Int. J. Man–Machine Studies (1983) **18**, 361–389

On the retino-cortical mapping

STEWART W. WILSON

Research Laboratories, Polaroid Corporation, Cambridge, Massachusetts 02139, U.S.A.

(*Received 26 August 1980, and in revised form 22 August 1982*)

Based on Hubel & Wiesel's physiological findings on the projection from retina to cortex, a schematic model of that stage of visual processing is constructed and its properties investigated. The projection or mapping appears to carry out an automatic "normalization of description" for the same object independent of retinal image size. This property suggests new concepts regarding (1) contrast sensitivity, (2) the nature and role of indirect vision, (3) the role of eye movements and (4) the recognition of patterns and the analysis of scenes.

Introduction

This article explores implications for human visual perception of the results of Hubel & Wiesel (1974), in macaque monkeys, on the relationship between receptive field size, magnification and eccentricity. A schematic model of the retino-cortical projection is constructed which seems roughly to include their results, and leads to an hypothesis about the general purpose of that stage of visual processing. The model introduces the notion of "data field" as a generalization of Hubel & Wiesel's "aggregate receptive field", and the notion of "message sending unit" (MSU), as a dedicated cell sub-assembly of which their "hypercolumn" is the prototypical example. The hypothesis about the retino-cortical projection is that it is arranged so that, in the presence of certain world constancies whose retinal images change with viewing distance, the output signals—in the model's terms, messages—of the primary visual cortex are constant and independent of viewing distance. The model is applied to an explanation of the contrast sensitivity function; to Aubert & Foerster's (in Helmholtz, 1962) classic experiment, and Lettvin's (1976) observations, on peripheral visibility; and to some basic questions of pattern recognition.

Physiology and model

In the view of Hubel & Wiesel (1979), the primary visual cortex is both anatomically and physiologically a regular array in which a standard unit of neuronal processing machinery, the hypercolumn, is repeated over and over. Each hypercolumn has some thousands of input fibers which project backwards via the lateral geniculate body to ganglion cells subserving a delimited region of the retina. Each cortical cell measured in a perpendicular penetration of the hypercolumn will respond to a characteristic stimulus over a locale within that delimited region; the precise response locale is termed that cell's receptive field. The fields vary in size for different cells in the hypercolumn by a factor of two or three. The variation is probably because different cortical cells compute different things: for example, the so-called "complex" cells by

0020-7373/83/040361+29$03.00/0

definition have a larger field than the "simple" cells. The receptive fields in one hypercolumn also tend to vary somewhat in position (the positions of their centers); the variation is about the same as the variation in (linear) field size; this is termed "scatter". Hubel & Wiesel call the "pile of superimposed fields that are mapped in a penetration beginning at any point on the cortex the 'aggregate field' of that point". In effect, each hypercolumn may be said to have an aggregate field.

In the model which will now be developed, the hypercolumn is regarded as a "message sending unit" (or MSU) and its aggregate field is regarded as an experimentally observed manifestation of the MSU's "data field", i.e. the retinal region over which the MSU collects stimulus information. Hubel & Wiesel say the hypercolumn has "perhaps 50,000" output fibers. Incorporating this, the model's MSU is an entity which receives input on some thousands of input fibers from its data field, computes simple overall properties of the stimulus (related to, but not necessarily the same as, those that Hubel & Wiesel have identified by probing single cells within the hypercolumn-MSU) and outputs the presence or absence of these properties—"the messages"—as a code on the 50,000 output fibers. In the ensuing discussion, the reader may certainly still think of "hypercolumn" for "MSU" and "aggregate field" for "data field". New terminology has been introduced because the model and the physiology are distinct, and to emphasize suggested function.

In developing the model further, let us go back to Hubel & Wiesel (1974) and examine particularly their fig. 6A (reproduced here as Fig. 1). We want to know more

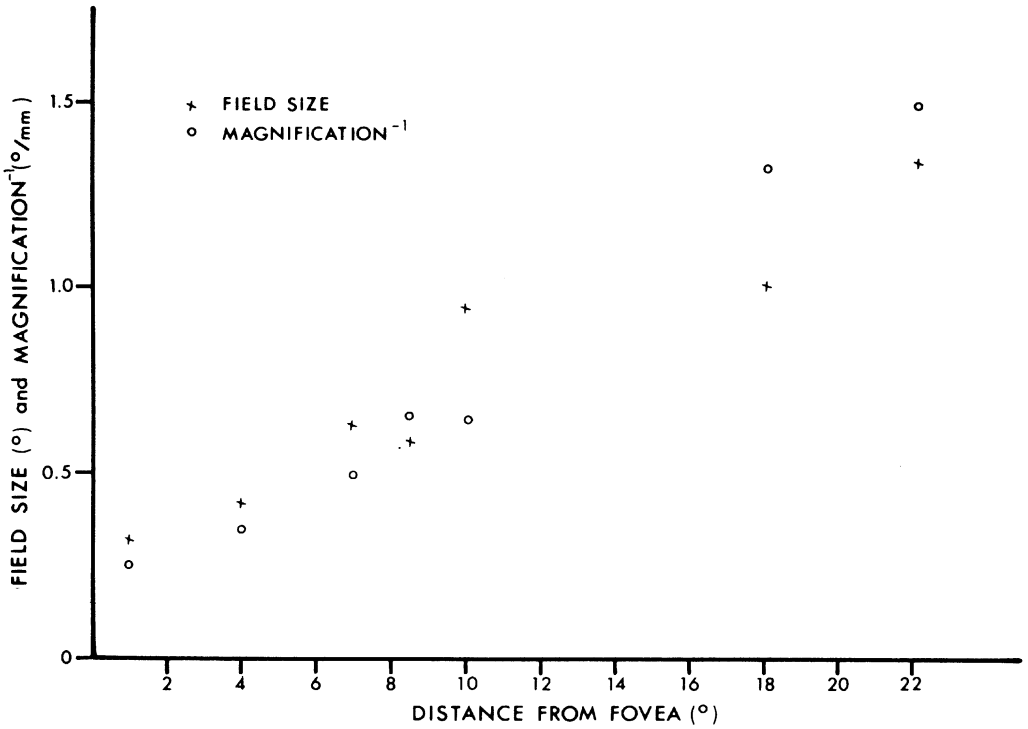

Fig. 6A Graph of average field size (crosses) and magnification^{-1} (open circles) against eccentricity, for five cortical locations. Points for 4°, 8°, 18° and 22° were from one monkey; for 1°, from a second. Field size was determined by averaging the fields at each eccentricity, estimating size from (length × width)$^{0.5}$.

FIG. 1. Field size and magnification^{-1} vs distance from fovea in macaque monkey. [From Hubel & Wiesel (1974), with original caption. Reprinted by permission of Alan R. Liss, Inc., copyright owners.]

precisely the relationship between retina and primary visual cortex. At a given eccentricity, how big is the aggregate field of a hypercolumn, and to what extent do the aggregate fields of adjacent hypercolumns overlap?

Figure 1, while containing just a few hard-won points, gives an answer. (1) Field size increases as a linear function of eccentricity, with a slope of about 0·05 deg/deg, and the field size close to the fovea is about 0·25°. ("Field size" is the average size of the receptive fields comprising the aggregate field at that point.) (2) The parallelism between field size and "magnification^{-1}" means that the spacing between aggregate fields for adjacent hypercolumns is such that the diameters of the two aggregate fields overlap by about 50%.

For the model, it will be assumed that field size vs distance from fovea in Fig. 1 really is a straight line; also that the overlap percentage is constant, independent of eccentricity. As noted, field size in Fig. 1 is about 0·25° at 0° "distance from fovea". Whether this is distance from the fovea's boundary or center is not made clear. The model is neater if one assumes that the slope of the curve outside the fovea would allow straight-line extrapolation to zero field size at exact fovea center.

Suppose now that one wants to diagram the input domain of the retino-cortical mapping on a piece of paper. For this, place the fovea center at the center of the paper and let radial distance from there represent angular distance to a given point on the retina. To show the mapping, one might place dots on the paper, one for each hypercolumn, wherever the center of the hypercolumn's aggregate field falls. But how should the dots be arranged? What is the correct pattern of their spacing?

The above interpretation of Fig. 1, plus the assumption that the arrangement is radially symmetric, gives the answer. One curve says field size (diameter, say) is proportional to distance. The other curve says overlap is a constant percentage of field size. Since overlap is linearly related to both field size and field separation, constancy of percentage overlap requires that the dots be laid out so that *the distance between adjacent dots is also proportional to distance from the center.*

Figure 2 shows the resulting arrangement. Imagine each dot to be encircled by a field which extends more-or-less up to the nearest neighboring dots. One can see that (1) field size is proportional to eccentricity; (2) the fields have constant percentage overlap; and (3) (a consequence of the first two properties) field *spacing* is proportional to eccentricity.

The pattern of Fig. 2 has a fundamental property which we can establish by placing an object on the pattern. In Fig. 3, imagine that the inner square is the outline of some general object which has been imaged on the retina. The object happens to be fixated at its center in this case. The various parts of the object are being examined, so to speak, by the data fields (aggregate fields) of the MSUs (hypercolumns) whose representative dots fall on the object. (It is assumed, as before, that the data field of a dot is a circle extending more-or-less up to the nearest neighboring dots.) The outer parts of the object are examined by fewer data fields per unit area than the inner parts. We can suppose that this object has surface detail of some kind, so that the various data fields see various things "under" them and the associated MSUs send out various messages.

