FRED S. ROBERTS AND PATRICK SUPPES

SOME PROBLEMS IN THE GEOMETRY OF VISUAL PERCEPTION

1. A GENERAL MODEL

In view of the long history of discussion about the nature of perception, both by philosophers and psychologists, it may seem foolhardy to propose to begin afresh with a general discussion. However, as we hope to show in this paper, many of the most fundamental problems about perception have not as yet been clearly settled and are just now receiving careful formulation. We believe that from a scientific standpoint the problems of perception are difficult, and we want to say at once that we do not propose to solve may of them here. Our purpose is mainly to set up a general model which we may use as a framework for discussion. In these general terms, we shall try to summarize a class of empirical observations about perception, organize some of the fundamental problems into sharply defined classes, and suggest one or two possible explanations.

1.1. *Physical Space vs. Perceptual Space*

We shall, as our title suggests, limit ourselves to a discussion of visual perception, and also to those visual phenomena involving perception of geometrical characteristics as opposed to such things as color, texture, and the like, although many of our remarks are more generally applicable. Our discussion for the most part will deal with binocular vision, although several of our explanations, notably those in terms of eye movements, will be monocular in nature. To begin with, we shall distinguish between *physical space* and *perceptual space*, the space from which we draw our 'conscious' perceptions. For the latter we shall also use the phrases *visual space* or *subjective visual space*. It seems sensible to take as physical space ordinary three-dimensional Euclidean space. This space we denote by $\mathbf{E}_3$, or simply by $\mathbf{E}$. (It is possible to argue about this choice of physical space.)

For perceptual space, we propose no a priori structure of a general nature. Indeed, we shall try to study how one might infer the geometric

Synthese **17** (1967) 173–201; © *D. Reidel Publishing Co., Dordrecht-Holland*

certain hyperbolic curves. Moreover, light arranged in physically straight lines is not always seen as straight by the observer. If a subject is asked to align two rows of lights in 'parallel straight lines' (parallel alleys), and then alternatively into 'lines with corresponding points equidistant' (distance alleys), the two resulting configurations are different, whereas in a Euclidean geometry they would be the same. If L and R denote the idealized centers of rotation of the left and right eyes respectively, then Figure 1 shows some of the configurations in the horizontal plane at eye level which are judged aligned. Figure 2 shows the parallel and distance alleys in the same plane.

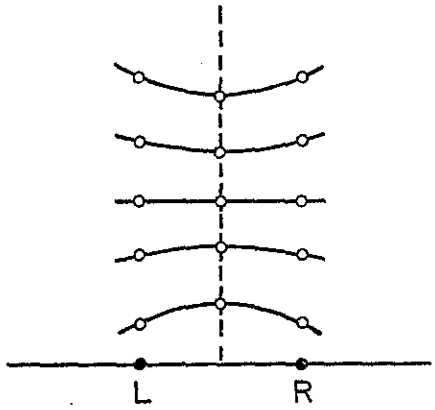

Fig. 1.

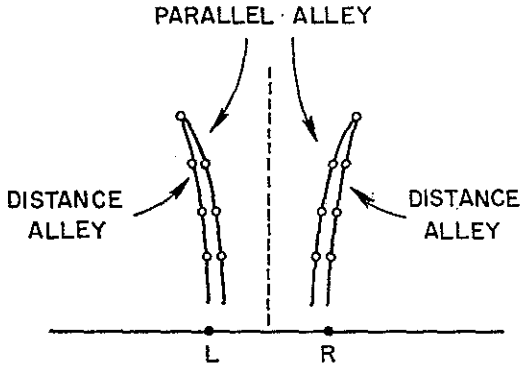

Fig. 2.

Helmholtz [10] obtained similar results by the use of after-images. Fixate at the center of a horizontal straight line at eye level in the frontal plane in physical space, and then shift your gaze rapidly to the center of a parallel line below it. The after-image of the first line does not coïncide with the second, but instead, the first line continues to appear straight

while the new line appears concave up. Conversely, the after-image coïncides with a physical hyperbola, concave down. Similar results hold if we move our gaze upwards or deal with vertical lines. Thus the physical curves 'seen as straight' appear to be hyperbolas that are convex toward the primary point of fixation.[2] (This includes the original horizontal line.) The curves seen as straight are at least qualitatively like those of Figure 3 below.

