence, and foveal fusion; only peripheral fusion of the disk is unsatisfied.

Eventually, as the observer attempts to fuse his entire visual field, his eyes return to fixate the disk and he has returned to his initial state (Figure 10a). To fuse the disk, the eyes first must converge. Convergence may be triggered in numerous ways. In terms of the model, these ways are methods of getting over the 'hump' between the two minima of e_v, or of eliminating it temporarily. In some cases, convergence may be delayed even when there is no hump, if it is slowed by a very shallow slope of e_v. In these cases, the observer need only keep looking at the unfused disk and his eyes will slowly verge on it. Alternatively, a saccadic eye movement may cause a spontaneous fluctuation in vergence that is large enough to bring vergence over the hump in the e_v function. Or, the eyes may find a location to fixate that produces a particularly strong vergence stimulus (a minimum of minima of h_v), which eliminates the e_v hump. Voluntary or spontaneous vergence-hunting motions may occur. Vergence may also be stimulated artificially, by placing negative lenses before the eyes; or it may be simulated by placing prism(s) before the eye(s).

For whatever reason, when the vergence marble finally arrives at the edge of the disk's e_v energy well, it rolls into it. Fusion of the

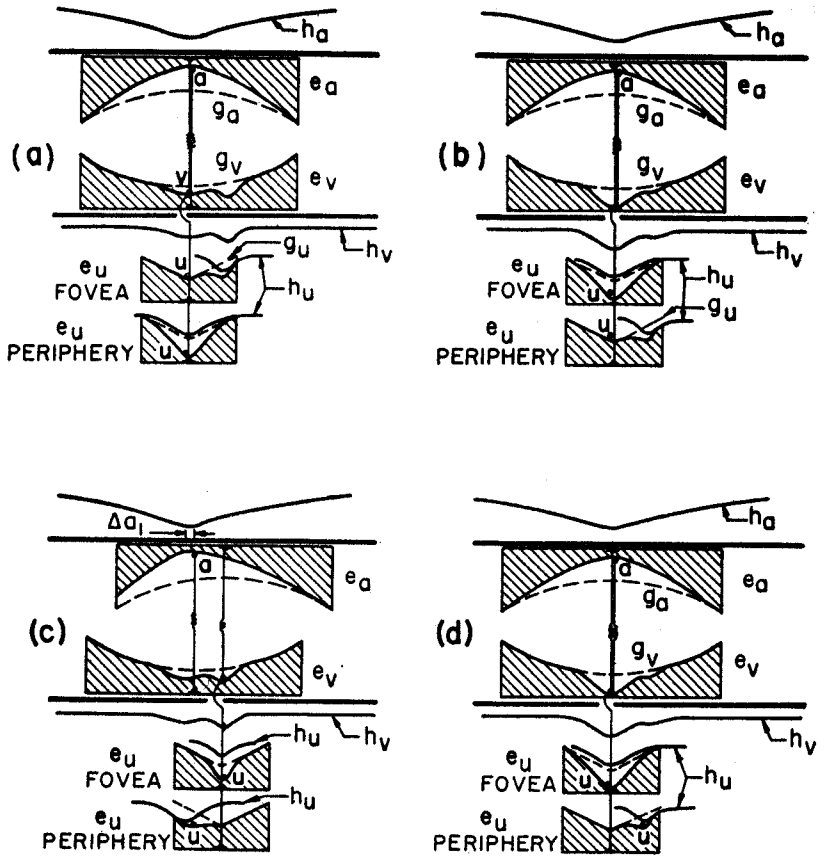

FIG. 10. Complete model of vergence, accommodation and fusion. The g_u fusion templates are attached to the v marble and move with it (as does g_a). The fusion image-disparity energy function $h_u(u, z)$ remains fixed. It adds to $g_u(u)$ to produce the e_u surface upon which the u marble rolls. A constant force (mechanism not indicated) propels the u marble downward. Each g_u template represents a different small area of retina. The minimum of h_u is drawn narrow for the disk, wide for the rectangle (see Figure 9). The h_v function is the sum of h_u functions, with the peripheral h_u reduced by $\frac{1}{2}$. (a) Initial viewing of the images of Figure 9a, fixation on disk: vergence and accommodation are appropriate for rectangle; accommodation is appropriate for disk also, but vergence is not and the disk is not fused (v and foveal u marbles are not in energy wells of the disk). (b) Fixation on rectangle: the disk, which still is unverged upon and unfused, now is in periphery. (c) Fixation on disk after vergence upon and fusion of disk: note that peripheral fusion of rectangle is maintained; note 'accommodation disparity' Δa_1, and negligible 'vergence disparity.' (d) Fixation on rectangle after (c): note that peripheral fusion of disk is maintained.

disk occurs momentarily before the equilibrium position of vergence on the disk is reached (Figure 10c). In verging to equilibrium, fusion of the rectangle may remain undisturbed because of the relatively slow vergence movement, although nonfusion of the rectangle is now an alternative stable state.

Figure 10d illustrates the equilibrium state of the system when fixation is transferred to the rectangle. In the model (and presumably in the observer), accommodation and vergence when both disk and rectangle are fused are the same as when only one or the other was fused (compare Figure 10a and 10d). Even the fusion energy function e_u is the same. The only difference is that in Figure 10d the fusion marble has attained the less accessible one of the two energy wells of the peripheral e_u function.

Once fusion of the disk has been established, it is possible to look back and forth from disk to rectangle and to verge on each. In the model, this requires that fusion of the disk, when it occurs, remove the hump in the e_v vergence functions of Figure 10a. This facilitating effect of fusion on vergence is provided for in the revised definition of vergence image-disparity energy (H_v), but it has not been indicated in Figure 10.

THE NEURAL MODEL

The physical model is formally the same for vergence, for accommodation, and for fusion. However, the neural mechanisms underlying these three systems differ from each other. The accommodation mechanism is essentially monocular; therefore, accommodation itself is of interest here mainly in terms of its interaction with vergence. (A neural mechanism of accommodation is proposed in Appendix A.) On the other hand, fusion and vergence are essentially binocular processes; both require the interaction of information from both eyes. The proposed neural process is basically the same for vergence and for fusion; the main difference being that fusion is computed separately in each region of the retina whereas vergence signals are summed over the whole retina. Thus, fusion serves as the kernel of the vergence system. In addition, fusion involves binocular rivalry and other interesting phenomena. The purpose of the neural model is to propose an analysis of the component processes of fusion; the extension to the processes of vergence is merely outlined.

The analysis is presented in terms of the functions performed by

hypothetical neurons. The hypothetical neurons resemble real neurons, but whether they resemble neurons of the visual system is an unanswered question. The analysis is entirely functional, so any exact resemblance would be fortuitous. Although there are important new reports of electrophysiological recording from binocularly sensitive neurons in the brains of cats and monkeys,[18] the neurophysiological data presently available are too fragmentary to provide either confirmation or disproof of the proposed model.

The core of the proposed neural mechanism is a *neural binocular field* (*NBF*) that performs the kind of binocular comparisons described by Equation 7. The use of a binocular-comparison field as an explanatory concept for depth perception can be traced to Kepler;[19] neural fields of this general kind are also familiar in other contexts.[20] Psychophysically, the existence of binocular point-by-point comparison has been amply documented by the demonstration of binocular depth perception in the absence of monocular cues.[21]

As diagrammed in Figure 11, the neural binocular field (NBF) is an internal three-dimensional representation of the external world. The x, y plane represents the dimensions perpendicular to the observer's line of view; the z dimension represents depth. At each level in the NBF, signals originating from the left and right eyes are represented with different translations relative to each other. Signals originating from corresponding points of the retinas intersect in the middle of the field; signals originating from points displaced temporally in the two retinas (i.e., signals produced by an object nearer to the observer than his fixation point) intersect below the middle of the field; signals from a more distant object intersect at the top of the field.

[18] H. B. Barlow, C. Blakemore, and J. D. Pettigrew, The neural mechanism of binocular depth discrimination, *J. Physiol. (London)*, 193, 1967, 327–342; T. Nikara, P. O. Bishop, and J. D. Pettigrew, Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex, *Exp. Brain Res.*, 6, 1968, 353–372; D. H. Hubel and T. N. Wiesel, Cells sensitive to binocular depth in area 18 of the macaque monkey cortex, *Nature*, 225, 1970, 41–42.

[19] See reviews in E. G. Boring, *The Physical Dimensions of Consciousness*, 1933 (reprinted 1963), and in L. Kaufman, On the operations underlying stereoscopic combination, SP-66-1, 1966, Sperry Rand Research Center (Sudbury, Mass.).

[20] W. S. Pitts and W. McCulloch, How we know universals, *Bull. math. Biophys.*, 9, 1947, 124–147.

[21] B. Julesz, Binocular depth perception of computer-generated patterns, *Bell Syst. tech. J.*, 39, 1960, 1125–1162.

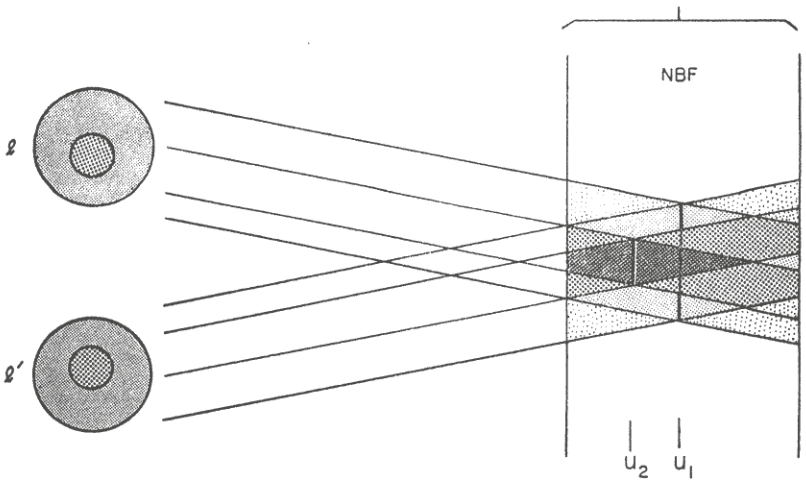

FIG. 11. Diagrammatic illustration of the Keplerian theory of stereopsis. A cross section of the neutral binocular field is indicated by NBF, at the right. Neural signals originate from images on the left and right retinas (l and l' respectively) and pass through the NBF along the straight lines drawn from l and l' to the NBF. When a region of the field receives inputs from both l and l', these inputs interact. Regions of interaction of the center and outer disks respectively are indicated by heavy and by medium stippling. Vertical lines, u_1 and u_2, indicate levels of the NBF at which exact binocular correspondences occur between the signals from outer and inner disks respectively of l and l'. These two different levels represent different distances from the observer (depth). The figure is so drawn that the direction of signal flow can be reversed. In this case, levels of binocular correspondence u_1, u_2 in the NBF are interpreted as cross-sections of external stimuli. The signals (light) flow from right to left to produce the images l and l'. This method of representation demonstrates how, in Keplerian theory, the contents of the NBF are a simple reflection of the external world.