Now, let us magnify the object on the retina somewhat so it now is represented by the outer square. The diagram reveals that for every data field now receiving input from some portion of the object, there was, prior to magnification, another data field

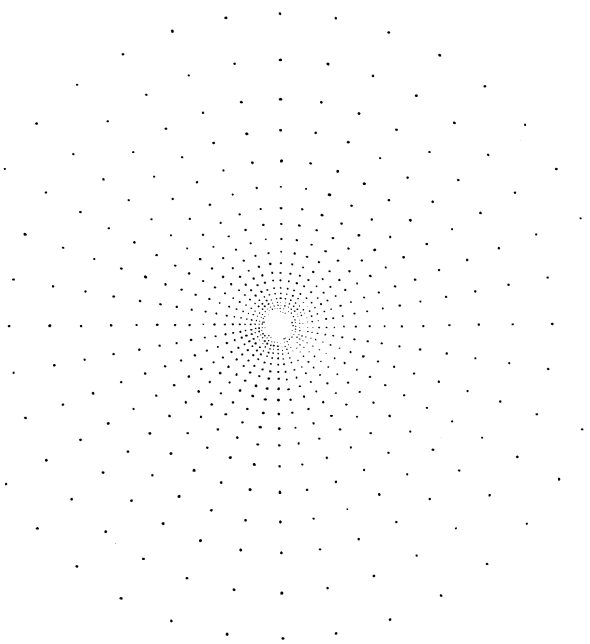

FIG. 2. Arrangement of MSU data field centers in retinal space.

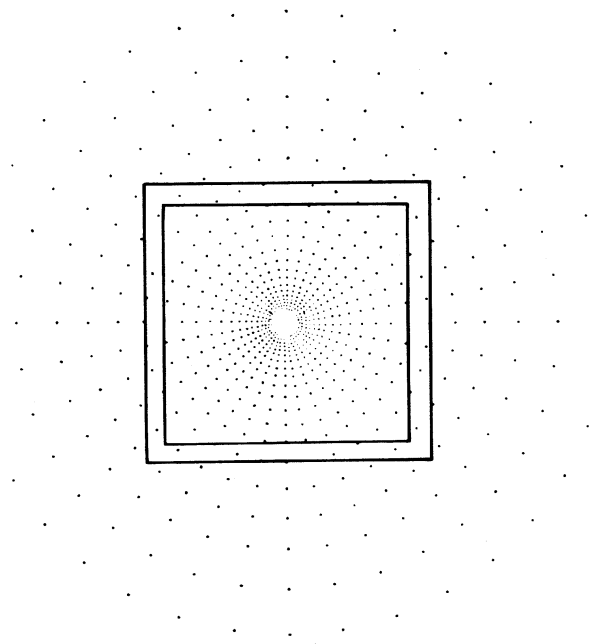

FIG. 3. Array of Fig. 2 with superimposed, centered images of a "generalized object" at two magnifications.

which received its input from precisely the same object portion. Thus, if all MSUs are similar (Hubel & Weisel's assumption for hypercolumns), the *ensemble of messages* output by the MSUs will not differ between the two magnifications. The particular set of MSUs *sending* the messages will shift slightly, but not the messages themselves.

It therefore appears conceivable that the retino-cortical mapping is a system designed to produce a similar output or, so to speak, to "tell the same story", whatever the

size of an object's image on the retina. The output message ensemble may be likened to a set of reports that speak about the various portions of an object, where "portion" has units something like "percent of overall object size". The effect is to extract information about an object's visual pattern, independently of its size in space or the distance at which it is seen. Such a property is consistent, of course, with daily experience. The appearance of an object is substantially independent of its accidental size on the retina.

Indirect vision

In the *Physiological Optics*, Helmholtz (1962) described experiments by Aubert & Foerster on "the precision of vision in the peripheral parts of the retina". In one experiment, they investigated the extent to which, during a brief flash, subjects could identify letters and numerals which occurred at an angle to the direction of vision at the moment of the flash. That is, they asked: what is the maximum off-axis (horizontal, in this case) angle at which a letter of a given size will still be just recognizable? Aubert & Foerster's main result was that, for letters of a given actual size, the maximum angle was proportional to the letter's visual angle, the ratio being approximately five to one.

It may be helpful to put this important result somewhat differently. Imagine the following special kind of eye chart which an eye doctor might have. On a large white card there is a small x on which one fixates, and then, somewhat to the side of the x, a single letter such as "E". Suppose that when you look at the x, you can indirectly see the E just well enough to tell that it is an E. What Aubert & Foerster discovered is that if you can just barely tell the E at this distance from the chart, you can move forward and backward (within fairly wide limits, always fixating the x), and the E will *still* be just barely identifiable.

Some flavor of the experiment may be gained by looking at Fig. 4. If the x is fixated, the appearance of the E will be substantially independent of the distance at which you hold the paper.

FIG. 4. "Eye chart" for indirect vision.

These phenomena are very important for the model of the retino-cortical mapping. In the "expanding diagram" of Fig. 2, suppose the fixated x to be at the center; the image of the E will then fall somewhere away from the center. It will be "under" the data fields of some number of MSUs—more if the E is close to the center, fewer if farther away—but, in any case, some definite number. Now, what will happen as, say, our patient approaches the eye chart, or we move Fig. 4 nearer? The answer is clearly that the E will move farther out from the center; at the same time, it will also grow larger. In fact, its distance and its size will grow by exactly the same factor.

But if this is true, a brief consideration of the properties of Fig. 2 will persuade one that the E, in its new position, will fall under the same number of MSU data fields as before. Put more generally, as long as the point of fixation does not change,

the number of MSU data fields "subtended" by an indirectly seen object is independent of distance.

Now, from Aubert & Foerster's experiment we can draw two important conclusions. First, since they found that recognizability in this situation is independent of distance, one can infer that the MSUs, as information processors, are functionally identical. For, if equal recognizability at all distances goes together with an unchanging number of "inspecting MSUs", then, since the actual MSUs doing the inspecting *do* change, it must be concluded that one MSU is just as good as any other MSU—i.e. they are functionally alike.

There is thus independent support for Hubel & Wiesel's assumption of the similarity of the hypercolumns. But from Aubert & Foerster one can infer something more. Since the "E" on the eye chart may be drawn at such a distance from the fixation "x" that it is *no longer* recognizable, it must be the case that each MSU has a finite and limited information processing capacity. Can we say anything about this capacity, that is, about the specific computations which the MSU performs?

In an article entitled "On seeing sidelong", Lettvin (1976) drew attention to the way in which the visibility of an indirectly seen object depends strongly on the presence or absence of other objects in its immediate neighborhood. Figure 5 gives examples

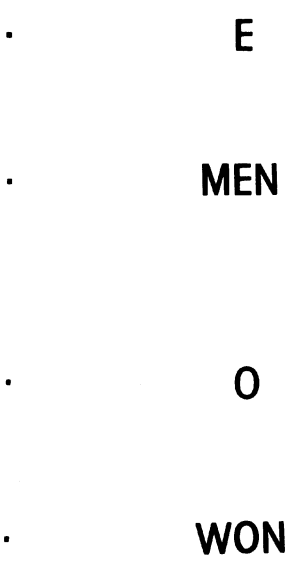

FIG. 5. "Eye chart" for indirect vision with isolated and embedded letters.

of an E by itself and between M and N, likewise an isolated and an embedded O. In each case the associated dot should be used as a point of fixation. After inspecting the figure, notice that the indistinctness of the embedded E and O is not a matter of insufficient basic visual acuity; for, if one fixates the letters instead, the dots are sharp.

The present model seems to shed some light on these observations. In particular, the notion of finite processing capacity in the MSU suggests that the MSUs that are "speaking" about an embedded letter and its environs may be unable to describe "cleanly" a situation of this degree of complexity; whereas, when the letter exists in isolation, the situation is inherently less complicated and the processing machinery may be more adequate to the task of description.

The above is rather vague; but the problem is subtle. Up to a point one might say it is just a matter of "resolving power": the embedded letter is harder to separate from its neighbors; without neighbors, we "see it better". But a more powerful notion, that the MSU is a primitive pattern recognizer, is suggested by the work of Hubel & Wiesel. In the hypercolumn they found cells sensitive to spots, cells sensitive to oriented bars, cells sensitive to a properly oriented bar anywhere within a relatively large area, etc. Let us assume that these stereotyped single-cell responses represent aspects of the MSU's calculation of its output message, and that the message itself is drawn from a quite limited vocabulary. Then, if the MSU sees a stimulus which has simple properties of the appropriate type, its output message will be strong and clear: e.g. "there is a corner", "there is a round patch", etc. But, if the stimulus is not easily summarized in the terms the MSU is designed to detect and express, its message will be indecisive—an outcome whose conscious consequence is the belief that we "can't see the form clearly".

The concept of the MSU as a primitive pattern recognizer implies that it is an information-reducer: the MSU operates on a stimulus so as to describe it in certain simplified or stylized terms. This leads to a further question. If, by detecting and talking in a language of simple forms, the MSU acts as an information-reducer, then for every output message there must be a set of physically distinct stimuli which will all produce the same output. Put another way, are there ways to vary a stimulus seen indirectly such that one would not notice any change? And, what variations produce pronounced changes? Experiments to answer these questions would give further clues as to the forms for which the MSU is looking.

Luminance gradients

To test and extend the ideas developed so far, we shall consider the major phenomena in a rather different area, that of the visibility of sinusoidal luminance gradients. Here, it is *not* always the case that the appearance of an object is independent of distance or subtended visual angle. Instead, there appear to be two regions: for objects of "low" spatial frequency, appearance and threshold visibility are independent of retinal image size; while for objects of "high" spatial frequency, retinal image size matters. Quotation marks have been used because the dividing line is not well defined in the literature; an alternative viewpoint will be offered here in which "spatial frequency" turns out not to be the central variable. In the process, the investigation will identify a widely useful characteristic parameter of the mapping.