Such results indicate that the primitive visual geometry differs from the physical geometry. Presumably the role of learning is to help us overcome this difference. Thus in the case of straightness, for example, we have to learn to see physically straight lines as straight.

2.2. *The Eye-Movement Explanation*

There are two approaches to the study of primitive visual geometry. One is to try to describe precisely the properties of this geometry, and the other is to try to explain why our primitive geometry is as it is.

The latter type of explanations, presumably, are physiological. In our present kinematical situation, it is natural to try to use eye movement as a basis for understanding primitive visual space. Such an approach goes back to Helmholtz [9, 10], and it is of interest to follow his presentation, concentrating on straightness. The idealized eyeball may for our purposes be considered a rigid body which rotates about a fixed point O. If we fix our gaze at any point P in 'external space', then OP will be called the *visual axis*. In particular, the point of fixation A when we are looking straight ahead toward the horizon will be called the *primary point*. It is natural to assume that the position $E(P)$ of the eyeball when the fixation point is P is completely determined by the primary position $E(A)$ and the visual axis OP.[3] This fundamental law of eye movement is known as *Donders' Law*. Thus, under Donders' Law, no matter how much we move our eyes, if we return to looking at the same point in external space, the eyeball returns to the same position. It is implicitly assumed that the correspondence $P \to E(P)$ is not trivial (i.e., not a constant map) and also continuous. Finally, the discussion is really limited to points within a reasonable angular distance of the primary point. Donders' Law, which postulates the existence of a correspondence $P \to E(P)$, should be distinguished from any particular law of eye position, which specifies for each point P the corresponding position $E(P)$.

We would like to give, independent of any particular law of eye position, an eye-movement definition of our primitive perception of straightness. Using the notion of alignment as motivation, we say with Helmholtz that a curve C in physical space is 'seen as straight' (in the primitive sense) provided that as we move our fixation point along C (scan C), successive portions of the curve are imaged on exactly the same elements of the central portion of the retina.

Using this definition, Horace Lamb [12] proves the following striking theorem:

THEOREM: *Under Donders' Law, it is not possible for every physically straight line (segment) to be 'seen as straight'. (More precisely, under any particular law of eye position, the class of those physical curves seen as straight does not include all straight line segments.)*

This theorem is a strong argument for the view that our primitive visual geometry, for physiological reasons, cannot be Euclidean, and so learning must enter. Thus, it seems likely we cannot perceive Euclidean straightness at birth. Because we think this theorem is very important, because it does not seem to be a well-known result, and because Lamb's proof is not particularly rigorous, we sketch a proof in an appendix.

It might be objected that the Helmholtz definition of primitive straightness corresponds more to constant curvature than to straightness. However, we would argue that there does not seem to be any distinction on the primitive level between these two concepts. And, even if the objection is well taken, Lamb's theorem (with the words 'judged as aligned' replacing the words 'seen as straight') remains just as startling.

Accepting the Helmholtz definition, and given the negative result of Lamb, it becomes of interest to calculate exactly what curves in physical space are seen as straight. To do this, we need a specific law of eye position. It is sufficient to describe how to find position $E(P)$ from point P and the primary position $E(A)$. Since the same visual axis corresponds to many points of external space, we may limit ourselves to points P on the surface S of a sphere about O surrounding the eyeball. We shall call S the *spherical field.* In particular, we may assume A lies on S.

Probably the simplest law of eye position is *Listing's Law,* which says that $E(P)$ may be obtained from $E(A)$ by a rotation of the entire eyeball corresponding to the great circle arc AP on the spherical field.[4] (That is,

in moving our eyes in a haphazard path from fixation point A to fixation point P, the end result is 'as if' we just rotated the eye directly.)

Helmholtz proved the following:

THEOREM: *Under Listing's Law, those curves in external space which are seen as straight are exactly those corresponding under projection from the point O to arcs of circles on the spherical field which pass through the point B diametrically opposite to the primary point A.*

In particular, in the plane perpendicular to the line OA, all of the curves shown in Figure 3 are seen as straight. This result agrees with the experimental data mentioned.