Up to this point, the model reiterates the general ideas of Koffka, Charnwood, Linksz, Julesz, Dodwell and Engle, Kaufman, and many others.[22] The model presented here embellishes on its predecessors in three respects. First, it contains two NBFs: a primary NBF for the fine-detail functions of binocular vision, and a sec-

[22] K. Koffka, Some problems of space perception, in C. Murchison (ed.), *Psychologies of 1930,* 1930, 161–187; J. R. B. Charnwood, *Essay on Binocular Vision,* 1951; A. Linksz, *Vision, Physiology of the Eye,* Vol. 2, 1952; P. C. Dodwell and G. R. Engle, A theory of binocular fusion, *Nature,* 198, 1963, 39–40, 173–174; L. Kaufman, Some new stereoscopic phenomena and their implications for the theory of stereopsis, this JOURNAL, 78, 1965, 1–20.

ondary NBF for coarse-detail functions. The interaction of these two systems is crucial for many phenomena of binocular depth perception. Second, the model analyzes binocular fusion into elementary subfunctions, performed by hypothetical neurons. Third, it accounts for fusion and for rivalry within the same system and in the same neurons; it suggests how neurons that yield binocular fusion in some circumstances can yield binocular rivalry in others.

Primary Neural Binocular Field

Inflow neurons. Figure 12 illustrates the four kinds of neurons proposed to constitute the primary NBF. The first of these are the inflow neurons (Figure 12a), which carry signals that originate from the left and right retinas. Their axons cross the NBF at opposite slants. Inflow signals are carried by their axon terminations to all depths in the field. The inflow neurons are assumed to con-

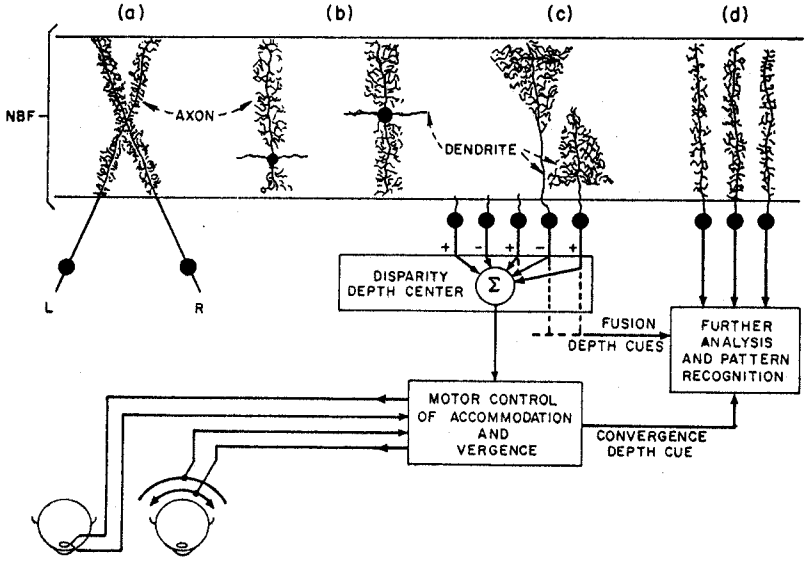

FIG. 12. The neurons of which the NBF is assumed to be composed and the operations subsequently performed on their outputs. (a) Inflow neurons, carrying inputs from left and right eyes. (b) Binocular correspondence-detecting neurons (BCDNs). (c) Level-detecting neurons. (d) Outflow neurons. The motor control of the lens (accommodation) and of the rectus muscles (vergence) is indicated at lower left.

nect to all other neurons in the NBF. Figure 12 is drawn so that the number of synapses (connections) between axons (outputs) of one neuron and dendrites (inputs) of another neuron in a given small area is proportional to the product of the density of axon terminations and dendrites within that area.

Binocular correspondence-detecting neurons (BCDNs). A BCDN- (Figure 12b) receives its input from only one level of the NBF and sends its output to all other levels. The function of the BCDNs is to limit outflow from the NBF to just one particular level when a binocular correspondence exists at that level. A BCDN receives inputs from inflow neurons, and it is silent unless its input indicates a correspondence between the left and right retinas. When its input does indicate correspondence, it produces a large output; this output acts to potentiate outflow at its own level and to inhibit BCDNs on all other levels. The precise nature of the inputs is considered below, in the section on rivalry. The transfer of signals from inflow neurons to outflow neurons is assumed to be possible only at the level that is not inhibited by BCDNs. Thus, within any small z column of the NBF, outflow from the NBF occurs only from the one level that has active BCDNs (monoactivity).

The monoactivity of a column of BCDNs depends on their gain and their threshold. The gain is assumed to be large; that is, in isolation a BCDN would produce a large output in relation to its input. When the BCDNs on one level have a significantly larger output than those on any other level, the BCDNs of this level silence those of all the other levels. The silencing occurs because of implicit positive feedback. As the BCDN with the largest input begins to inhibit the other BCDNs, they in turn inhibit the original BCDN less; it increases its output still further, which in turn acts to diminish other outputs still further, and so on, ad infinitum. When the most active BCDN succeeds in suppressing other BCDNs to below their threshold, it no longer receives inhibitory signals from them; it then acts as though it were in isolation. This monoactive interaction permits only one level in any vertical column of the NBF to be active at one time. Appendix B describes and analyzes a neural network that has these properties; in particular, it demonstrates that when neurons are appropriately interconnected, their normal physiological characteristics are adequate to produce the monoactive interaction.

If it is assumed that BCDNs in the middle level of the NBF have

a slightly lower threshold (or equivalently, receive a slightly higher level of spontaneous activity) than other BCDNs, then when only one eye is stimulated, the middle level is the only active level. The bias in favor of the NBF's middle level corresponds to the g_u function of the physical model for fusion; recall that $g_u(\Delta u)$ expressed the tendency for fusion to occur at $u = 0$ (middle level) in preference to other depth planes.

In general, the computation performed by the BCDNs corresponds to the $l - l'$ comparison operation of the h_v and h_u image-disparity functions (Equation 6) of the physical model. Different levels of the NBF correspond to different values of Δv and Δu at which $l - l'$ is computed. The active level of the NBF corresponds to the plane of fusion of the physical model; the outflow from this level corresponds to the subjective binocular image.

Level-detecting neurons. The third kind of neuron detects the active level of the NBF, that is, the depth plane at which binocular correspondence occurs. The function of the level-detecting neuron could be combined with that of the BCDN, but I have chosen to designate it separately. Level-detecting neurons (Figure 12c) receive their input from BCDNs. The number of connections between a BCDN and a level-detecting neuron increases with the BCDN's distance from the center of the NBF, so that a binocular correspondence at an extreme level of the NBF produces a big input to the level-detector and hence a big output.

The outputs of level-detecting neurons are evaluated at a disparity-depth center. Level-detecting neurons that receive signals primarily from the bottom of the NBF feed excitatory signals to the disparity-depth center; neurons from the top of the NBF feed inhibitory signals. Thus the inputs to the disparity-depth center indicate the depth levels of binocular correspondences.

Information about the fusion depth of each small area of the NBF is passed through the disparity-depth center to subsequent stages for further processing. In addition to transmitting local depth information, the disparity-depth center sums the combined outputs of the level-detecting neurons over the NBF. This summed output signal (plus the equivalent output of the secondary NBF) is used to control vergence; it is the net contribution of retinal disparity to the control of vergence. In the physical model, the image component of the *force* that controls vergence is $\partial h_v/\partial v$ (the image's contribution to the slope of the e_v surface under the ver-

gence marble); this force corresponds to the summed output of the disparity-depth center.

As defined, the disparity-depth center produces a signal inversely proportional to depth, that is, a convergence signal. As a rule, the nervous system provides functions in paired opposites; therefore, a more complete diagram would include an oppositely signed depth center to provide a divergence signal. The factor g_v of the physical model, which represents motoric and all influences on vergence other than disparity cues, is implicitly represented in Figure 12 by the box labeled Motor Control.

Outflow neurons. The outflow neurons (Figure 12d) continue the analysis of shape, texture, and color in the NBF. These neurons are organized perpendicularly to the NBF and they receive inputs equally from all depths. Their input is significantly simplified by the BCDNs, which limit the inflow-outflow transfer to a single level of the NBF. If it were not for BCDNs, the outflow neurons would receive the summed signal of various planes of comparison, a blur of useful and useless information. The outflow neurons carry no binocular depth information. At subsequent stages of analysis, however, binocular depth information (from the level-detecting neurons) is recombined with the shape, texture, and color information of the outflow neurons.

We may note here that the neurons which control accommodation are not represented in Figure 12. To provide the h_a defocus signal, any combination of contour-sensitive neurons (e.g., on-center, off-surround) whose output was added at an 'accommodation center' would suffice (see Appendix A).

'Binocular-rivalry neurons': BCDNs I–IV. When antagonistic stimuli are presented to the two retinas (e.g., l = horizontal, l' = vertical stripes), then in some areas of the visual field an observer sees the stimulus from one eye; in the remaining areas he sees the stimulus from the other. For many stimulus pairs, there are no areas of the field in which stimuli from both eyes are seen simultaneously. This 'either-or' interaction between stimuli to the left and right retinas is called binocular rivalry, a phenomenon so closely related to binocular depth perception that one can hardly discuss one without the other. Here, the features of the NBF are elaborated to account for these rivalry interactions.

Only one generic type of BCDN has so far been proposed, to detect binocular correspondences in the NBF. Here it is further pro-

posed that the NBF has *four* generic types of BCDNs, organized as two pairs, and that reciprocal interactions between these pair-mates on the *same* level of the NBF account for binocular rivalry. (The interactions between BCDNs on different levels are as described above: an active level suppresses all other levels, and this interaction accounts for fusion and depth phenomena.)

To illustrate the four BCDN types, a nomenclature for describing neuronal inputs is helpful. Let F represent the function transmitted by an input, and let $\sim F$ represent the complementary function. For example, F may represent input to the BCDN from an on-center, off-surround receptive field as described by Kuffler and by Hubel;[23] $\sim F$ then represents an off-center, on-surround receptive field. Other examples of antagonistic receptive-field functions would be a vertical light-dark boundary (F) and a dark-light boundary ($\sim F$), a vertical bar (F) and a horizontal bar ($\sim F$), and so forth. Further, let L represent an input from the left eye and R an input from the right eye. Let + represent an excitatory input connection and − an inhibitory input connection. A BCDN is inactive when it receives no input signals or when it receives only inhibitory signals. When it receives a signal through an excitatory input, it becomes active. If both excitatory and inhibitory signals are received, the activity is less than that produced by only excitatory signals.

Table I lists the four kinds of BCDNs for a particular function (F). Types I and IV are F detectors; they are active when the eyes receive a stimulus satisfying the function (F)—in this example, a white spot on a dark surround. When only the left eye receives the

TABLE I

INPUTS OF THE FOUR TYPES OF BCDNs

Type	Eye	Message	Input
I	L	F	(F, L, +) and ($\sim F$, R, −)
II	L	$\sim F$	($\sim F$, L, +) and (F, R, −)
III	R	$\sim F$	(F, L, −) and ($\sim F$, R, +)
IV	R	F	($\sim F$, L, −) and (F, R, +)

Pair A, BCDNs I and III; Pair B, BCDNs II and IV.