A number of people (see McCann, Savoy, Hall & Scarpetti, 1974) have investigated the "high" frequency region. McCann and his co-workers (McCann *et al.*, 1974; McCann, 1978) extensively studied the "low" region and showed that it had to be treated quite differently. There are many experiments and the subject is conceptually tricky to navigate. For simplicity we shall describe a typical experiment and use it in explaining both the phenomena and the new viewpoint.

In general, subjects look at displays or "targets", usually square, in which the luminance varies as a sinusoidal function across the target in, say, the horizontal direction. Figure 6 shows the luminance profile of a typical target. The target may have a uniformly luminous surround: sometimes the surround is black, sometimes it is equal in luminance to the average of the sinusoid, etc. "Contrast" in the display is

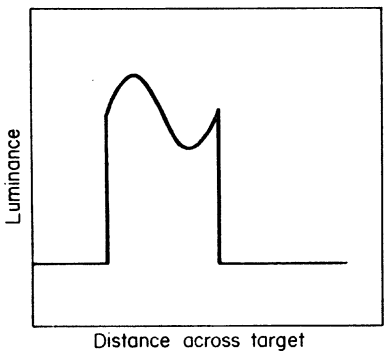

FIG. 6. Luminance profile of a sinusoidal target.

defined by McCann as $(L_{max}-L_{min})/L_{max}$, where the Ls are luminances. "Spatial frequency" means the rate of variation of the target luminance, in cycles per degree measured at the eye. "Retinal gradient" is defined as contrast divided by the retinal angle between L_{max} and L_{min}. "Number of cycles" is just the number of cycles through which the luminance varies in going across the display; it need not be an integer.

Now, suppose in our experiment we use a fixed target size and average luminance, but allow ourselves to change the target's contrast, its number of cycles and the distance at which we view it (angular size on the retina). We will begin with a low number of cycles, say 0·5, set the contrast to zero, and then gradually increase contrast until we just see the luminance pattern. This contrast value is the "threshold contrast" and its reciprocal is "contrast sensitivity".

We repeat the experiment at a different viewing distance and, perhaps surprisingly, find no difference in threshold contrast. We realize that we might have guessed this result: the ensemble of messages output by the MSUs is concerned with the pattern of an object, and is independent of the object's size. Certainly this target, once it is visible above threshold, *looks* the same at any distance. It would be natural to expect the threshold value to be independent of distance, as well.

We now increase the number of cycles in the target, using 1, 2, 3 and 4 cycles. In each case we find a threshold independent of viewing distance, but the threshold becomes lower as the number of cycles increases. Contrast sensitivity is increasing with number of cycles. However, as we proceed above about 4 cycles, contrast sensitivity stops increasing; instead, it plateaus for a while and then begins a steady decline as number of cycles goes higher and higher. Furthermore, we find a point somewhere above 4 cycles where, for any given number of cycles, contrast sensitivity falls as we move away from the target, instead of staying constant as before. Figure 7 shows the contrast sensitivity curve we would typically have obtained in the experiment. The widening band at the upper end shows the effect of distance; the lower lines correspond to the observer being farther from the target.

How, in the light of the retino-cortical mapping, can we explain this overall curve? Two new concepts are needed. The first is "object frequency", which will be represented by f_o. Object frequency means number of cycles per object and is intended to indicate the amount of detail of an object independent of object size, viewing distance, etc. For these targets, object frequency is equivalent to number of cycles. (A more complex target or object would have an "object frequency spectrum", based on the Fourier spectrum of a geometrically similar object whose longest dimension was unity.)

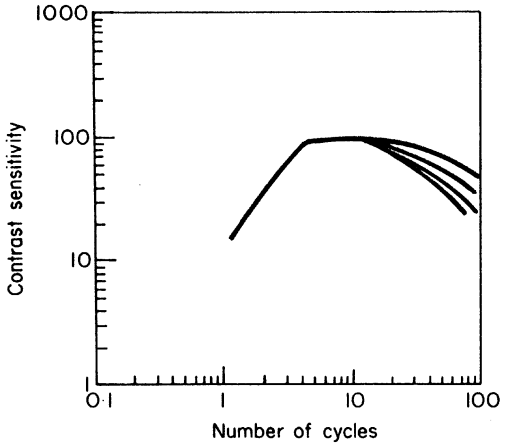

FIG. 7. Schematic contrast sensitivity curves as plotted against number of cycles in the target. Curves for different viewing distances superimpose at low number of cycles, diverge for higher number of cycles.

For the other new concept, recall the expanding data field array of Fig. 2. Note how the field centers form rays and that there is a certain angle between adjacent rays. In fact, if fields overlap 50%, then twice this angle (in radians) is equal to the ratio of field size to distance from the center. Define this ratio as the array's characteristic angle, C_M. The explanation of Fig. 7 will center around the relationship between f_o (a property of the object) and C_M (a property of the perceiver).

Suppose we are observing at the low end of the curve, with, for example, $f_o = 1$ (one cycle). Figure 8 is intended to show the situation on the data field array. There is an outer square, representing the retinal image of the display as seen at one distance,

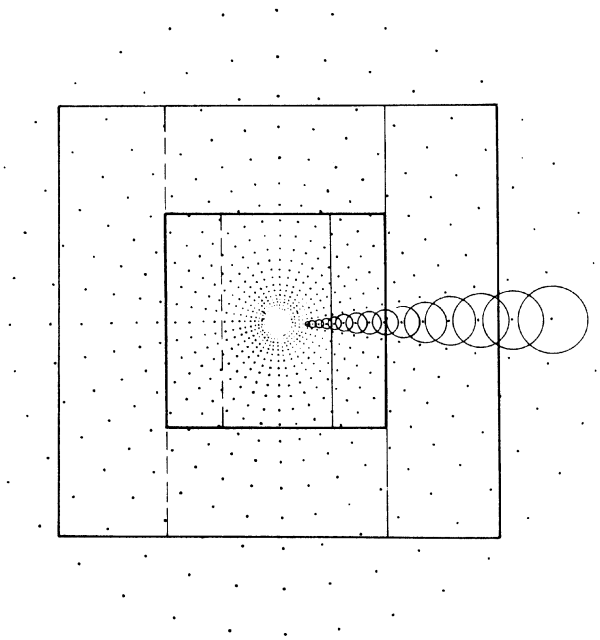

FIG. 8. Sinusoidal luminance gradient target containing one cycle ($f_o = 1$) of sinewave superimposed on the retinal array of Fig. 2. Outer square: the target as seen at one distance; inner square: as seen at twice the first distance. Vertical broken and solid interior lines are minima and maxima, respectively, of luminance. Circles are data fields drawn in for one sector of the array.

and an inner square, representing the image of the display at twice the first distance. Each square has a solid and a broken vertical line. These indicate, respectively, the peak and the valley of the luminance sinusoid. Finally, a sector of the data field array has its fields explicitly drawn in; note that the sector has angle C_M.

Considering first the outer square, let us ask: which field of those drawn in is most strongly stimulated by the sinusoid, in the sense of having the greatest luminance change across it? Clearly it is the one next to the square's edge. That field is both the largest in the sector (of fields completely on the square) and it is positioned close to a place of maximum rate of change of luminance. Thus it is the field (of those in the sector) with the maximum absolute luminance change across it.

Suppose we now gradually turn the display's contrast up from zero. Suppose also that among the computations that any MSU makes is one which notes the overall change in luminance across its data field, and suppose that the MSU sends out a message when that change exceeds a certain ΔL. Under this assumption, it will be the MSU having the field discussed in the previous paragraph which, as we turn the overall display contrast up, first reports that it "sees something". The hypothesis is that this is the point we call "threshold"; that is, threshold is the point at which this "trigger field", as it will now be called, first reports. The contrast at threshold will be that contrast value which results in the luminance change across the trigger field exceeding the field ΔL.

Returning to Fig. 8, we may note that the trigger field diameter is rather small relative to a half-cycle of luminance; if f_o were 2 instead of 1, the trigger field would have a larger luminance change across it and threshold would occur at a smaller value of overall contrast. Since contrast sensitivity is the inverse of threshold contrast we might therefore expect contrast sensitivity to be a rising function of f_o in this part of the curve, and a glance at Fig. 7 shows this to be the case. This part of the curve is thus characterized by the trigger field "seeing" a larger and larger portion of the total luminance change as f_o increases.

Turning to the inner square of Fig. 8, it may be observed that changing the viewing distance has not materially changed the situation. The trigger field is a different one, to be sure, but it is still the one next to the square's edge and it still sees the same fraction of the display. This confirms the result that contrast sensitivity, in this part of the curve, is independent of distance.

Let us move on to Fig. 9. Here $f_o = 4$ and we have reached a critical point. The trigger field is again at the square's edge, but for the first time it exactly covers a half-cycle of luminance. The trigger field, regarded as a detector, has reached maximum efficiency: it spans the display's full luminance change. Thus we would not expect contrast sensitivity to go higher than the value it has here. What, in fact, would be the case if f_o were, say, 5? A slight mental extrapolation of Fig. 9 shows that *the trigger field begins moving inwards*; the field at the edge will now cover *more* than a half-cycle and so will be less efficient than one a little way in from the edge. This latter field, in fact, being smaller, can still cover exactly a half cycle of the higher f_o. The trigger field thus changes or "moves", but still retains maximal efficiency. We would thus expect the contrast sensitivity curve to have a peak which is in fact a plateau, at least for some range of f_o. Figure 7 shows that this in fact occurs.