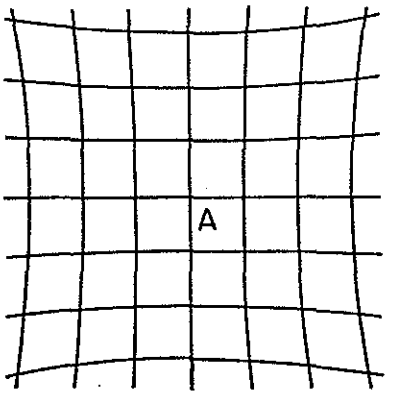

Fig. 3.

It should be remarked, as a final comment here, that our judgments of straightness in 'real-life' situations can be made without eye movement. We can even recognize straight lines which are flashed on a tachistoscope so fast that no eye movements can be made. How then can eye movements be used to account for the perception of straightness if the perception of straightness can be accomplished without eye movements? The answer here is that it is only our learned concept of straightness which can be perceived without scanning. Before learning, we require eye movements to perceive alignment (cf. Hebb [8]). These observations indicate then, not that our eye-movement definition of primitive straightness is misguided, but rather that learning plays a crucial role. For it seems that we can see a certain familiar configuration on the retina and immediately infer that it is straight without scanning at all (cf. Platt [15]). We describe a specific mechanism for such inferences in Section 3.3.

2.3. *Recovering the Primitive Visual Geometry*

The other approach to primitive visual space is to try to recover its geometry from certain observables. We are interested in studying what structures are compatible with our primitive visual perceptions, what relations are meaningful in our primitive visual space, etc. Hence we are using the word 'geometry' in a very general sense. The problems involved here are what observables to choose and what properties to study and derive.

For example, one whole collection of observables are our judgments of comparative distance, alignment or betweenness, parallelism, etc. We shall limit our discussion to these with the remark that choice of appropriate observables for the study of primitive visual space is very much an open question. These particular concepts all make sense in an abstract metric space. Blank [3], following Luneburg [13], has investigated to what extent primitive visual space is a metric space. He starts with observed relations Q^* and B^*, the comparative distance and betweenness relations; i.e., Q^* consists of all quadruples (x, y, u, v) of points of visual space so that the distance between x and y is observed to be smaller than the distance between u and v; and B^* consists of all triples (x, y, z) so that x, y, and z are observed to lie on a line, with y between x and z. It should be noted that Q^* and B^* do not necessarily agree with the corresponding Euclidean relations.

Given a metric d, we may speak of its comparative distance and betweenness relations, Q_d and B_d, defined respectively as $\{(x, y, u, v): d(x, y) < d(u, v)\}$ and $\{(x, y, z): d(x, y) + d(y, z) = d(x, z)\}$.

Blank proves a representation theorem of the following form:

THEOREM: *Under a set of axioms A on the relations Q^* and B^*, there exists a metric d on primitive visual space such that $Q^* = Q_d$ and $B^* = B_d$. Moreover, such a metric is unique up to a similarity transformation.*

We quote this theorem, but without a detailed list of the axioms, because we feel that it is an example of the *type* of consideration involved here. Namely, it is an attempt to recover the geometry of visual space from certain observed relations. We are not asserting that a subject consciously makes judgments of numerical distance. The numerical measure is, in the words of Blank, "something superimposed upon his visual experience". It might, on the other hand, be a reasonable approach merely to describe the primitive visual space by describing the various

observed relations, without trying to give numerical representations.

The specific Blank axiomatization has several serious drawbacks of which Blank is aware. For example, the observed relation $xy < uv$ defined by $Q^*(x, y, u, v)$ is probably not transitive on 'pairs' as it must be if Q^* comes from a metric. Such difficulties stem in large part from what happens near the threshold of discrimination. This is typical, then, of the difficulty in studying primitive visual geometry: local geometric and topological properties are obscured.

Beals, Krantz, and Tversky [2] list a set of axioms based on the relation Q^* alone, which may be compared with Blank's, although the axioms do not refer specifically to visual space. Their representation theorem is also essentially the same. These axioms encounter the same difficulty as do Blank's.