[23] S. W. Kuffler, Discharge pattern and functional organization of the mammalian retina, *J. Neurophysiol.*, 16, 1953, 37–68; D. H. Hubel, Integrative processes in central visual pathways of the cat, *J. opt. Soc. Amer.*, 53, 1963, 58–66.

stimulus, Type I is active; when only the right eye receives the stimulus, Type IV is active. Types II and III are the corresponding $\sim F$ detectors.

All four types of BCDNs are assumed to be present in every small region of every level of the NBF. The function of the BCDNs is to detect binocular correspondences and then, by means of strong inhibition of BCDNs at all other levels, to restrict outflow from the NBF to the level at which the correspondence occurred. The four types of BCDNs are assumed to *enable specific* inflow-outflow connections: left-eye BCDNs (Types I and II) enable transfer of information from left-eye inflow neurons to outflow neurons, and right-eye BCDNs (Types III and IV) enable right-eye inflow-outflow transfer.

The four BCDN types are organized into *strong reciprocally inhibiting* pairs within any given level of the NBF. Two kinds of BCDN pairs are proposed here: *A* pairs, composed of BCDN Types I and III; and *B* pairs, composed of BCDN Types II and IV. The two partners in each pair receive identical inputs except that excitation and inhibition are reversed. Within a level the two members of a pair interact in basically the same way as two different levels interact: a member is fully active or silent, and only one member of the pair can be active at one time. The active member is usually the member with the larger input, although previously established activity may persist even when the inputs are changed. When neither member has an input that exceeds the threshold (ϵ), then both members of the pair are silent.

Finally, it is assumed that—in addition to their strong reciprocal interconnections with their partners in a pair—the four kinds of BCDNs have similar but much *weaker reciprocally inhibiting* interconnections with neighboring opposite-eye neurons on the same level. For example, BCDN I inhibits neighboring BCDN IIIs and IVs weakly, in addition to strongly inhibiting its BCDN-III partner. The effect of this interaction is facilitory, in the sense that as soon as one member of one pair has achieved dominance, it helps to silence ambivalent opposite-eye neighbors and thereby to facilitate the quick spatial spread of dominance of BCDNs of the same eye throughout that level of the NBF.

To illustrate the operation of BCDN pairs, we consider first the case where both eyes receive the same stimulus. Then, one BCDN of each pair receives an excitatory input, and the other member receives an inhibitory input. For example, when the stimulus on both

retinas satisfies F (when it is, say, a bright spot on a dark background), the left-eye input of BCDN I (which has an excitatory connection) is excited, and the right-eye input of BCDN I (which has an inhibitory connection) is inhibited and carries no signals. Therefore BCDN I is excited. The excitatory input of BCDN III is silent (it requires a dark spot), and the inhibitory input is active; therefore, BCDN I is active; BCDN III is inhibited. In Pair A then, BCDN I dominates BCDN III. Similarly, in Pair B, BCDN IV dominates BCDN II. The active BCDN I's enable left-eye inflow-outflow connections in the neural field, and the active BCDN IVs enable right-eye inflow-outflow connections; all inflow-outflow connections are enabled. And when the stimulus on both retinas satisfies $\sim F$ (when it is a dark spot on a light surround), BCDN II dominates I and BCDN III dominates BCDN IV; again, all inflow-outflow connections are enabled.

When the stimulus to one retina satisfies F and the stimulus to the other retina satisfies $\sim F$, however, then only one set of inflow-outflow connections is enabled. For example, suppose that the left eye receives a bright spot and the right eye receives a dark spot. BCDN I receives excitatory signals from the left eye and inhibitory signals from the right eye. BCDN III has the opposite inputs. Which of BCDN I and III dominates Pair A depends on the relative strength of the two stimuli, on past stimulation, and on bias factors. However, only one member can be active at one time. If BCDN I is the active member of Pair A, then the left eye's inflow-outflow connections are enabled; if BCDN III is active, the right eye's inflow-outflow connections are enabled. Normally, Pair B would enable the connections of the eye that was not enabled by Pair A. In the case of antagonistic stimuli ($L = F$; $L' = \sim F$), however, both members of Pair B (both BCDN II and BCDN IV) fail to receive excitatory input signals. Therefore neither member of Pair B is active, and only one eye's inflow-outflow connections are enabled—those of the eye chosen by Pair A.

Secondary Neural Binocular Field

The secondary NBF is assumed to be functionally similar to, and composed of the same types of neurons as, the primary NBF, the main difference being that the sizes of the receptive fields of its input neurons are about 10 to 20 times greater. The depth range of the primary NBF is estimated to be about $\pm\frac{1}{8}$ deg., that of the

secondary NBF about ±2 deg. These depth ranges are very approximate characterizations of central vision; in both NBFs the depth range increases rapidly with peripheral angle.[24] The depth signals from the primary and secondary NBFs add in proportion to the magnitude of the depth; thus the primary NBF signals are merely a perturbation on the larger depth signals from the secondary NBF.

The fine-detail output of the primary NBF contains information about stimulus shape and texture, information which is used in subsequent decoding of these stimulus properties, such as in pattern recognition. The outflow of the primary NBF ultimately determines the contents of perception—*what* we see. The coarse-detail output of the secondary NBF—except for the depth signal—does not serve pattern recognition so directly. Metaphorically, we may say the secondary NBF is concerned with the metalanguage of perception; it directs recognition processes to particular regions of the visual field, but it is the outflow of the primary NBF that is processed. Because the information dealt with by the secondary NBF is so coarse, the secondary NBF computes binocular correspondence between stimuli that would be treated as different by the primary NBF. For example, at reading distance, the primary NBF clearly discriminates between the letters A and B; the secondary NBF treats them as identical 'blobs.'

Predictions of the Neural Model

There are two kinds of outputs from the primary and secondary NBFs that can be related to behavioral data: the *depth signal,* which derives from both NBFs, and the *pattern outflow,* which derives from the primary NBF. Below, the depth signals for fusion and vergence derived from the neural model are compared with those derived from the physical model; pattern outflow is considered next; and the combinations of depth signals and pattern outflow are considered last.

Depth signals: Fusion. The active level of an NBF corresponds to the level ($\triangle u$) at which fusion occurs in the physical model; a signal indicating this level is transmitted by the level-detecting neurons (Figure 12c). The computation of binocular correspondence

[24] Ogle (n. 17).

by the BCDNs corresponds to h_v (and h_u) in Equation 6 of the physical model. The layer of BCDNs with greatest binocular correspondence corresponds to the fusion value of the h_u function in the physical model, and the spontaneous input to the middle layer of the NBF corresponds to the g_u bias function. The addition of g_u and h_u to form e_u in the physical model is represented in the neural model by the combination of the spontaneous and stimulus-produced inputs to the BCDNs; the particular energy well of e_u into which the marble representing fusion rolls depends on past stimuli and responses, just as does the particular active layer of the NBF.

The multistability of fusion in stabilized vision is the work of the BCDNs. Activity at any level of an NBF—once it is established—inhibits activity at any other level (monoactivity). The normal mechanism for establishing activity at a new level is a sudden change in the entire input, such as that caused by a saccadic eye movement or a blink, which in effect resets the NBF to a neutral starting position and allows the level which contains the strongest stimulus to establish itself.

In stabilized vision, activity can be established at a level (u) far from the center of the primary NBF by slowly shifting a fused L and L' laterally. Sooner or later, the level of maximal correspondence ($\triangle u$) will contain a shallower energy well than that produced by the bias in favor of the center level, but fusion at $\triangle u$ continues to predominate until a sudden change causes a reevaluation of stimulation at all levels (see Figure 10d).

The physical model for fusion assumes that the image disparity of all possible fusion depths (u) is computed. The marble representing fusion rests in a particular energy well because of its past history. The neural suppression of one level in the NBF by another is equivalent to the suppression of certain fusion computations, that is, to the elimination of all but one of the energy wells—the one in which the marble rests. This suppression does not grossly disturb the isomorphism between the neural model and the physical model because the marble representing fusion can rest in only one energy well at one time, and therefore the presence or absence of the others in the neural model is not of first-order importance.

Depth signals: Vergence. The signals for controlling vergence are the summed depth signals from the primary and secondary NBFs, weighted according to the signals' retinal location, with

foveal values having the greatest weight. The summing and weighting is carried out at the disparity-depth center (Figure 12). This signal (h_v) is combined with the motor-bias component (g_v) to give the net vergence-control signal (e_v).

The multistability phenomena illustrated in Figure 4 (that the eyes may fail to verge on a stimulus when it is presented suddenly, although they can verge on it when it is produced by slowly transforming a previously verged stimulus) result from the combination of motor-bias forces (g_v) with image-disparity forces (h_v). However, complete failure of vergence (Figure 5) results when the disparity of the stimuli to the left and right retinas exceeds that for which binocular correspondence is computed in the NBFs. This and other interactions of fusion with vergence were subsumed in the physical model by Equation 12. Vergence failure usually does not occur outside the laboratory because the link between accommodation and vergence provides a vergence stimulus for naturally occurring objects.

Pattern outflow: Binocular rivalry. The neural model predicts that when stimuli to the two retinas are in correspondence, the outflow from the primary NBF contains the total pattern message from both eyes. When the retinal stimuli are not in correspondence, the outflow from the primary NBF contains messages only from one eye; the chosen eye depends on the stimulus configuration (past and present), on ocular dominance, and on accident. For a given pair of rivalrous stimuli, the condition of one eye being dominant is a stable condition, which, in the model, would persist indefinitely until a perturbation such as a blink or an eye movement changed the inputs.[25] Insofar as activity tends to fatigue BCDNs, fatigue as well as perturbations of input would contribute to alterations of dominance.

By assuming that the function (F) used by BCDNs in the primary NBF is on-center and off-surround, many additional facts of binocular rivalry are accounted for. For example, when, in a given retinal neighborhood, one stimulus contains contours and the other does not, the stimulus with contours dominates. This observation is predicted by the model because F is a contour-sensitive function; the contour input to the BCDNs enables inflow-outflow transfer in

[25] N. J. M. Levelt, Note on the distribution of dominance times in binocular rivalry, *Brit. J. Psychol.*, 58, 1967, 143–145.

its eye, whereas the noncontour input to the other eye does not enable transfer.