Returning to Fig. 9, we note that the point at which maximum contrast sensitivity is first reached can be mathematically characterized. It occurs for f_o such that the

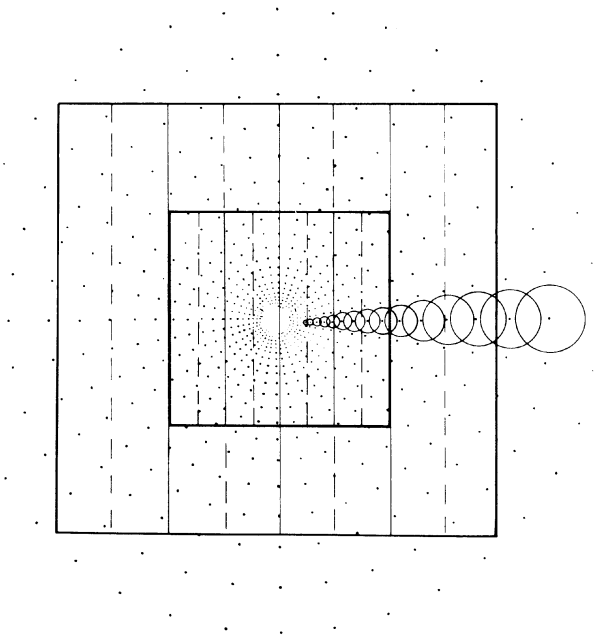

FIG. 9. As Fig. 8 except that $f_o = 4$.

product $f_o C_M$ becomes unity. This may be derived by asking: when is the trigger field diameter first equal to a half cycle of the display? This can be answered directly from Fig. 9. The trigger field diameter is C_M times half the side of the square. The display's half cycle is the full side of the square divided by f_o, all times one-half. Setting these equal gives $f_o C_M = 1$.

For f_o greater than $1/C_M$, we have concluded from the model that contrast sensitivity should hold constant, at least for a while, and indeed it is observed to do so. But Fig. 7 shows that the curve eventually declines. How can we explain this?

The answer comes in examining distance effects. In Fig. 9, doubling the distance (the inner square now applies) changes nothing since the new trigger field still sees the same portion of the display. In effect, everything scales uniformly. As f_o increases, the trigger field will move in from the edge, but the scaling for different distances will not be affected. Or will it?

What happens when the trigger field is so far in that viewing from a greater distance shifts the properly scaled position of the trigger field *onto the fovea*? In Fig. 10, the dot diagram has been enlarged and there is now a "fovea", i.e. a region where data field size and spacing are no longer proportional to distance from the center. The scaling rule cannot continue as before, since the cells are finite in size. A convenient assumption for the model is that the foveal data field size is constant and equal to the field size of the fields just outside the fovea.

Now consider the outer square in Fig. 10. Here $f_o = 8$ and the trigger field is more than half-way in toward the center. If we double the distance, the new trigger field, under scaling, would want to fall in the fovea. Thus, under the assumption above, the most efficient detecting field available has the efficiency of the field at the fovea's edge, and this efficiency is *lower* than that of the trigger field before the distance was doubled. This means that contrast sensitivity is beginning to be affected by distance, which is what we should have according to Fig. 7.

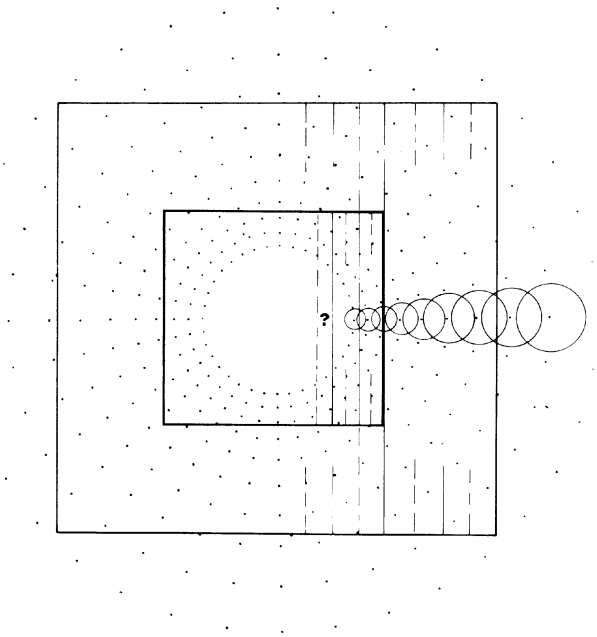

FIG. 10. An enlarged version of Fig. 8 with $f_o = 8$ and the central disk considered to be the fovea.

We can be quite precise about the onset of these effects. If the fovea's half-angle is F, then the angular size of the field at the fovea's edge is FC_M. On the other hand, for a target of angular size S and object frequency f_o, the angular size of one half-cycle is $S/(2f_o)$. The distance begins to matter when these are equal, or for $S = 2f_oC_MF$. If the fovea half-angle is 0·5°, then we may write $S = f_oC_M$. The meaning of this is that if you do the experiment of Fig. 7 with a target S degrees in size, the contrast sensitivity curve will start to decline at $f_o = S/C_M$.

We can get something more out of this formula. It tells the conditions under which the most efficient available gradient detector of the model becomes fixed in size: i.e. the size of a foveal data field. But the implication of the most efficient detector *being always the same angular size* at this point on the curve and beyond is that threshold can henceforth only be a function of the luminance change over a fixed angular increment; that is, a function of spatial frequency. Thus the condition $f_o = S/C_M$ marks the beginning of the region where contrast sensitivity depends solely on spatial frequency. In fact, since f_o/S is by definition spatial frequency f_s, we may also describe the critical point as occurring at $f_s = 1/C_M$, a formula which will be useful later.

Let us look at some real data! Figures 11 and 12 are from McCann (1978). Figure 11 plots contrast sensitivity vs number of cycles (f_o) for targets of various angular sizes. We note that the two smallest targets were smaller than the fovea. Let us ignore them for the moment. The three curves for larger targets stay together up to about $f_o = 4$. The model says this point occurs for $f_o = 1/C_M$; therefore we shall take $C_M = \frac{1}{4}$ rad. Next, the theory says the curves will start to fall from their plateaux at $f_o = S/C_M$. Plugging in, we have the curve for the 16° target turning down at $f_o = 64$; that for the 7·6° target starting down at 30·4; and the drop for the 2·7° target beginning at 10·8. The actual facts are not extremely different from these values.

Figure 11 is actually a replot of the original experimental data shown in Fig. 12. In Fig. 12, contrast sensitivity is shown vs spatial frequency (f_s) in cycles/degree. The

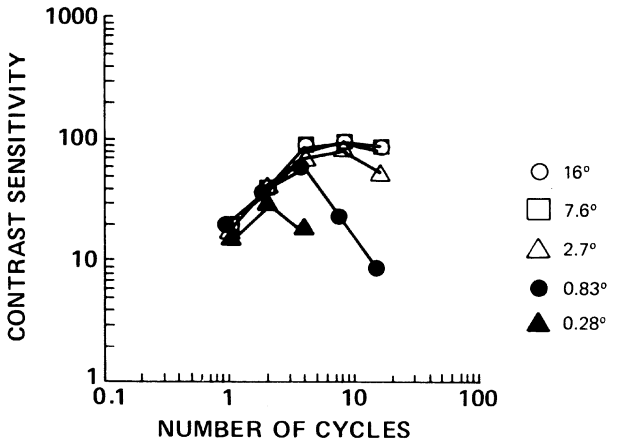

FIG. 11. Contrast sensitivity vs number of cycles (f_o) for five differently sized targets. Data are the same as for Fig. 12. [From McCann (1978). Reprinted by permission of the Society of Photographic Scientists and Engineers, copyright owners of *Photographic Science and Engineering.*]

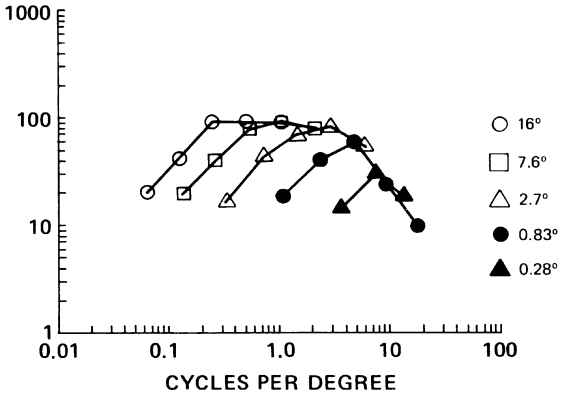

FIG. 12. Contrast sensitivity vs spatial frequency (f_s) for five differently sized targets. [From McCann (1978). Reprinted by permission of the Society of Photographic Scientists and Engineers, copyright owners of *Photographic Science and Engineering.*]

model predicts that contrast sensitivity will begin to depend on spatial frequency—regardless of target size—for $f_s = 1/C_M$ (for a fovea half-angle of 0·5°), or $f_s = 4$. That seems to be quite near where the curves do gather together in Fig. 12.

Let us return now to Fig. 11, and the two curves for targets smaller than the fovea. Assuming still that data field size is constant in the fovea, and equal to the data field size at the fovea's edge, we may ask: how large actually is it? From the other curves, we have $C_M = \frac{1}{4}$. If the half-angle of the fovea is taken to be 0·5°, then the field size on this argument must be 0·125°. We would thus expect the fovea to be maximally sensitive to displays with half-cycle widths of 0·125°. In Fig. 11, the peak in the 0·83° target curve occurs at 3·5 cycles. The half-cycle width at this point is therefore 0·12°. The 0·28° target curve peaks at 2 cycles. The corresponding half-cycle width is 0·07°. Given our uncertainty about what is actually going on in the fovea, the agreement of these numbers with 0·125° is encouraging.