To go a step further, a good description of primitive visual space might be some sort of coordinatization, together with a set of functions relating the physical and psychological coordinates. Blank obtains a coordinatization by adding several powerful axioms, allowing him to prove:

THEOREM: *Primitive visual space is a Riemannian space of constant negative curvature, i.e., a hyperbolic space.*

The result of this theorem is actually the *starting point* of the Luneburg theory. Using the extremely specific geometry implied by the theorem, Blank and Luneburg coordinatize visual space, write down an explicit metric in terms of these coordinates, and investigate experimentally the relation between psychological and physical coordinates. We feel, once again, that these results are more significant for their approach to the study of primitive visual space than for their detailed accuracy as a description of it, although it should be added in all fairness that the hyperbolic geometry explains several of the experimental results described earlier.

As a final comment, it should be noted that one of the advantages of both the Blank and the Beals, Krantz and Tversky axiomatizations discussed above is that they allow for a simple algorithm to recover the metric from the observables. This is not true of the additional Blank axioms required for the representation as a hyperbolic space, and hence of his metric based on the explicit psychological coordinates, without knowing the non-explicit relation between physical and psychological coordinates. We feel in general that such representations *should* be constructive in nature, allowing us to approximate the primitive visual

geometry by simple algorithmic procedures from the observed data.

3. LEARNING

We would like to turn next to an investigation of how our primitive visual space becomes modified through learning. We assume that a primary role of learning is to overcome the difference between primitive visual space and physical space. (It is immediately clear that it is not possible to completely overcome this difference. To a large extent, for example, we cannot avoid restrictions on our visual acuity and hence the existence of thresholds in our visual space.)

It should be noted that perception after learning involves both learned and primitive factors. By learned perceptual behavior, we mean perceptual behavior after learning, thus including the invariant primitive factors. An investigation of learning in perception should, we feel, be divided into two problems. First is the problem of giving a precise description or definition of the learned perceptual behavior. And second is that of suggesting how we develop a mechanism for exhibiting this behavior. We have been disappointed not to be able to find many mathematically oriented papers that discuss these problems.

3.1. *Perception of Constancy*

We shall restrict ourselves to discussion of the so-called 'perceptual constancies'. The fact of perceptual invariance has been commented upon by philosophers and psychologists for a long time, and for good reason. The phenomena of size constancy and shape constancy are basic to our very ability to move around in the world, to relate to our environment, and so forth. It is almost inconceivable to imagine what would happen if we could not identify an object seen in different positions or orientations, under different conditions, and transformed in various ways. It seems correct to claim that to a very large extent, these constancies are learned, but the precise sense in which they are learned needs to be formulated with care. Proceeding with our twofold approach, we would like to start by giving a precise definition of 'constancy' and 'perception of constancy'. Our discussion should be viewed not so much as a definite proposal but rather as a hopefully fruitful framework for discussion.

Any description of the phenomenon of constancy leads to the idea that

there is a certain group[5] G of transformations of the physical space under which figural identity is preserved.[6] Namely, if $\mathbf{S}$ and $\mathbf{S}'$ are subsets of physical space $\mathbf{E}$, then $\mathbf{S}$ and $\mathbf{S}'$ are identified if and only if there is a transformation t in G so that $\mathbf{S}=t|\mathbf{S}'$. (The notation $t|\mathbf{S}'$ means the restriction of t to $\mathbf{S}'$. By abuse of notation, we shall hereafter write $\mathbf{S}=t(\mathbf{S}')$). We might prefer to think of G as a group of transformations on a selected collection F of subsets of $\mathbf{E}$, called the 'relevant configurations'. The idea of 'squareness' might then be looked at in the following platonic way: we have in mind an 'ideal' square in a standard position in physical space. Given a configuration in physical space, we identify it as a square if and only if it can by a series of allowable transformations be superimposed on or made to coïncide with this ideal square. Similar remarks might be made for 'right angle', 'circularity', etc.

For every group of transformations we have a different type of constancy. 'Size constancy', for example, might correspond to taking the group of congruence (or distance-preserving) transformations of Euclidean three-space.[7] A certain type of shape constancy arises by taking G to be the group of rotations of E. We view it as an empirical problem to identify the relevant groups G for different types of significant constancies, and prefer to speak in this generality here.

The group G determines an equivalence relation $\equiv_G$ on the subsets of $\mathbf{E}$, defined by: $\mathbf{S} \equiv_G \mathbf{S}'$ if and only if there is a transformation t in G so that $\mathbf{S}=t(\mathbf{S}')$. We think of this equivalence relation as the constancy and we shall refer to $\equiv_G$ as the 'constancy of type G'. Then our 'perception of constancy' is the identification of two figures equivalent with respect to G.