The model further predicts that a contour stimulus (say, a ring) presented to one eye enables inflow-outflow from the same eye not only in the neighborhood of the contour but also for the whole inside of the ring. It predicts this even for rings whose radius exceeds the range of lateral BCDN-pair interactions—see the paragraph on weak BCDN interactions—because lateral spread is a self-propagating phenomenon when no resistance is encountered (i.e., a rivalrous visual stimulus). Levelt summarizes many observations to prove that common contours enable binocular summation in the area within the contours.[26] For example, when a black disk on a white surround is presented to the left eye and a uniform field, white or dark, to the right eye, the disk dominates and is seen unaltered. But when a disk is presented to the left eye and a ring to the right eye, a disk is seen and its observed brightness is intermediate between that of the presented disk and that of the field within the ring. In the neural model, inflow-outflow is enabled within both the disk and the ring, so that the net outflow is the combination of both inflows, dark and light. The model does not specify how subsequent stages respond when they receive messages from the primary NBF that an area is both light and dark, but gray—the average—is a reasonable solution.

The model predicts in some detail what is seen in such complex stimuli as horizontal stripes to one eye and vertical stripes to the other. There are two areas at the intersection of the edges of the stripes where the BCDN pairs receive conflicting inputs from each eye (see Figure 13). Once one retinal stimulus (l), say, the vertical stripes, has dominated the other in one boundary conflict, lateral interaction tends to cause dominance of that stimulus to spread throughout the whole stripe and into any adjacent stripes. If the stripes are wide enough, however, the contours of horizontal stripes (l') from the other eye can establish little 'islands' of dominance within the vertical stripes (Figure 13). Brightness signals from the two eyes are combined interior to these islands because both BCDN pairs are active. Where l and l' present conflicting contours to the two eyes, only one member of one BCDN pair is active. Figure 13 indicates these areas of rivalry.

[26] N. J. M. Levelt, *On Binocular Rivalry*, 1965, Institute for Perception RVO-TNO (Soesterberg, The Netherlands).

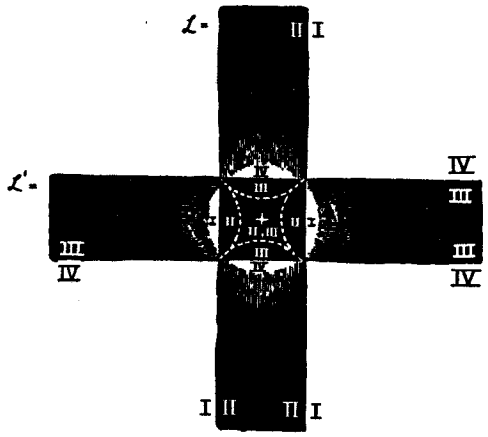

FIG. 13. Diagrammatic representation of the visual appearance of binocular rivalry. The stimulus to the left eye, *L*, is a vertical black bar on a white surround; the stimulus to the right eye, *L'*, is a horizontal black bar. The Roman numerals indicate the type of BCDNs stimulated in each area of the view. In the center and in the background, a BCDN member of each of the *A* and *B* pairs is stimulated; these areas are not seen in rivalry. In other areas, members of only one pair are stimulated; these are areas of rivalry. (After Helmholtz [n. 3], 496.)

The binocular rivalry discussed in the paragraphs above is within a depth plane. Rivalry *between* depth planes is another matter. Consider two objects, one behind the other, yet so placed that both objects are seen by each eye. The reader can arrange such a demonstration for himself with two pins. Stick the pins into a pencil about an inch apart, parallel to each other and perpendicular to the pencil. Hold the pencil against an illuminated surface at reading distance so that one pin is directly behind the other, as in Figure 14. As long as one pin is directly behind the other (pencil pointing between the eyes), both pins cannot be fused simultaneously; a double image is very obvious. When the pencil is rotated so that it points outside of the eyes, the angular distance between the depth planes of the pins is not changed much, but the pins can now be fused. This demonstration illustrates that even when both objects produce distinct retinal images in each eye (so there is no loss of information), an object cannot be seen as being behind another object at the same point of the visual field.

In the model, the inability to fuse two objects when one is directly behind the other results from the monoactivity of a column

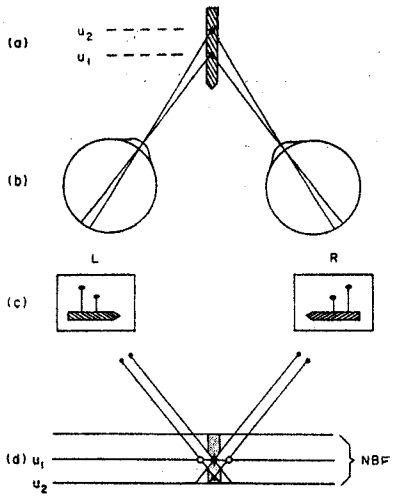

FIG. 14. A display in which two objects, one directly behind the other, nevertheless produce distinct images on each retina. (a) Top view of two pins in a pencil; u_1, u_2 indicate the depth planes. (b) Retinal projections of the stimuli. (c) Views of the stimulus as seen by each eye (these views are mirror images of the retinal projections). (d) Projection of the stimuli (c) onto the NBF: fusion of the near pin on level u_1 shown by an filled circle; the unfilled circles indicate heterogeneous fusion of a projection of the near pin with a projection of the far pin; shading indicates the area within which potential fusions are suppressed because of fusion at u_1; the filled circle at u_2 (in shaded area) indicates the point of binocular correspondence of images of the far pin, a point at which fusion is suppressed.

of BCDNs (i.e., outflow is restricted to only one level of the NBF). This rivalry between different levels of an NBF is fundamentally different from binocular rivalry between neurons at the same level of the NBF, where the within-pair interactions of BCDNs come into play.

Depth and outflow: Binocular combinations. A binocular view contains the outflow of the primary NBF (the contents) plus the combined fusion-depth values assigned to each x, y location by both NBFs. The binocular view is cyclopean because the two eyes combine to give a single outflow at each x, y location of the NBF; the contents of a binocular view therefore are basically the same as those of a monocular view. The important difference between monocular and binocular vision is the possibility of nonzero fusion-depth values in binocular vision.

To determine the contents and depth values of the cyclopean view in the most general case, consider a pair of stimuli (L and L') of the type illustrated in Figure 15. The letters A, B, X, Y, and A′, B′, X′, Y′ in Figure 15 denote the stimulus letters A, B, X, Y, or any other identifiable patterns. The stimuli X-X′ and Y-Y′ function as a reference frame to fix vergence; it is assumed that binocular correspondences of X-X′ and Y-Y′ occur in the middle layer of the NBF. When X = X′ and Y = Y′, then X-X′ and Y-Y′ are

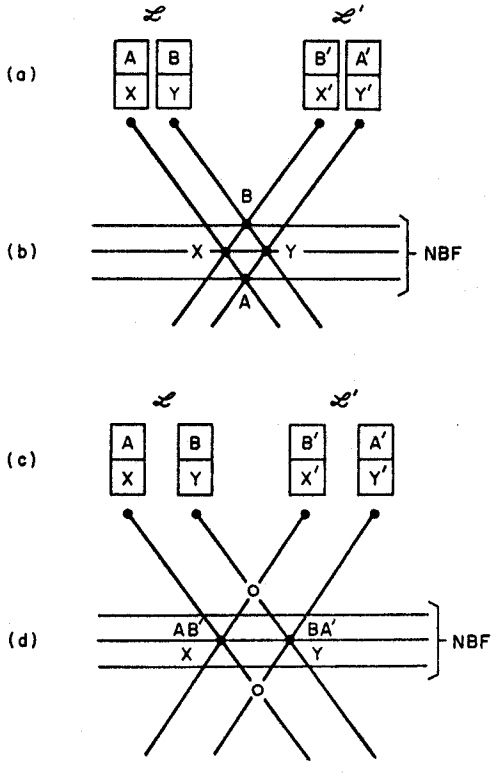

FIG. 15. Stimuli (a and c) to demonstrate binocular rivalry and/or depth; L and L' are the images to the left and right eyes respectively. The representations of the stimuli in the primary NBF are shown in (b) and in (d). (b) Monoactivity within each column of the NBF precludes seeing both A and B; only one is seen and in depth relative to X and Y. (d) A or B′ (A∩B′) is seen together with A′∩B, according to the outcome of rivalry in the central layer of the NBF. When binocular correspondence is detected in the secondary NBF (indicated by open circles), then A∩B′ and/or A′∩B is seen in depth relative to X and Y.

each seen as a single stimulus at a depth determined by vergence and incidental cues. Which of A or B is seen? What is its depth and lateral position relative to that of X and Y? The answers depend on the nature of A and B, and on the lateral separation between them. The answers to these questions are completely general, because for any given region of any binocular stimulus, there is a choice of A, B, X, Y and of A′, B′, X′, Y′ that makes the reduced stimulus configuration of Figure 15 prototypical of any given region of the complex stimulus.

—*Binocular combinations of identical stimuli.* In the first group of cases, B is assumed to be a null stimulus (i.e., equal to the background). This eliminates complications due to binocular rivalry and means that the specific questions and answers involve binocular combinations of identical stimuli (A and A′). When the lateral separation between X and Y is very small (less than $\frac{1}{3}$ deg. in central vision), the binocular correspondence between A and A′ falls within the primary NBF (Figure 15b). In this case, a single stimulus (A-A′) is seen, above and behind X and Y, laterally between them. Not only does A-A′ appear laterally between X and Y, but, by varying the relative intensity of the stimulus to the left and to the right eye, the perceived lateral position can be made vary continously between its two extremes: with A laterally directly above X, as when only the left eye's stimulus of Figure 15a is presented; or A′ directly above Y′, as when only the right eye's stimulus is presented. This displacement effect is predicted by the physical model of fusion as follows.

Let u_1, u_2, u_3, respectively be the values of u at the minima of $g(u)$, $h(u)$, and $e(u)$. Whenever $u_2 \neq u_1$, then $u_1 \leq u_3 < u_2$ or $u_1 \geq u_3 > u_2$; u_3 is displaced from u_2 toward u_1 by the amount $\triangle u = u_3 - u_2$. The shift $\triangle u$ depends on the shallowness of the minimum of $h(u)$; the shallower the minimum, the greater $\triangle u$. The deepest minima of $h(u)$ occur when l and l' are of equal intensity; the minima become shallower when l is unequal to l' in intensity, and they vanish entirely when either l or l' vanishes. Changing the value of $\triangle u$ corresponds to sliding the point of fusion along one of the l-l' projections in the NBF, and hence to a lateral shift as well as a shift in depth. In fact, this lateral mobility of a fused object is the best empirical criterion of fusion in the primary NBF.[27]

[27] The question of whether fusion can cause a lateral displacement of the fused object has received conflicting answers in the literature. Large displacements invariably are the result of uncontrolled vergence and suppres-

Consider now what happens when we separate X and Y. When the lateral separation between A and A′ exceeds the $\frac{1}{8}$ deg. half-depth of the primary NBF, binocular correspondence between A and A′ cannot occur within the primary NBF. In this case, the middle layer of the primary NBF dominates; A and A′ are represented there separately (Figure 15d). Unless one eye is strongly dominant, A is seen above X-X′ and A′ is seen above Y-Y′. However, if A and A′ are large relative to the lateral displacement between them, they overlap and binocular rivalry may cause one to partially or completely suppress the other. For this reason, singleness of vision is not a sufficient criterion for binocular correspondence in the primary NBF.