The number of data fields

It is of some interest at this point to calculate the model's estimate of the total number of data fields or MSUs in the human system. A calculation is now feasible because, from our interpretation of the sinewave gradient data, we have a value for C_M, the characteristic angle of the data field array (Fig. 2). The calculation is useful as a cross-check on the theory developed so far, and because it brings out strikingly the rapid decline in data field density as one goes from the fovea into the periphery.

In Fig. 2, the dots represent field centers. Assuming a 50% overlap of fields, the angle C_M is, as explained earlier, twice the angle between adjacent "rays" of field centers. Figure 2 can also be seen as consisting of "rings" of field centers. The first step in the calculation is to notice that each ring has the same number of fields. Then, since there are 2π radians in a ring, if the angle between field centers on adjacent rays is $C_M/2$ rad, there are

$$4\pi/C_M \text{ fields per ring.} \tag{1}$$

How many rings are there? Let us imagine that the blank spot in Fig. 2 is indeed the fovea, and that the smallest ring shown is right at the fovea's edge (we shall count foveal fields later). Call this the "zeroth" ring and let its radius be R_0 degrees (of visual angle).

What is the radius R_1 of the next ring? From the earlier discussion, we want the separation between field centers to be proportional to distance R from the center of the array. A convenient way to state this is

$$R_{n+1} - R_n = a R_n, \tag{2}$$

where a is some constant. Therefore the distance between the first two rings is given by

$$R_1 - R_0 = a R_0, \tag{3}$$

which says that

$$R_1 = R_0(1+a). \tag{4}$$

By inspection, we see that this and equation (2) imply that, in general,

$$R_n = R_0(1+a)^n. \tag{5}$$

To get the value of a, we refer to Fig. 2 and observe that we want the actual distance between rings to be essentially the same as the actual distance between rays at that point. At R_0, the distance to the next ring is, from equation (3), given by aR_0. But also at R_0, the distance between rays is $(C_M/2)R_0$. Thus we have $a = C_M/2$, and equation (5) becomes

$$R_n = R_0(1+C_M/2)^n. \tag{6}$$

This gives us the radius of the nth ring. But, looked at the other way round, equation (6) also says that within radius R_n, there are exactly n rings. Thus, to find the number of rings out to radius R, we solve equation (6) for n and get

$$N(R) = \frac{\ln (R/R_0)}{\ln (1+C_M/2)}, \tag{7}$$

where N(R) has been written for n. To get the total number of fields on these rings, we multiply N(R) by expression (1).

The fovea must be treated differently. The assumption has been that the foveal data field density is constant and equal to the density at the fovea edge. To get the latter, we first find the area of a field at the edge. By definition of C_M, the field diameter is R_0C_M and thus the area is $\pi(R_0C_M)^2/4$. Ordinarily, one would just invert this to get the density. But since the fields are assumed to overlap by 50%, the density is actually four times greater. Thus the density becomes

$$\sigma_F = \frac{1}{\pi}\left(\frac{4}{R_0C_M}\right)^2 \tag{8}$$

fields per square degree.

From this we can get the number of data fields in the fovea,

$$N_F = \pi R_0^2 \sigma_F = \left(\frac{4}{C_M}\right)^2, \tag{9}$$

which, it is interesting to note, is independent of the fovea's size.

Combining equations (9), (1) and (7), we get a complete expression for the number of fields out to radius R:

$$N(R) = \left(\frac{4}{C_M}\right)^2 + \frac{4\pi}{C_M}\frac{\ln(R/R_0)}{\ln(1+C_M/2)}. \tag{10}$$

One further expression is interesting to have, namely the density of fields *at* the radius R. By the same reasoning used in deriving equation (8), this is

$$\sigma(R) = \frac{1}{\pi}\left(\frac{4}{RC_M}\right)^2, \tag{11}$$

or, alternatively,

$$\sigma(R) = \sigma_F\left(\frac{R_0}{R}\right)^2 \tag{12}$$

fields per square degree [an expression which may be integrated to give an alternative, but almost identical, expression for N(R) as equation (10)]. For the constants in equations (10)–(12) we may choose $C_M = \frac{1}{4}$ as found from the gradient data, and a fovea half-angle $R_0 = 0{\cdot}5°$.

Table 1 shows N(R) and $\sigma(R)$ for values of R from 0·5° to 90°. In carrying R out to 90°, we are of course quite beyond the limits of the "flat retina" approximation of Fig. 2. However, the highly asymptotic behavior of N(R) beyond about 10° suggests that recalculating for a spherical geometry would not yield important differences at this stage. In Table 1 it is striking that within the first 10° or so, N(R) has climbed a major part of the way toward its 90° value. Correspondingly, the field density falls at such a great rate that beyond about 10° it takes a large increment of angle to add an appreciable number of new fields to the total.

This section began with the aim of estimating the total number of fields. From Table 1, a reasonable figure might be about 2200 fields, or roughly the value for R = 45°. But the calculation depends, and perhaps sensitively, on the values used for R_0 and

TABLE 1

Number of data fields $N(R)$ between the center of vision and eccentric angle R (in degrees), and density $\sigma(R)$ of data fields (in degrees^{-2}), with $R_0 = 0{\cdot}5°$ and $C_M = 0{\cdot}25$ rad

R	N(R)	σ(R)
0·5	256	326
1	552	81·5
2	848	20·3
5	1239	3·26
10	1534	0·82
15	1708	0·36
30	2003	0·09
40	2126	0·05
45	2176	0·04
50	2221	0·033
60	2299	0·023
70	2365	0·017
80	2422	0·013
90	2472	0·010

C_M. To see the effect of changing them, we may calculate N(45°) for values of C_M and R_0 differing by factors of two on either side of 0·25 rad. and 0·5°, respectively, as shown in Table 2. There is nearly a factor of 20 range between extremes.

To check whether these numbers are anywhere near correct we may again make a connection with the work of Hubel & Wiesel. The number of hypercolumns times the area of cortex per hypercolumn (a constant) should give the total area of primary visual cortex. Hubel & Wiesel (1979) seem to have settled on about 1 mm^2 per hypercolumn, in monkeys. (The hypercolumn has halves for each eye, but that does not affect this calculation.) If the human system also has 1 mm^2 hypercolumns, then Table 2 suggests that the total primary area—both sides of the brain—should be somewhere between 500 and 9600 mm^2. Hubel & Wiesel give 3000 mm^2 as an approximation to the actual total human area. This figure is not greatly different from

TABLE 2

Values of $N(45°)$ under various assumptions for R_0 and C_M

		C_M (rad) 0·125	0·25	0·5
R_0 (degrees)	0·25	9635	2472	649
	0·5	8485	2176	571
	1	7336	1881	492

the "middle value" (2176) in Table 2, which lends support to the model and to the value of $C_M(\frac{1}{4})$ which has been employed.

A further element of support comes by noting, in Table 1, that the number of data fields in the fovea is about one-tenth of the number for the whole retina. This is in good accord with the fact that the fraction of optic nerve fibers serving the fovea is also about one-tenth.

The cortical image

So far in this study of the retino-cortical mapping, we have not considered how the information originating at the retina might be physically arranged on the cortex. That is, we have discussed the hypercolumn-MSUs, their data fields and how the data field centers are arranged on the retina (Fig. 2); but we have not inquired how the MSUs are actually laid out in the cortex.

One might first propose that the MSUs are just arranged randomly, a notion which is not so odd as it might seem. After all, if the overall system needs to keep straight "what is next to what" in the visual world, it need not go so far as to use actual physical adjacency in the cortex for this purpose. The MSUs could instead be arranged irregularly or even randomly, with subsequent cortical areas "wired" to the primary area in such a way that adjoining parts of the visual world are treated as adjoining, even though their MSUs are not physically next to each other.

This hypothesis seems unnecessarily complicated. It would seem, in fact, that a very good thing to do would be to keep adjacent visual world information adjacent in the primary cortex precisely because, if adjacency is important, subsequent wiring would be greatly simplified. Thus we would expect that the MSUs would be laid out so as to preserve the connectedness of the visual world.

How then should the MSUs be arranged? A way to approach the question is to think of the dots in Fig. 2 as being of constant, fixed size, and to imagine shoving them together until each just touches its neighbors. The resulting mosaic could represent the corresponding hypercolumn-MSUs in the cortex: MSUs of constant size and spacing as found by Hubel & Wiesel. Thus we might guess that the cortical arrangement simply condenses Fig. 2 down to equal spacing (it would separately condense the two halves of Fig. 2 to form the left and right primary cortexes).

But unless the condensing is done in just the right way, connectedness will get lost in the process. This can be seen by recalling, in Fig. 2, that each "ring" contains the same number of dots. If rings are simply pushed together toward the center, the dots of one ring must begin to interpenetrate the dots of adjacent rings since the total perimeter over which the dots are spread keeps growing and the number of dots per ring does not. Interpenetration of rings evidently destroys connectedness.

Is there another way of pushing the dots of Fig. 2 together? We have already numbered the "rings" of Fig. 2 from zero upwards. Number also the "rays" starting with zero for the ray at 12 o'clock and continuing up to the ray at 6 o'clock. (Number correspondingly clockwise and counterclockwise for the two visual world halves.) Define a *rectangular* co-ordinate system—a sort of checkerboard—in which the x-axis represents ring number and the y-axis represents ray number. The dots for each visual world half will now map onto that half's checkerboard preserving connectedness.

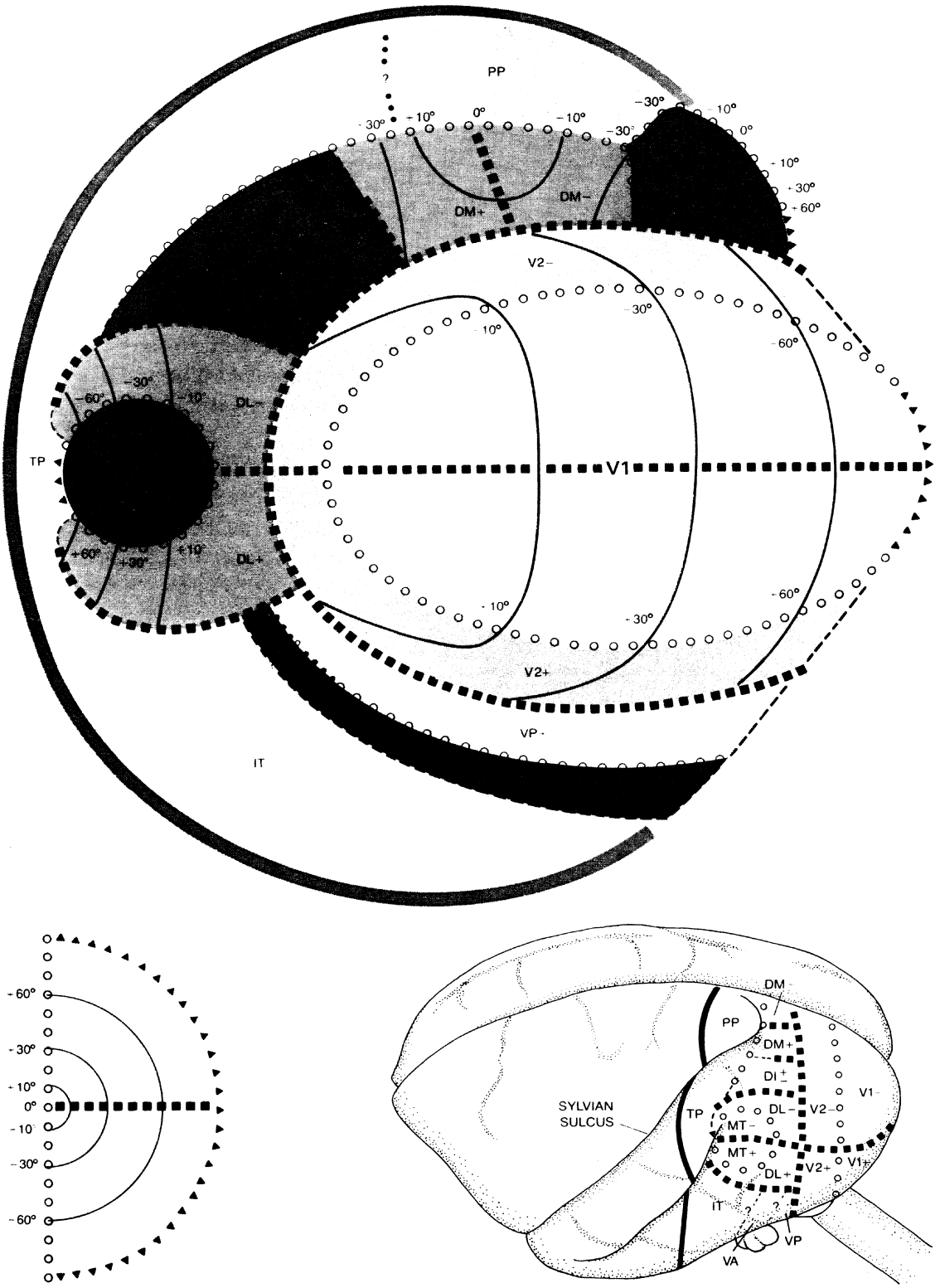

FIG. 13. Visual cortex of the Owl Monkey showing retino-cortical mapping relationships found by John M. Allman of California Institute of Technology and Jon H. Kaas of Vanderbilt University. Please see text for explanation. [From "Thinking about the Brain" by F. H. C. Crick. Copyright © 1979 by Scientific American, Inc. All rights reserved.]

Furthermore, because they form a checkerboard pattern, we have not lost the required equal size and spacing of the cortical MSUs.

Crick (1979) shows an illustration of the retina-to-cortex mapping in the owl monkey, as measured by Allman & Kaas. The illustration is reproduced here as Fig. 13. The large elliptical area labeled V1 is the (left-hemisphere) primary visual cortex. In the lower left part of the figure appears a representation of the rignt visual field. By relating the contours formed, respectively, of squares, triangles and circles in the two regions, and by relating the 10°, 30° and 60° rings, one can get an idea of the actual mapping.

To a first approximation it would seem, remarkably enough, that in the owl monkey Nature is doing something very like the "ray-and-ring to checkerboard" mapping described above. If we assume that the left-hand end of the cortical ellipse contains the foveal field MSUs (which would map without spatial distortion), and if we allow for a little rounding at the ellipse boundary, then it looks very much as though rays map into parallel horizontal lines and rings map into parallel vertical lines.

Whether the spacings are just right, as well, we cannot definitely tell from the diagram. That is, (1) do equally spaced rays go into equally spaced horizontal lines and (2) do exponentially spaced rings go into equally spaced verticals? On the first question, the diagram provides almost no information. On the second, it is ambiguous. Nevertheless, there is enough shown to suggest that deeper investigation may provide positive answers to both questions. It would appear that the owl monkey mapping lends support to the model that has been developed.

To gain insight into the properties of the ray-and-ring to checkerboard mapping it is helpful to see its results "pictorially". Quotation marks are used because, while the mapping's input—an image on the retina—is a sort of picture, the mapping's output is certainly not one. Rather, the output is an array of messages (in a largely unknown language) emanating from the hypercolumn-MSUs of the primary cortex. Nevertheless, the first thing we need to comprehend about the mapping is the way in which it transforms spatial relations: that is, if an object on the retina has a certain shape, what is the spatial "shape" of the transformed object (i.e. its pattern of activity) on the cortical surface?

To answer this, we can map what is perhaps the simplest feature of an image: the local average intensity or gray level. With the aid of a computer, we can effectively superimpose the array of Fig. 2 on a photograph, centering the array on some chosen point (which becomes the center of vision). For each dot in the array we can compute the average gray level in a circular neighborhood about the dot. We let the diameters of these neighborhoods grow proportionately to distance from the center of vision, so that each neighborhood corresponds to an MSU data field.

Next, we transfer the gray level averages to an output array of cells which are indeed arranged like a checkerboard. In transferring we obey the rule of the mapping: the information from the field centered at ray i, ring j goes to the cell at row i, column j. To "see" the results of the mapping, we construct a display consisting of rows and columns of small squares, then set the gray level of each square equal to the numerical value in the corresponding cell of the output array. The display can be photographed and compared with the original image.

Figure 14(a) shows the results of this process for an input picture which is a close-up of a woman's face. The retinal array was centered on the highlight in the right eye

FIGS. 14(a)–(d). Left column: retinal images of a woman seen at four distances. Right column: corresponding cortical "pictures". Fixation point is the highlight in her left eye.

(her left eye). The curious bow-tie shape of the output cortical display arises as follows. Each half corresponds to the primary visual cortex on one side of the brain. Further, each half consists of a rectangular part and a triangular part. The rectangular part is the rectangular array just described (for half the visual field). The triangular part corresponds to half of the fovea (discussed below).

To aid visual interpretation the two halves have been arranged in the picture with the foveal triangles adjacent. In fact, according to Fig. 13, the foveal areas are at opposite ends of the two cortical strips. In addition, the cortical images are inverted with respect to the correct orientation as given in Fig. 13. Finally, the cortical half on our left in Fig. 14(a) belongs in the right hemisphere, etc. If we were to rotate each cortical half through 180° and look through the back of the picture, the cortical and image orientations would be correct as seen from behind the brain.

The foveal mappings come about as follows. A constant density of MSU data fields has been assumed in the fovea, equal to the density at the fovea's edge. Consistent with this, R. C. Crumrine (who created this and the following cortical pictures) spaced the foveal field centers at equal intervals along equally spaced rings. But this implies, for example, that on a ring half way between the center and the fovea edge there are half as many fields as are on the ring right at the edge. Thus, the number of fields per ring is proportional to distance from the center. Now, if the fovea is split vertically and the semicircular rings of each half are straightened into columns, the resulting figure is a triangle. In Fig. 13, the cortical area associated with the fovea is roughly triangular.

One further point on technique: in the photographs, the right-hand half of the cortical image is shorter than the left half because the rest of the right half maps mostly black (the black area on the woman's left). By not displaying that part it was possible to enlarge the image slightly.

Figures 14(a)–(d) show retinal–cortical pairs for the woman as seen at four different distances, centered on the highlight in the right eye. Looking at the cortical images, we see all parts of the woman's face. The distortion is severe: one eye is gigantic, the nose–mouth line has become practically horizontal, etc. Only the other eye, by accident (it is on a horizontal ray with respect to the fixation point), looks nearly normal.

Despite the distortion, one quickly perceives the striking regularity of the series from Figures 14(a) to 14(d). It looks as though the face—or at least the part from the nose–mouth line leftward—maintains the same size and (distorted) shape on the cortex, even though the retinal image size changes by a large factor. Only as the retinal image becomes so small that much of the face falls on the fovea (Fig. 14(d)) does the cortical image change markedly. We note, further, that the cortical face slides horizontally along the cortex as the retinal face changes size.

It would appear, therefore, that one of the properties of the retino-cortical mapping is to transform image size variations about a fixed position on the retina into positional variations of an image of fixed size on the cortex. Put somewhat differently, the group consisting of contractions and dilations ("zooming") about a fixed point is mapped into a one-dimensional translation group.

A second property of the retino-cortical mapping is seen when the retinal image is *rotated* about the center of vision. We recall that retinal "rays" go into horizontal lines on the cortex; thus rotation makes the cortex image slide vertically. For example, if the input image of Fig. 14(a) were rotated clockwise about the right eye highlight, the cortical image would rise in the left half and fall in the right half. Lines of the

image that disappeared from one half of the cortical image would re-emerge in the other half. Thus, rotations about a fixed point are also mapped into a one-dimensional translation group which is orthogonal to the first property's translation group.