When the lateral separation of X and Y is less than the 2 deg. half-depth of the secondary NBF, the apparent depth of A is determined by the depth signal from the secondary NBF. Suppose both A and A′ are seen, one above each of X and Y. Whether or not A or A′ or both appear in depth behind X-Y depends on their lateral separation, on the past history of stimulation, and on their size and texture. With large separations, two stable states (depth, no depth) are possible in the secondary NBF, but the depth sign of the A-A′ fusion in the secondary NBF may not always be attached to both the A and A′ representations in the primary NBF. Therefore, the same stimuli (A-A′) may appear behind X-Y on some occasions and in the plane of X-Y on other occasions. In brief, then, one or both images of A are seen because of their separate representations and possible rivalry in the primary NBF; neither, one, or both of the images appear in depth depending on whether binocular correspondence was detected in the secondary NBF. (These relations between the objects of Figure 15 are summarized in Table II, in the middle cell of the top row).

The last cell of the top row of Table II indicates the visual contents when the separation between X and Y exceeds the limit at which correspondences can be calculated in the secondary NBF. Then, in both NBFs, the only active layer is the middle one and no depth difference is seen between A and X-Y. Because A is represented twice in the middle layer, it is seen doubly. In all rows of

---

sion. This is demonstrated by briefly intensifying a steady l and l' alternately, and noting the reappearance of the suppressed elements. For background, see H. Werner, Dynamics in binocular depth perception, *Psychol. Monogr.*, 49, 1937, No. 2 (Whole No. 218); L. Kaufman, Suppression and fusion in viewing complex stereograms, this JOURNAL, 77, 1964, 193–205.

Table II depth disappears when the disparity exceeds the range of the NBFs, and in all cases it reappears (if it appeared in any cell of the row) when vergence movements are permitted.

Because the receptive fields of cells composing the secondary NBF are assumed to be very large, partial correspondences may be detected at disparities far exceeding the range of the secondary NBF. For example, suppose the extreme layers of the secondary NBF compute correspondences at disparities of 2 deg. If the diameter of receptive fields also were 2 deg., then stimuli separated by 4 deg. would still stimulate the edges of the receptive fields at the extreme layers, and a partial correspondence would be detected there. Brief flashes of simple stimuli separated by over 4 deg. do, in fact, yield appropriate localizations in depth (i.e., far or near) and they induce vergence movements in the appropriate directions,[28] but these results do not directly indicate the range of the secondary NBF.

—*Binocular combinations of rivalrous stimuli.* In this second group of cases, B is assumed to be a nonzero stimulus, different from A. This means that the question of what is seen involves binocular combinations of rivalrous stimuli (A and B—assuming that they are sufficiently different to 'rival' each other, a condition which it is obviously impossible to fulfill for very small X-Y separations). The results with rivalrous stimuli are basically the same as for stimuli that contain only A, except that wherever A was seen, only one of A or B is seen, depending on the local outcome of the A-B rivalry. These results are summarized in the second row of Table II.

When the difference between A and B is too small to be resolved by the secondary NBF (e.g., two different printed letters are interpreted as similar 'blobs'), then there are three possible levels of binocular correspondence in the secondary NBF: A-A′; A-B′, B-A′; and B-B′. Of these, the A-B′, B-A′ correspondence is most likely to dominate because it is on the middle level of the NBF. When the middle level dominates, relative depth between X-Y and A or B is lost. The last two cells of the second row of Table II summarize these results.

When A and B are so similar that they are taken to be identical

[28] D. E. Mitchell, Qualitative depth localization with diplopic images of dissimilar shape, *Vis. Res.*, 9, 1969, 991–994; G. Westheimer and D. E. Mitchell, The sensory stimulus for disjunctive eye movements, *Vis. Res.*, 9, 1969, 749–755; D. E. Mitchell, Properties of stimuli eliciting vergence eye movements and stereopsis, *Vis. Res.*, 10, 1970, 145–162.

TABLE II

CONTENTS OF THE CYCLOPEAN BINOCULAR VIEW FOR THE STIMULI SHOWN IN FIGURE 15

Stimulus	Disparity range (central vision)		
	Primary NBF ($\pm$ 1/8 deg.)	Secondary NBF ($\pm$ 2 deg.)	Beyond NBFs
A = A′, B = B′ = 0	(*d*, AA′, X/Y)	(*nd* $\cap$ *d*, A, X) + (*nd* $\cap$ *d*, A′, Y)	(*nd*, A, X) + (*nd*, A′, Y)
(A = A′) $\neq$ (B = B′) $\neq$ 0	(*d*, AA′ $\cap$ BB′, X/Y)	[(*nd*, A $\cap$ B′, X) + (*nd*, A′ $\cap$ B, Y)] $\cap$ [(*d*, A, X) + (*d*, A′, Y)] $\cap$ [(*d*, B′, X) + (*d*, B, Y)]	(*nd*, A $\cap$ B′, X) + (*nd*, A′ $\cap$ B, Y)
(A = A′) $\approx$ (B = B′) $\neq$ 0	(*nd*, AB′, X) + (*nd*, A′ B, Y)‡	(*nd*, A $\cap$ B′, X) + (*nd*, A′ $\cap$ B, Y)†	(*nd*, A $\cap$ B′, X) + (*nd*, A′ $\cap$ B, Y)†
A = B = B′, A′ = 0 (Panum)	(*nd*, AB′, X) + (*d*, BB′, X/Y)	(*nd*, AB′, X) + (*nd* $\cap$ *d*, B, Y)	(*nd*, AB′, X) + (*nd*, B, Y)

Note: Each row represents a different angular separation (disparity) between X and Y. Each column represents a different stimulus content, A or B. The subjective visual contents are indicated by terms (extending from a left to a right parenthesis). Each term is composed of three elements. The first element represents depth: *d* indicates that the content (A or B) is seen in a different depth plane than X-Y: *nd* indicates no depth (i.e., the same plane). The second element represents content: $\cap$ indicates that either A or B is seen but not both; AB indicates that a concatenation of A and B is seen. The third element indicates location (lateral position) of the content: X/Y indicates that it is seen between X and Y; X indicates that it is seen directly above X; Y indicates that it is seen above Y; X $\cap$ Y indicates that it is seen above X or above Y but not both. To find the contents of a particular cyclopean view, sum the applicable terms. The table is not exhaustive, but all indicated combinations of elements are possible, though not necessarily for the same choice of A and B.

† Difference between A and B resolved by primary NBF, unresolved by secondary NBF.

‡ Difference between A and B unresolved by primary or secondary NBF.

not only by the secondary NBF but also by the primary NBF, then A fuses with B′ and B with A′ on the middle level of the primary NBF and no relative depth is seen. The first cell of the third row in Table II represents this situation. A-B′ fusion without depth is not the only stable state; but it takes some ingenuity to achieve the others. For example, if the intensity of B-B′ is initially set to zero, then fusion occurs on A-A′. By slowly intensifying B-B′, a second state (A-A′ fusion with depth) can be maintained, at least temporarily.

Because of its historical interest, Panum's limiting case[29] is described in the last row of Table II. In Panum's limiting case, the stimuli, A, B, and B′, are identical lines (e.g., the letter I), and A′ is null. The AB′ line is always fused in the X-Y plane. Whether the B-B′ binocular correspondence occurs in the primary or secondary NBF depends on the X-Y disparity, and it determines the lateral shift (if any) of B; in either case, it is seen behind X-Y.

The previously untested predictions of Table II have been verified experimentally by the author, using stimuli of the kind introduced by Kaufman following Julesz's method.[30] The demonstrations and the predictions have much in common with those of Verhoeff, Asher, and Kaufman.[31] It was Kaufman who first obtained and correctly interpreted evidence for a coarse-detail function (a 'blob' detector) in binocular correspondence detection.[32] Some of Kaufman's results are considered below; the neural model accounts for all the details.

Binocular combinations: The Kaufman figures. Figure 16ab illustrates two stereoscopic image pairs (L_1, L_1' and L_2, L_2') taken from Kaufman (n. 22). Only Figure 16a yields a stable stereoscopic depth illusion. Figure 16cd illustrates the representation of these stimuli in the secondary NBF. (Because of the large separation between characters, binocular correspondences at nonzero levels occur only in the secondary NBF.) Partial correspondences (any character of L with any other character of L') are indicated by open circles; exact cor-

[29] K. N. Ogle, Special topics in binocular spatial localization, in H. Davson (ed.), *The Eye*, 4, 1962, 349–407.

[30] Kaufman (n. 27); Julesz (n. 21).

[31] F. H. Verhoeff, A new theory of binocular vision, *Arch. Ophthal.*, 13, 1935, 151–175; H. Asher, Suppression theory of binocular vision, *Brit. J. Ophthal.*, 37, 1953, 37–49; Kaufman (n. 22).

[32] L. Kaufman, On the nature of binocular disparity, this JOURNAL, 77, 1964, 393–402; Kaufman (n. 22).

respondences (a character with its exact match) are indicated by filled circles. With the stimuli of Figure 16a, exact binocular correspondences occur only on one level of the NBF. When a BCDN becomes active on that level, activity quickly spreads to the limits of the region of exact correspondence, and beyond, suppressing the other levels.

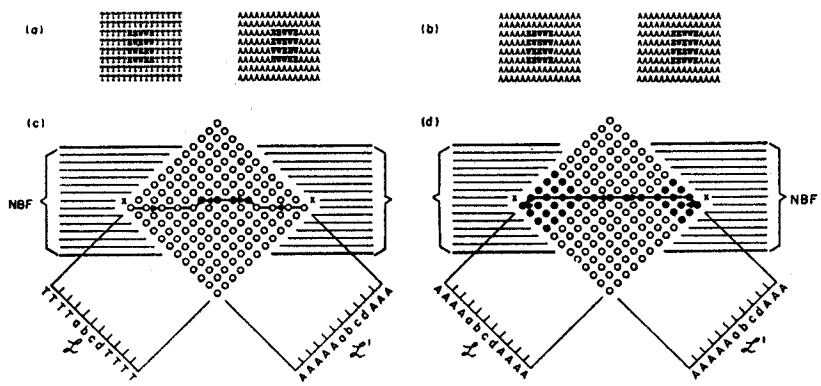

FIG. 16. Two stereogram pairs (a and b) from Kaufman (n. 22) and their representations (c and d) in the secondary NBF. Filled circles indicate exact correspondences, unfilled circles indicate coarse correspondences (i.e., of different letters); Xs indicate projection onto middle layer of outermost letters (for these, there is no binocular correspondence in the middle layer). Arrows indicate the direction of spread of activity within a layer; the lines through the arrows indicate the final, stable depth perception. For simplicity, only 12 of the 15 columns of the stereograms are indicated in (c) and (d).

Partial correspondence is possible at many levels throughout the secondary NBF. However, one BCDN that is always activated by a partial correspondence is the outermost one, because only this BCDN suffers no competition from correspondences at other levels. Since the uncontested, outermost partial correspondence is one level removed from the level of exact correspondence, it initiates a counterspread of activity throughout its level. The activity initiated by the partial correspondence at the edge spreads inward from the edge while the activity initiated by exact correspondences spreads outward from the center; ultimately they meet. The exact correspondence is a stronger input to the BCDNs than the partial one, so that wherever a conflict arises, the exact correspondence predominates. Therefore, in Figure 16c, the activity initiated by outer correspondences is excluded from the center rectangle. There are two

stable states: the one illustrated in Figure 16c (which predominates) and the initial state (illustrated in Figure 16d).

For the stimuli of Figure 16b, activity initiated by an exact correspondence in the center of the stimulus spreads almost to the very edge (Figure 16d). Thus in Figure 16a, the center rectangle is seen in depth against its entire surround; in Figure 16b, only the end columns are seen in depth. By careful stereoscopic observation of Figure 16b the reader may be able to observe the temporal spread of activity from the center outwards: he also may observe briefly various semistable depth configurations.

The foregoing explanation accounts for the presence or absence of relative depth in Kaufman's demonstration. The content of what is seen (i.e., which letter) is determined in the primary NBF. Only the center level is active there because the disparity between adjacent letters exceeds its limits. Since the eyes verge on the inner letters, these fall on the center level. To note which letters are projected onto the center level of the primary NBF, the reader may refer to the center level of the secondary NBF (Figure 16c, 16d); there is no need for a separate diagram. For the stereogram of Figure 16a, which letter is seen depends on the A-T rivalry ($A \cap T$) at the center level of the primary NBF. The end columns are an exception because only the stimulus from one eye is represented at the center level of the NBF (crosses, Figure 16c). Thus the left-end column is A, and the right-end column is T.

In the stereogram of Figure 16b, the character A is seen at all its projections onto the center level. As for lateral position, all letters line up in vertical rows; the letters of the end columns (which can be seen in depth) do not appear to be shifted horizontally. (Lateral position is determined in the primary NBF, i.e., by the projections onto the center level in these illustrations.) These phenomena are subsumed in Table II in the second column (secondary NBF), the second ($A \neq B$) and third ($A \approx B$) rows. That is, A is partially confused with B in the secondary NBF and therefore the phenomena characteristic of both $A \neq B$ and $A \approx B$ occur. Most of Kaufman's other interesting demonstrations fall in the same cells of Table II.

Binocular combinations: The Hochberg stereograms. Hochberg published several interesting stereograms,[33] all of which use rival-

[33] J. Hochberg, Depth perception loss with local monocular suppression: A problem in the explanation of stereopsis, *Science*, 145, 1964, 1334–1336;

rous stimuli with this property: that whenever binocular rivalry causes the image in one eye to dominate completely, the illusion of depth is lost. Since Kaufman's stereograms are counterexamples to this rule (only A or T is seen, yet depth persists; Figure 16a), we are faced with specifying the critical difference between the two kinds of stereograms. The main difference seems to be that the rivalry-inducing features of Hochberg's demonstrations are quite coarse, so that rivalry occurs in the secondary NBF as well as in the primary NBF.

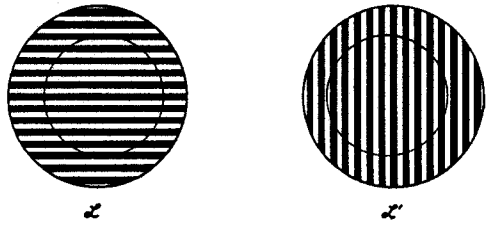

FIG. 17. Stereogram from Hochberg (third reference, n. 33) after Kaufman and Pitblado. The inner circles are seen in depth except when binocular rivalry of one pair member over the other is complete.

For example, to produce the stereogram of Figure 17, a horizontal grid is added to the left member and a vertical grid is added to the right member of a stimulus pair that would normally produce a stable illusion of depth. The relative coarseness of the grid feature explains the occasional loss of depth, because large conflicting features cause failure to compute the image-disparity minima in the secondary NBF.

Feature coarseness does not explain the correlation between suppression of one eye's input in the primary NBF and depth loss (suppression of an input in the secondary NBF). Significantly, however, the best operation for producing the correlated effects of suppression and depth loss is reducing the stimulus intensity to one eye.[34] Filtering the light input obviously affects both NBFs, which suggests that peripheral causes—such as vergence changes or other

J. Hochberg, On the conditions of stereopsis-suppression, *Science*, 146, 1964, 800; J. Hochberg, One view through two eyes: A theory of binocular combination and some supportive experiments, paper read at the meeting of the Psychonometric Society, October, 1964.

[34] Kaufman and Pitblado, cited by Hochberg (see the second reference in n. 33).

eye movements—may account for the correlation and, at any rate, need to be eliminated before considering explanations at higher levels.

Binocular combinations in stabilized vision. Fender and Julesz's experiments on the "extension of Panum's fusional area in binocularly stabilized vision" were described above, in the section on examples of the physical model for fusion. In these experiments, relative depth perception persisted in stimuli that had been displaced by 2 deg. on the retina (i.e., the left eye's stimulus moved left by 1 deg. and the right eye's stimulus moved right by 1 deg.). These conditions fall in the middle cell of Table II (secondary NBF, $A \neq B$). The only discrepancy between the prediction of Table II and the observation is that Table II predicts that two images, one of l and one of l', will be seen, whereas visibility of only one was reported. To account for this result, we must carefully examine the procedure.

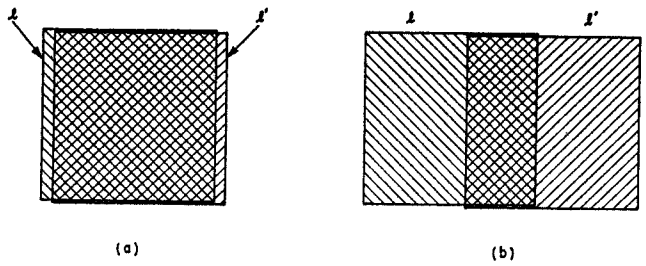

FIG. 18. Cyclopean view of the relative position of the stimulus to the left (l) and right (l') retina in the stabilized-vision experiment of Fender and Julesz (n. 4). The width of the stimuli is 3.4 deg. The orientation (b) illustrates the point at which singleness of vision is lost; (a) illustrates the point at which fusion reoccurs.

Figure 18a illustrates the stimulus conditions at the beginning of a trial, and Figure 18b shows them at the point where singleness of vision was lost. If the stimulus of Figure 18b were presented by itself, both l and l' would be seen, yet when the stimulus of Figure 18a was transformed slowly into that of Figure 18b, subjects saw only one square, an effect which Fender and Julesz explained in terms of an extended fusion area in stabilized vision. Both results can be explained without introducing any new concepts: the crucial detail is Fender and Julesz's inadvertently cunning use of binocular

rivalry. For the stimulus of Figure 18a, binocular rivalry determines which of l or l' is seen. Once the lateral disparity between l and l' is about $\frac{1}{4}$ deg. (top-to-bottom depth of the primary NBF), binocular correspondence no longer can be computed within the primary NBF. The active level of the primary NBF is the center level; wherever l and l' conflict on that level, rivalry determines which is suppressed. Because of the dense detail of l and l', it is unlikely that any contour from l can fall against a neutral area in l', or conversely. Contours fall against contours, so that when l' is suppressed, for example, there are no stable islands where contours of l' can break the suppression by l. However, depth relations between portions of l and l' persist because they are computed in the secondary NBF.

As the lateral disparity between l and l' is increased, the suppressed stimulus remains suppressed even after it no longer is competing against contours. This phenomenon can be observed in normal vision, although the model predicts it would be more readily produced in stabilized vision where eye movements and blinks—events that would tend to perturb a delicately achieved suppression state—are eliminated. The 2-deg. limit of single vision (not fusion) observed by Fender and Julesz measures the limit at which the binocular correspondence no longer can be maintained in the secondary NBF (see Figure 5). The change of state in the secondary NBF from binocular correspondence (at $\triangle u = 2$ deg.) to rivalry (at $\triangle u = 0$) triggers the failure of the extraordinary suppression. At this point, the dominant stimulus is seen in its entirety; the other reappears in the regions where it stands alone (Figure 18b). Both stimuli are seen at loci appropriate to their retinal locations.

In brief, the depth that results with binocularly stabilized images is accounted for by depth signals from the secondary NBF, and the singleness of vision (apparent fusion) is accounted for by binocular rivalry. No new principles are involved.[35] In fact, similar sup-

[35] Professor Fender (personal communication) himself has often observed the sudden loss of apparent fusion at the critical separation of binocularly stabilized images. He states that the sensation is one of sudden 'jumping apart' of two images from a single source. The proposed suppression hypothesis, he thinks, would predict that at the instant when suppression fails, the suppressed image merely reappears without any 'jumping apart.' The observation of 'jumping apart' and the assumption that suppression precludes 'jumping apart' causes him to reject the suppression hypothesis. To support the hypothesis it is proposed that the loss of binocular correspondence in the secondary NBF, a loss which is coincident with reemergence of the suppressed stimulus, produces a sensation of 'jumping apart.'

pression phenomena can be observed in normal, unstabilized vision by displacing stimuli laterally beyond the limits at which the eyes diverge to verge on them. In unstabilized vision, binocular rivalry is quite apparent, since only partial suppression is normally achieved.

Other functions in the NBF. So far, only two functions of the input have been considered: coarse detail and fine detail. Both of these functions can be accomplished by center-surround types of receptive fields of the appropriate size. Undoubtedly, many other functions of the input are computed. For example, the well-known red-green antagonism in binocular vision suggests the existence of a function pair for green (F) and red ($\sim F$). So far too, it has been assumed that the fine-detail and coarse-detail functions of the primary and secondary NBFs are noninteracting and that the fine-detail system of the primary NBF controls other inputs, such as color and brightness. At present, however, the relations to each other of different function systems—even the specification of such systems—must be regarded as an open question for future theory and experiments.

Two processes that have erroneously been attributed to the NBFs are a change in the mapping of corresponding points with asymmetric vergence (to account for stability of the apparent normal plane) and a change in the receptive field size with convergence (to account for changes in apparent size).[36] In fact, corresponding points, as measured by nonius lines are so remarkably stable[37] that remapping is excluded, and we must look elsewhere for explanations of these perceptual phenomena. The apparent location, orientation, and size of objects depend in a complex way on the interpretation by subsequent analyzers of information available from the NBFs *and other sources* rather than on any restructuring of the NBFs themselves.