The mapping has behaved rather neatly so far. It is now time to acknowledge an apparently disastrous property: cortical image size and shape are not at all constant with movement of the center of vision. Figure 15 shows the result of applying the

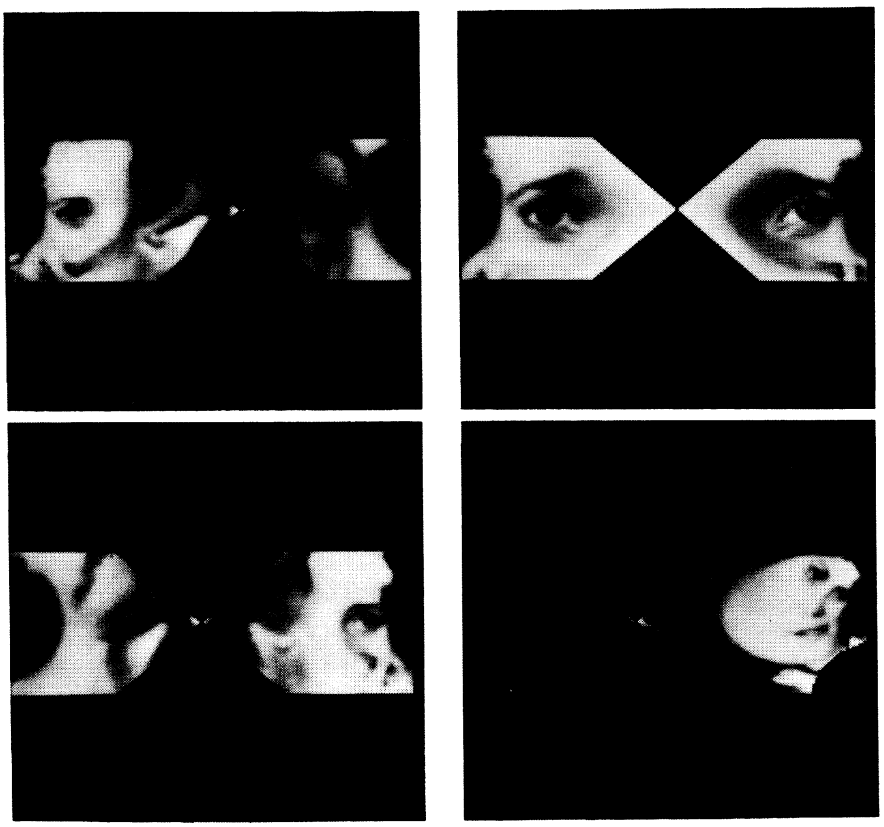

FIG. 15. Cortical "pictures" of the woman of Fig. 14 with four different fixation points. Clockwise from top left, fixation points are: (her) left eye, bridge of nose, right earring, right eye. The woman is seen at the same distance as in Fig. 14(b).

mapping to the input picture of Fig. 14(b), with the center of vision on the right eye, the nose, the left eye and an earring. Clearly, the cortical image of any chosen facial feature varies wildly with fixation point. Is the visual system really equipped to handle the same object under so many different cortical incarnations? Are the size and rotation properties of the mapping worth such a price? In short, do we really have this kind of mapping in our heads? The experimental evidence cited so far indicates that we do. However, there is another kind of evidence, in contrast quite speculative, which also supports the case for the mapping. The next section will attempt to suggest how the visual system, now considered in its role as a perceptual or pattern recognizing system, could find the characteristics of the mapping natural and advantageous.

Pattern recognition

The size and rotation properties just discussed would certainly seem to offer advantages for pattern recognition, for they suggest "normalization" processes: i.e. the reduction of a range of variation to a standard form. However, there appears to be more to the mapping's contribution than just this. With the addition to the model of a small amount of further structure, and by making a connection with voluntary eye movements, we will obtain some important further results.

To prepare the ground for the augmented model, the reader is asked to participate in some simple seeing and looking experiments. For equipment one needs only a large (and unfamiliar) book or magazine having pictures; *Life* or *Paris Match* are excellent for the purpose.

In the first experiment we will make some observations on peripheral vision. Place the magazine closed on the table in front of you and slide your hand somewhere between the pages. Now gaze at your index finger as though you could see it through the magazine (you will find this to be possible) and, with your other hand, open the magazine to that place. Without moving your eyes, observe what you can—and cannot—make out in the periphery.

Most people are able to see very little, except near the finger tip. At the same time, certain *kinds* of things seem nevertheless to stand out, even when quite far from the center of vision. To make what is possibly an oversimplification: the things that stand out are *edges*, particularly those that close on themselves to form a round object. However, to be visible peripherally as a distinct entity an edge or round object must be relatively separate from other similar edges or forms, the more so as the edge is farther from the center of vision.

This has a familiar ring based on the earlier discussion of indirect vision. But doing the experiment shows convincingly that when fixating a point, one's ability to describe most of the rest of a scene is really exceedingly limited and, apart from lightness and color, is restricted to a vocabulary consisting essentially of "edge", "round" and a few degrees of textural coarseness.

For the second experiment, find in the magazine the image of a relatively simple, though still interesting, object such as an automobile or a large capital letter. Look at it and notice that when it is your intention to see such an object as a whole, you tend automatically to direct your eyes so that the object is more-or-less centered around your line of vision. This seems particularly true with letters; to view a letter "as a whole" from an off-center gaze point seems peculiar and even uncomfortable. But the same holds to a greater or lesser degree with other objects, especially those having familiar shapes. There seems to be something special about centering.

In the third experiment, look again at the previous image; suppose it was that of a car seen from the side. Look at some *part* of the car, say a wheel or the rear side-window. You may find that as you now center on the part, the car as a whole, although perceptually still present, has in a strange but nevertheless definite sense moved into the perceptual periphery. It is as though the act of centering on an object or object-part has the corollary effect of imposing a certain ordering or emphasis ranking on other objects or "object-wholes" in the scene.

The retino-cortical mapping model will now be extended in a way that reflects the foregoing experiments and also provides a greater degree of definiteness about the

normalization processes sketched earlier. Some of the assumptions to be made will have the character of hypotheses that should eventually be testable.

Human pattern recognition surely implies learning. The vast majority of the patterns we respond to were learned at some point. Learning evidently involves the generation of some sort of description of the objects learned. The arrangement or relationship of descriptions in the overall structure of memory is not known. Yet it seems reasonable to suppose that every new object attended to gets described in some language, the description may be stored, and when that object or one like it is subsequently recognized, a match has in some sense been made with the stored description.

The *first assumption* in extending the model is that *there is a high degree of similarity between (a) the "vocabulary" used for these stored descriptions, (b) the vocabulary of the MSU reports and (c) the minimal distinct visual elements that we are able to make out peripherally.* Specifically, the stored description is assumed to use the MSU report vocabulary directly, and it is further suggested that we can guess the elements of this common vocabulary by looking carefully in peripheral vision. When we look there, the minimal distinct forms seem to be the edge and the round object, and we also have crude sensations of texture. [Lightness and color are certainly visual elements; but they seem to be non-locally determined, as discussed by Land (1977), and will be omitted here.]

The *second assumption* is that, while for well-learned objects we may have a number of stored descriptions, *there is a principal stored description which consists of the ensemble of MSU reports that the mapping produces when the object is centered about the line of vision.* This ensemble, of course, by the basic properties of the mapping and of the MSUs, is size-normalized; that is, it gives the same cortical picture independent of retinal size. Note that this means that not only the spatial extent of the cortical "picture", but also the specific ensemble of reports within it, is independent of retinal size.

It is worth asking at this point about the grain of the principal stored description. In other words, how finely detailed is the description? Since, according to the first assumption, the MSU reports constitute the elementary vocabulary, this question amounts to asking: how many reports make up the description? Or, since for a centered object the density of MSUs near the object's edge is least—and therefore might best be said to measure graininess—what is that density?

We can get at the answer by supposing first that the object is exactly circular. Then, we may note that by the definition of the angle C_M, there will be something of the order of $4\pi/C_M$ (depending on overlap, etc.) fields inspecting the perimeter of the circle regardless of the size of the circle. What if the object is not circular? In this case, a little experimentation with the array of Fig. 2 will persuade one that, unless the object is exceedingly irregular, the number $4\pi/C_M$ still gives a good approximation to the number of fields on the object's perimeter.

Thus we have the important result that *the grain of all stored descriptions is the same*, and is proportional to $1/C_M$, the reciprocal of the MSU array's characteristic angle. This has an interesting consequence. Suppose an object has a local part which is too small or too detailed to be made out well when the object is seen in a centered view. This would be the case if the angle which the part subtended with respect to object center were less than C_M. To see the part better the eye might move to and center on it. The striking thing is that now the eye will see the part just as well as it

previously saw the whole object—since, as just found, all descriptions have equal grain. The system seems designed to treat parts and wholes with equal processing efficiency.

A further observation related to grain shows an important pattern recognition advantage of the mapping as contrasted with a visual system having uniformly spaced and sized fields. Suppose one is shown a relatively interesting object, for example a doll, at a distance of 10 ft. Suppose one centers on it, one's MSUs inspect it and their reports are duly filed away. A week later the doll is seen again, but at a distance of 2 ft. The MSUs again inspect and report, the set of reports exactly matches the set that was stored, and the doll is recognized. At both ends of the process the MSUs have gone through a perfectly routine procedure. Although the scales differ by a factor of five, the results match.