Evolution of Binocular Depth Discrimination

Some of the interactions between accommodation, vergence, and fusion are best understood in terms of their evolutionary development. Of the three, according to Walls, accommodation was

[36] W. Richards, Spatial remapping in the primate visual system, *Kybernetik*, 4, 1968, 145–156.
[37] Ogle (n. 8).

evolved first and most universally.[38] The most versatile of the accommodation mechanisms is that for comparing blur in different retinal areas. It requires the object whose depth is in question to overlap at least two predetermined retinal areas merely to make a determination of closer or further than the current value of accommodation; even then, the determination of blur is highly dependent on the number of contours within each area. Furthermore, when pupil diameter is reduced in bright daylight, blur becomes an extremely insensitive indicator of depth because of increased *depth of focus*. Therefore, an accurate accommodative depth system would ultimately depend on movements of the lens and of the eye to track depth, and would use the accommodation values as the depth estimate. Unfortunately, systems that require both the eye and the lens to track are too inefficient, too insensitive (relative to the systems described below), and too slow—and therefore seem not to have evolved. Accommodation thus failed to develop into a depth-discrimination mechanism, and depth determination fell to other systems.

Head movements provide extremely good cues for the monocular discrimination of depth. The two successive images to the one eye correspond exactly to the two different views of different eyes. The evolution of a depth mechanism based on head movements would suffice for animals that are not endangered by having to move their heads and for animals that capture food with their heads by moving toward it until they encounter it. But animals that use paws, hands, or trunks to capture food need *binocular* depth perception to avoid avoidable trial-and-error reaching. (In fact, such animals have the most binocular overlap in their monocular fields—an indication of dependence on binocular vision.)

The evolution of stereoscopic depth discrimination required, first, overlapping visual fields (which originally evolved to avoid blind spots); second, a binocular registration mechanism (the NBF, which originally evolved to deal with the binocular overlap); and ultimately, vergence movements. To show how these systems evolved together, consider the following example. Suppose a species has evolved an NBF that is ten levels thick, and that an animal can discriminate two depths if their levels in the NBF differ by one level. If the total NBF thickness of ten levels corresponded to 15 deg., the animal could discriminate two depths that differed

[38] Walls (n. 13), 247 ff.

by 1.5 deg.; if the total depth range of the NBF were 5 deg., the animal could discriminate .5 deg. As the NBF evolved toward *the limits of visual acuity and became capable of ever-finer dis*crimination, it also covered an ever-smaller depth range. The principle is the same as in a microscope: the higher the power, the smaller the range of view. A small NBF depth range implies a critical dependence on accurate centering in the NBF. Therefore, to evolve a sensitive NBF, precise vergence movements were necessary.

A small NBF range also means that vergence is not computed for objects outside this range. At this point, the link of accommodation to vergence becomes an important supplement to the NBF. Thus the development of sensitive binocular depth discrimination required the concomitant development of accurate vergence (to center objects in the NBF) and of linked accommodation (to supplement the range of the NBF).

Conclusion

The neural model represents only a tiny subject of the functions carried out in the visual system and does so in terms of neural components compatible with our present-day knowledge of neural processes. It is not unique: similar functions could be carried out in other ways. Nor does the gometric arrangement of parts of the model—multilayered NBFs—necessarily correspond to neuroanatomy; the model is an hypothesis only about topology (i.e., connectivity).

For all its limitations, the model has one characteristic that is of very general interest. Namely, one function performed by the BCDNs—limiting activity to a single level of the NBF—is representative of a universal kind of function in the nervous system. The BCDNs perform an extreme filtering action that restricts the whole range of potential NBF outputs to output from only one level; it makes only some very selected aspect of the stimulus available for further processing. Since it is inefficient for a system to decide in advance just which selected aspects of every possible stimulus are to be processed, the selection of input for processing must be under dynamic control; the results of previous processing—the previous outputs—must control the future inputs. Therefore, extreme filtering invariably is accompanied by feedback control

of the filter; these twin processes produce the phenomena of multiple stable states.

For example, the lens of the eye can be focused exactly on but a single depth plane, an input restriction that places greater weight on the depth plane accommodated and, together with the feedback control of accommodation, leads to the multiple stable states of accommodation. Similarly, the vergence angle between the eyes determines the input from which binocular image disparity is computed, and in turn image disparity controls vergence. Again, filtering and feedback control lead to multiple stable states. Of course, vergence movements are just a minor subclass of eye movements. In general, the direction in which the eyes are pointed determines the input they receive, and this input ultimately determines the direction in which they point. Eye movements represent the ultimate in peripheral filtering under feedback control—a truly multistable system.

The BCDN interactions in the NBF are especially interesting because the filtering is not mechanical but neural. When a binocular correspondence is detected at one level of the NBF, the level becomes active and output from the NBF is restricted to that one level (is filtered). The choice of level depends on responses to previous inputs (is feedback controlled), and the result is multiple stable states of fusion. Other examples of visual processes that exhibit extreme multistability, presumably because of feedback-controlled neural filtering, are binocular rivalry (seeing the stimulus from one eye in a region of the visual field precludes seeing a conflicting stimulus from the other eye in that region) and figure-ground phenomena (seeing an ambiguous figure in one mode precludes simultaneously seeing other modes; seeing a particular figure-ground relation, as on a map, precludes simultaneously seeing other figure-ground relations). Even elementary interactions between successive visual stimuli can be interpreted as exhibiting this multistable 'either-or' characteristic.[39]

In nature, there normally is no reason for organisms simultaneously to execute two antagonistic responses, and organisms cannot do it. Therefore, there is no reason to preserve information that ultimately would lead to antagonistic responses. The neural binocular field is an early locus for the rejection of antagonistic information; many of its paradoxical properties follow therefrom.

[39] G. Sperling, Bistable aspects of monocular vision, *J. opt. Soc. Amer.*, 50, 1960, 1140–1141.

APPENDIX A

PROOF THAT BLURRING INCREASES IMAGE-DEFOCUS ENERGY (h_e): THE RELATION OF h_e TO RECEPTIVE-FIELD STRUCTURE

Let $l(x, y)$ be any luminance distribution such that $l(x, y) \geq 0$. Let $b(x, y)$ be any blur function such that $b(x, y) \geq 0$ and $B = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b(x, y)\, dx\, dy = 1$. When $l(x, y)$ is blurred by the function $b(x, y)$ the resulting blurred image $[l_b(x, y)]$ is given by

$$l_b(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} l(x', y') b(x - x', y - y')\, dx'\, dy'. \qquad \text{[A-1]}$$

The two-dimensional Fourier transform of $l(x, y)$ is

$$L(\omega, \sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\omega x + i\sigma y} l(x, y)\, dx\, dy.$$

The Fourier transform $\mathfrak{B}(\omega, \sigma)$ of $b(x, y)$ is defined similarly. Let ∇^i represent the generalized Laplacian $[(\partial^i/\partial^i x) + (\partial^i/\partial y^i)]$. Then Parseval's theorem gives the following equality, which also defines $G(l, i)$:

$$G(l, i) = \iint \left(\nabla^i l(x, y)\right)^2 dx\, dy = \iint (\omega^i + \sigma^i)^2 |\, L(\omega, \sigma)|^2\, d\omega\, d\sigma\,. \qquad \text{[A-2]}$$

$G(l_b, i)$ gives a similar relation in terms of $|L_b(\omega, \sigma)|$, where $L_b(\omega, \sigma) = L(\omega, \sigma)\mathfrak{B}(\omega, \sigma)$. Proof of the theorem that blurring increases h_e follows immediately from Equation A-2 if we can show that $|L_b(\omega, \sigma)|^2 \leq |L(\omega, \sigma)|^2$; to show this it is necessary only to show that $|\mathfrak{B}(\omega, \sigma)|^2 \leq 1$, since

$$|L_b|^2 = |L|^2\, |\mathfrak{B}|^2. \qquad \text{[A-3]}$$

Using E to denote expectation, and recognizing that b is a probability distribution function, gives

$$|\mathfrak{B}(\omega, \sigma)| \equiv |E(e^{i\omega x + i\sigma y})| \leq E\, |e^{i\omega x + i\sigma y}| = E\, |1| = 1. \qquad \text{[A-4]}$$

In fact, the middle equality occurs only when $b(x, y) = \delta(x - a_1)\, \delta(y - a_2)$; that is, when there is no blurring, merely exact reproduction or translation.

Equations A-3 and A-4 in conjunction with Equation A-2 give

$$G(l_b, i) \leq G(l, i). \qquad \text{[A-5]}$$

In the text, h_a was defined as $-G(l, 2)$. As Equation A-5 unambiguously states, blurring l to produce l_b can only decrease G or, equivalently, increase h_a. (The proof of Equation A-5 is due to M. Sondhi.)

Receptive fields. The definition of h_a in terms of ∇^2 is perfectly logical in terms of receptive fields. The antagonistic center-surround type of receptive field corresponds to a 'blurred' ∇^2.[a] That is, if the center portion of the receptive field were infinitely narrow, and the surround were an infinitely small surrounding annulus which subtracted from the center, the receptive field would correspond exactly to ∇^2. In fact, the center and surround portions of the receptive field have a finite extent; this extent corresponds to a particular blurring (b) of the output of a network of receptive fields. The transformation performed by the receptive field is thus approximated by $b\nabla^2$. Because differentiation (∇^i) commutes with convolution (blurring), the finite center-surround dimensions of the receptive fields can be treated as blur in the stimulus, and the fields themselves considered to be perfect differentiators ($\nabla^2 b$). The defocus of the lens adds additional blur to the image; it is this variable source of blur, h_a, which is the stimulus for accommodation.

The use of the square $\nabla^i l(x, y)^2$ in Equation 8 of the text and Equation A-2 is for mathematical elegance. The replacement of $(\nabla l)^2$ by $|\nabla l|$ might perhaps bring the expression closer to physiological plausibility (separate 'on' and 'off' channels convey the positive and negative values of ∇l), but it would not materially alter the main results (Equation A-5). That is, when the absolute value replaces the square, Equation A-5 becomes

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\nabla^i l_b(x, y)|\, dx\, dy \leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\nabla^i l(x, y)|\, dx\, dy \qquad \text{[A-6]}$$

where l_b is a blurred image of l.

In Equation A-6, strict equality holds for $i = 0$; equality may still hold even when $i \geq 1$ for certain combinations of stimuli (l) and blur functions (b). The condition for inequality is that blurring cause some light from an area where $\nabla^i l > 0$ to spill into an area where $\nabla^i l < 0$ (or vice versa). For $i \geq 2$, this spillage always occurs in physically realizable images with nontrivial blur functions,

[a] For an early insight into this problem, see E. Mach, Ueber die physiologische Wirkung räumlich vertheilter Lichtreize, IV, *Sitzungsber. math.-naturwis. Classe Kaiserl. Akad. Wissensch.*, 57 (1868), No. 2, 11–19.

and so the absolue value, as well as the square, could serve in the evaluation of blur. Finally, as a practical matter, for special $l(x, y)$ which are themselves very 'blurred,' the additional effect of defocus may be negligible; in these cases, the theory predicts, it will be impossible to accommodate the images.