Very different would be the situation with a uniform array of equal sized fields. Now, the set of reports for the doll's larger image is in general entirely different from before. To achieve recognition, it is necessary to *interpret correctly* parts of the image as being the same as parts before, except five times larger. But how can this be done since the system does not know what scale factor to apply? One might propose dividing by the doll's overall size so as to get a "normalized" description. However, the doll would typically be presented against a varied background, which would make determining its size—prior to recognizing it—essentially impossible. For the retino-cortical mapping, these problems do not arise.

The *third and fourth assumptions* in extending the model address (1) the fact that size invariance still leaves images which vary in position on the cortex—thus it is not quite obvious how to recognize them by simple matching and (2) the fact that off-center images of an object differ tremendously from the centered image—thus it is not obvious how they can be recognized at all.

A solution to the first problem is to *assume the existence of a process which scans the cortical image in time.* Imagine a pair of vertically oriented "scanning bars" which are initially superimposed at the middle of the bow-tie cortex of Fig. 14, but then separate and move at constant speed until they reach the ends of the figure, at which point they return instantly to the middle and repeat. As the bars move over the cortex they read out, in parallel, the MSU reports of each successive column. From the viewpoint of retinal space, the scanning would be equivalent to a circle which starts at the center of vision, dilates, reaches a maximum radius, then shrinks instantly to the center and repeats. The rate of dilation would be exponential to correspond to a uniform rate of scanning on the cortex.

Now, imagine that the cortical scanning process feeds a multi-channel shift register, where the number of channels equals the number of cortical rows and the length equals the number of columns. If we now scan each of the cortexes in Fig. 14, the distorted image of the woman's face will eventually pass any particular point along the shift register at which we might choose to set up an observation station. If we had a "template" of the face, and kept watch over a section of the shift register as large as the template, it is clear that the template would eventually match the reports in the shift register. Thus, the woman would be recognized, independent of retinal size and cortical position. If we wished to know the retinal size, we could get that from the phase of the scanning process at the instant of match. (The reader will note that we have not assumed the existence of a second scan, orthogonal to the first, for

normalizing image rotation. Two scans at once would certainly be complex. Fortunately for the model, human visual perception seems only very weakly rotation invariant.)

The scanning assumption is similar to a proposal made by Wiener (1948) to account for size invariance. Referring to an idea of McCulloch (Pitts & McCulloch, 1947), Wiener suggested that the different layers of the cortex, when successively excited by a scanning process, would function upon the retinal image like a variable zoom lens, with some further station performing a template match when the image zoomed through the correct standard size. In the light of Hubel & Wiesel's work, we would find this model hard to accept today. But the scanning assumption of the present paper is clearly a relative in some degree of the scanning proposal made by Wiener and McCulloch. Furthermore, Wiener's suggestion that the scanning frequency might be related to the alpha-rhythm remains provocative.

Returning to the shift register, let us imagine that there is a standard section of it where *all* templates of learned objects wait looking for their own images to go by. Suppose, once again, that we are looking at the wheel of a car. What will occur at the shift register's matching area as the scan proceeds through its cycle? Early in the cycle the cortical image of the wheel will go by. Since that image is centered and presumably very well learned, it will be strongly matched by an appropriate template. Later, the image of the outline of the car will pass (it may be helpful to think of the dilating retinal circle), but this will not be what was earlier termed the car's "principal description", since the car as a whole is not centered.

Will the car's image be matched at all? It will be assumed that the template system makes some response, but a weaker one than it would make to the car's principal description. In fact, it is proposed that the car's off-center image is matched by a template which would fit a variety of quite different cars as seen from that fixation point. *The underlying assumption is that we indeed have templates for off-center objects, but they are less and less specific in their descriptions as the corresponding object is located more and more toward the periphery.* At the extreme, one has small objects, well out in the periphery, which are matched by a rudimentary blob template.

This idea is hard to accept. If true, how for example do we recognize our own car—as we surely do—on those occasions when we happen to see it off-center? The suggestion is that we recognize it from a combination of a generalized off-center template match and the context of the situation. If we were sitting in a darkened room and unexpectedly had the car briefly flashed on a screen in such a way that our point of fixation was the wheel, we might be quite unsure whether it was our own automobile, or even our own make. Our main response would be to try to move the eyes so as to achieve center and "see better".

From another point of view, the notion that we need to understand how we are able to recognize objects anywhere on the retina may well be a non-problem generated by confusion with the fact that we can see sharply anywhere in the field of view, if we move our eyes there. In reality, we do not see things at all sharply indirectly. Our ability to *recognize* indirectly—when not contextually supported—may be quite limited, as well.

Let us return to the cortical scanning of the scene with the car. At the beginning of the scan, we said there was a strong match with the wheel template. There follows a weaker match with an off-center generalized car template. Let us imagine the car is seen from a distance standing before the door of a center-entrance colonial house.

Although we are looking off-center at the car, we will still be essentially on-center with respect to the house, since the latter is much larger. Consequently the next event in the scan will be a strong "center-entrance colonial" match (assuming that style to be familiar). If the house has a small detached garage next to it, this will appear in the scan as a rather poorly differentiated blob. Should the eye now shift, say to the garage blob, the whole hierarchy will change, with the garage now finding strong match and the other objects becoming secondary.

In general, as the eye moves voluntarily (or even just wanders) about a scene, each important form, regardless of its size and the presence of other forms, will eventually be articulated and recognized. It is only necessary that at some point the form be more-or-less centered. The above four assumptions thus lead to a model in which a whole scene is analyzed in a simple, orderly, and psychologically plausible way. (In the light of these considerations, it is amusing to re-view the venerable "My Wife and My Mother-In-Law" picture reproduced here as Fig. 16; it is hard to see one woman while centered on the other.)

FIG. 16. "My Wife and My Mother-in-law" created by W. E. Hill and published in *Puck* in 1915.

For simplicity, the matching process has been described in terms of "templates", but it is clear that a more powerful and general recognizing and learning structure could be involved, perhaps along the lines of the so-called production systems (Davis & King, 1976). Briefly, a production system is a set of condition–action rules in which when a rule's condition is satisfied, it takes the specified action—subject perhaps to a competition with other currently satisfied rules.

In the present context, the condition part of a rule could be a template to match against a cortical pattern; the action part could be a specific eye movement (to see better, say). But also, the action part could be the generation of an internal message, and the condition part a template looking to match some such message. Thus a potential for chained and complex behavior would arise. Learning would be the setting up of new templates, perhaps by manipulating existing "good" ones, all in accordance

with experience, including a reward structure. Along this line, an adaptation of ideas such as those of Holland (1975) to a computer or robot version of the present model would seem especially interesting to pursue experimentally.

Other work

That the retino-cortical mapping has the mathematical form we have found has also been concluded by Weiman & Chaikin (1979), and by Schwartz (1977). Weiman & Chaikin principally propose a computer architecture for picture processing based on the complex logarithmic mapping, the formal properties of which they analyse extensively. They go on to present physiological and perceptual evidence that the retino-cortical mapping embodies the same function. Schwartz directly motivates the complex logarithm physiologically, presents evidence that it may be widely utilized in the nervous system, and suggests developmental implications. Both papers briefly discuss implications for perception and pattern recognition. While there are similarities between the present paper and the papers of these authors, it is perhaps indicative of the richness of this new field that the amount of overlap among all three is quite small.

Harmon (1960) built a pattern recognition machine which dealt with size (and orientation) variations of letters and geometric figures by means of a dilating circular scan. Although it did not have the logarithmic property, Harmon's machine is clearly a direct ancestor of the present concept.

John McCann showed me the Hubel & Wiesel data on field size and eccentricity at the time I was investigating the exponential MSU array (Fig. 2) in connection with peripheral vision and size-independence in pattern recognition. His idea that the discoveries of Hubel & Wiesel might relate to various size-independent phenomena, including the visibility of low-frequency gradients, caused me to take the MSU model much more seriously.

I am grateful for helpful conversations with Jerome Lettvin, Barbara Sakitt and Charles Stromeyer.

References

CRICK, F. H. C. (1979). Thinking about the brain. *Scientific American*, **241** (3), 219–232.

DAVIS, R. & KING, J. (1976). An overview of production systems. *Machine Intelligence*, **8**, 300–332. New York: Wiley.

HARMON, L. D. (1960). Line-drawing pattern recognizer. *Electronics*, 39–43 (2 September).

HELMHOLTZ, H. V. (1962). *Treatise on Physiological Optics, Part II*, pp. 39–42. New York: Dover Publications.

HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, Michigan: University of Michigan Press.

HUBEL, D. H. & WIESEL, T. N. (1974). Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology*, **158** (3), 295–305.

HUBEL, D. H. & WIESEL, T. N. (1979). Brain mechanisms of vision. *Scientific American*, **241** (3), 150–162.

LAND, E. H. (1977). The retinex theory of color vision. *Scientific American*, **237** (6), 108–128.

LETTVIN, J. Y. (1976). On seeing sidelong. *The Sciences*, 10–20 (July/August).

MCCANN, J. J. (1978). Visibility of gradients and low spatial frequency sinusoids: evidence for a distance constancy mechanism. *Photographic Science and Engineering*, **22** (2), 64–68.

MCCANN, J. J., SAVOY, R. L., HALL, J. A., JR & SCARPETTI, J. J. (1974). Visibility of continuous luminance gradients. *Vision Research*, **14**, 917–927.

PITTS, W. & MCCULLOCH, W. S. (1947). How we know universals. The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, **9**, 127–147.
SCHWARTZ, E. L. (1977). Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, **25**, 181–194.
WEIMAN, C. F. R. & CHAIKIN, G. (1979). Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, **11**, 197–226.
WIENER, N. (1948). *Cybernetics*, Ch. VI. New York: Wiley.