APPENDIX B

ANALYSIS OF NEURON-PAIR INTERACTIONS

In terms of its function, a neuron is defined by its input-output relations; these are approximated by a nonlinear differential equation of the form

$$dy(t)/dt + (A + B(x) + C(y))y(t) = x(t). \qquad [B\text{-}1]$$

Here, $x(t)$ is the input, $y(t)$ is the output; A is a constant, B is a function of the entire input (feedforward inhibition), and C is a function of the entire output of all the neurons at the same level (feedback inhibition).[b]

Elementary neuron pair. Consider the pair of neurons illustrated in Figure B-1a. These neurons interact only by mutually inhibiting each other; the constants (k) of mutual inhibition are assumed to

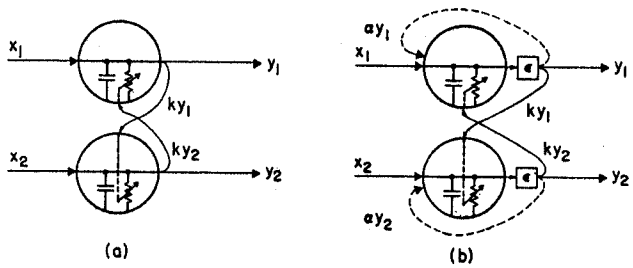

FIG. B-1. Reciprocally interconnected neurons. (a) Elementary neuron pair: each neuron consists of an RC stage; the value of R is controlled by the output of the neighboring neuron. (b) Neurons with positive feedback: threshold, ϵ, for output has been added, and there is an excitatory feedback loop, αy_1, from the output back into the input.

[b] To relate Equation B-1 to cellular parameters, consult W. Rall, Theoretical significance of dendritic trees for neuronal input-output relations, in R. F. Reiss (ed.), *Neural Theory and Modeling*, 1964; G. Sperling and M. M. Sondhi, Model for visual luminance discrimination and flicker detection, *J. opt. Soc. Amer.*, 58, 1968, 1133–1145.

be equal; and there is assumed to be no threshold and no self-inhibition. Thus the term $B(x) = 0$; $C(y) = ky$; the dimensional constant (A) is taken to be 1.0. The analysis becomes quite simple if it is restricted to steady-state outputs $(dy/dt = 0)$ in response to step inputs $[x(t) = \text{constant}, t > 0]$. In fact, very little generality is lost thereby. For such a neuron pair, Equation B-1 reduces to

$$y_1 = \frac{x_1}{1 + ky_2} \qquad \text{[B-2]}$$

and

$$y_2 = \frac{x_2}{1 + ky_1}. \qquad \text{[B-3]}$$

The inputs (x_i) are restricted to positive values and labeled so that $x_2 \geq x_1 \geq 0$. Define Δy as $y_2 - y_1$ and Δx as $x_2 - x_1$. Multiplying out Equations B-2 and B-3, and subtracting, gives

$$\Delta y = \Delta x. \qquad \text{[B-4]}$$

This result holds for all values of k.

Adding Equations B-2 and B-3, and reducing the result algebraically, gives

$$(y_1 + y_2) = -\frac{1}{k} + \sqrt{(\Delta x)^2 + 2(x_1 + x_2)/k + 1/k^2}. \qquad \text{[B-5]}$$

For special conditions (when $k \gg 0$), Equation B-5 can be substantially simplified:

$$k\Delta x \gg 2(x_1 + x_2) \Rightarrow \begin{cases} y_1 \approx 0 \\ y_2 \approx \Delta x \end{cases} \qquad \text{[B-6]}$$

and

$$k\Delta x \ll 2(x_1 + x_2) \Rightarrow \{y_1 \approx y_2 \approx \sqrt{x/k}. \qquad \text{[B-7]}$$

Equation B-6 holds for $x_1 \neq x_2$ and Equation B-7 holds for $x_1 \approx x_2$. These results show that the elementary neuron pair is a system for performing subtraction. The system is all the more remarkable because the interaction is not subtractive but divisive (the inhibition terms appear in the denominators of Equations B-2 and B-3). The output Δy always equals the input Δx, and when k is large, the smaller output is approximately zero. For large k, at most one

of two neurons is active, the one with the larger input, and this neuron's output is approximately Δx.

Elementary neuron N-tuple. When more than two neurons interact mutually in the same way as the neuron pair (i.e., as a neuron N-tuple), Equations B-2 and B-3 generalize to

$$y_i = \frac{x_i}{1 + k \sum_{j \neq i} y_j}. \qquad \text{[B-8]}$$

For large k and N the following properties are derived from Equation B-8. Let x_1 be the largest input and let all other inputs (x_i) be equal. Two cases are considered: first, when $x_1 > \sum_2^n x_i$ (see Equation B-9) and second, when $x_1 \leq \sum_2^n x_i$ (see Equation B-10).

$$x_1 = Nx_2 + \gamma \text{ and } \gamma > 0 \Rightarrow \begin{cases} y_1 \approx \gamma \\ y_j \approx \frac{x_2}{k\gamma} \approx 0 \end{cases} \qquad \text{[B-9]}$$

$$x_1 = Nx_2 \Rightarrow \begin{cases} y_1 = x_1^{2/3} k^{-1/3} \to 0 \\ y_j = x^{1/3} k^{-2/3} N^{-1} \to 0 \end{cases} \qquad \text{[B-10]}$$

When the largest input (x_1) is less than the sum of other inputs, the output also yields a power of k ($k^{1/2}$) in the denominator, much as in Equation B-10. Thus, the neuron N-tuple retains the desirable features of the elementary pair only when one neuron has a larger input than all the other inputs combined. Otherwise, all outputs depend on k and go to zero as $k \to \infty$. (The author is indebted to M. Sondhi for the derivation of Equations B-9 and B-10.)

Neurons with positive feedback. Consider an elementary neuron pair, like that of Figure B-1b. Each neuron feeds a part of its output, αy, back into its input, where it acts like an added input signal. (The threshold is ignored, temporarily.) The equation describing this modified pair is

$$y_i = \frac{x_i + \alpha y_i}{1 + k y_j}. \qquad \text{[B-11]}$$

By defining β as $1/(1 - \alpha)$ for $\alpha < 1$, it is a simple matter of algebraic manipulation to reduce Equation B-11 to

$$y_i = \frac{(\beta x_i)}{1 + (\beta k) y_j} = \frac{x'}{1 + k' y_j}. \qquad \text{[B-12]}$$

The right-hand term of Equation B-12 shows that by redefining input as $x' = \beta x$ and the inhibitory coefficient as $k' = \beta k$, the basic form of Equation B-2 remains unchanged by positive feedback. The net effect of positive feedback is to multiply k and $\triangle x$ by β. This is no small effect: as α approaches 1, β approaches ∞; for $\alpha \geq 1$, β is infinite. Thus, a relatively small amount of positive feedback can enormously increase the strength of interaction (k') between neurons, and their gain (β).

Effect of threshold (ϵ). Neuronal thresholds arise naturally from the two-stage conversion of input to output: a neuron's input is converted, first, into an internal voltage and, second, into an output spike rate. Spike rates typically are proportional to internal voltage, but as rates cannot fall below zero, there is a threshold voltage. If the internal resting voltage of a neuron (zero input) is ϵ below its threshold-for-firing voltage, then inputs will not be reflected in the output until they perturb the resting voltage by ϵ (Figure B-1b). The output of a neuron with a threshold and positive feedback is given by

$$y_i = \max\left[\frac{x_i + \alpha y_i}{1 + k\sum_{j \neq i} y_j} - \epsilon, 0\right]. \qquad \text{[B-13]}$$

Rather than develop the full complexities of Equation B-13, only one simple case is here considered, the case where $\alpha \geq 1$. Consider first a neuron acting in isolation: $x_i = 0$, $i \neq j$. (Note that $x_i = 0$ implies $y_i = 0$.) Because $\alpha \geq 1$, any input $x_i > \epsilon$ causes $y_i \to \infty$, or more realistically, $y_i \to Y$, the maximum possible output. A neuron (i) acting in isolation is either fully turned on ($y_i = Y$) or silent ($y_i = 0$) depending on whether $x > \epsilon$ or $x \leq \epsilon$.

Monoactivity in an interacting system of neurons with thresholds and positive feedback. Let X be the maximum possible input; let Y be the maximum possible output; let $\alpha \geq 1$. Consider first what happens when all inputs except x_i are zero; that is, $x_j = 0$, $j \neq i$. This is the same case as neuron i acting in isolation: $y_j = 0$ (for $j \neq i$) and either $y_i = Y$ (for $x_i > \epsilon$) or $y_i = 0$ (for $x_i \leq \epsilon$).

When more than one neuron receives nonzero inputs, the first neuron (i) whose input (x_i) exceeds ϵ turns fully on ($y_i = Y$). The condition for monoactivity of a network of N-neurons is quite simple; it is that any active neuron has sufficient output to

prevent all other neurons from turning on. For example, if the first neuron (i) whose input (x_i) exceeds ϵ were to produce an output (Y) of sufficient magnitude to keep all the others silent, this would be a highly stable state. It would persist until the input changed. When x_i finally subsided to ϵ, another neuron could take over; it would silence the others, and so on. Sufficient conditions for this kind of monostable interaction are developed below.

Neuron j is maintained in perpetual silence by neuron i when

$$\frac{x_j}{1 + kY} \leq \epsilon. \qquad \text{[B-14]}$$

(Note that the denominator of Equation B-14 contains kY, and not βkY as suggested by Equation B-12, because it is assumed that Y already is the largest value of y_i. The numerator of Equation B-14 does not contain β because it is assumed that neuron j's threshold is not exceeded; therefore there is no positive feedback.) Rearranging Equation B-15 gives

$$(1 + kY) \geq x_j/\epsilon = m. \qquad \text{[B-15]}$$

The left side of Equation B-15 is the total inhibition of neuron j (by neuron i); physiologically, it represents the factor by which neuron i must change the effective conductance of neuron j to silence it. The factor $m = x_i/\epsilon$ is the number of times by which x_j exceeds the threshold (ϵ). Let $X/\epsilon = M$. To guarantee monostability, the effective neuronal conductance change (left side of Equation B-15) must exceed the largest m, giving $(1 + kY) > M$. When this condition is satisfied, neuron x_j is silenced by x_i even when x_j is receiving its maximum possible input.

The value of M depends largely on ϵ; when ϵ is large, M is small and easily achieved. To avoid the possibility of having no active neurons when ϵ is large, it can be assumed that some neurons (i.e., neurons of the middle level of the NBF) receive spontaneously active inputs; these effectively reduce the ϵ requirement without changing M.

To obtain bounds on M, consider the two factors which determine the effective change in neuronal shunting conductance (the left-hand term of Equation B-15). Let the factor $f_1 = (1 + kY)$ be the direct increase in conductance wrought by inhibition. When the site of excitation is separated from the site of inhibition, the effectiveness of the inhibition can be increased; the increased

effectiveness is represented by the factor f_2. Assuming the attainable values of f_1 and of f_2 are about 3 gives a net $f_1 f_2$ of about 9ϵ for the range of opposing inputs that can be totally silenced by an active neuron. For inputs ranging up to 9ϵ, this system is monostable and unbiased. If one or more inputs exceed ϵ, one (and only one) neuron is active; otherwise, all are silent. This system has the properties assumed for it in the neural model.