Pitts and McCulloch [14] suggest a mechanism for perception of constancy which is of some interest, though we choose not to pursue it in detail here. Generalizing their ideas, we might suppose, as in Section 2, that the image eventually becomes stabilized. Then, it is reasonable to assume that corresponding to each physical stimulus S there is a sensory image $\varphi(S)$. This might be thought of as a retinal image or a firing pattern on the neurons in 'area 17' of the cortex. If Φ denotes the class of all $\varphi(S)$, then, we suggest, there is induced on the space Φ an equivalence relation $\approx_G$ corresponding to $\equiv_G$. And the constancy in question is 'computed in the brain' by means of a functional F with the property that $F(\varphi)=F(\varphi')$ if and only if $\varphi \approx_G \varphi'$. Pitts and McCulloch give a physiological interpretation of F.

3.2. *Elementary Properties and Concepts*

Our definition of constancy is actually independent of whether the perception of constancy is a learned or innate process. But it is clear our constancies are to a large part learned, and so we would like to turn to an explanation of how such learning might come about.

We begin by distinguishing two types of equivalence classes under a given constancy. One is the class of 'elementary properties' and the other is the class of 'concepts'. We shall propose that a constancy is acquired through the learning of numerous elementary properties and concepts which are invariant under it.

We cannot be extremely precise here, but what we have in mind for the elementary properties are such things as straightness, parallelism, perpendicularity, and roundness. On the other hand, a concept may be thought of as a collection of elementary properties, in a sense to be formalized in Section 3.4. For example, 'squareness', 'consists' of the elementary properties 'two pairs of parallel lines', 'four right angles', 'four equal line segments', etc.

We feel that a precise determination of the elementary properties will have to be to a large part experimental in nature. (It should be noted, by the way, that a determination of the elementary properties depends on the constancy. Thus, if the underlying constancy group G is the group of all rotations, then any two Euclidean straight lines are equivalent under G and the property of being a straight line is a candidate for an elementary property – it corresponds to one equivalence class. Similarly, if G consists of just parallel displacements of $\mathbf{E}$ into itself, then the property of being a horizontal straight line is a candidate for an elementary property).

Experimental determination of elementary properties should center around what basic properties we use to organize our perceptions. Any study of visual perception must come to grips with the vast information-processing problems involved in organizing our perceptions. How do we pass from a mass of perceptual input to an organized conscious perception? How do we arrange our percepts into meaningful parts? How do we select what aspects of the stimulus are consciously perceived?

Our organization of figures seems to be greatly influenced by our learning, although there are probably some innate or primitive factors involved here too. We propose to identify the elementary properties with

the meaningful units into which complex percepts are arranged. There is a particularly fruitful and relatively new source of data in the light of which we can be a little more specific in our ideas here. This source is the collection of experiments in which even involuntary eye movements are eliminated. By various means, the image is stabilized on the retina.[8] If this happens, the conscious perception soon fades. After a time, however, the image alternately reappears and fades out in various 'meaningful' units. Pritchard [16] describes some of these phenomena in detail. From a geometric standpoint, some of the organizing factors appear to be straightness, parallelism, and similarity, all in the Euclidean sense. That is, straight lines, parallel lines and planes, and similar figures, usually appear and disappear together. Also, contiguity, symmetry, convexity, boundary or 'closedness', and angle, among other factors, appear to play a role. Adjacent curves and closed figures, for example, appear as wholes. These are the types of concepts Gestaltists often use.

To give some examples, given an array consisting of rows of squares, the parts reappearing together are usually horizontal, vertical, and diagonal rows. A Necker cube breaks up into single lines, parallel lines, a pair of parallel planes or a pair of adjacent lines or planes. Finally, given a triangle and a circle, either these appear singly or alternatively adjacent boundaries appear together. Summarizing, in some sense 'simpler' figures act as units. (Simplicity also seems to play a role in length of reappearance time.)

These data indicate that our elementary properties should be divided into two classes, primitive and learned. Those such as Euclidean straightness, parallelism and the like are learned, while such factors as contiguity, boundary and closedness are probably primitive. A second observation here, and one that we made earlier, is that our perception of the elementary properties can be accomplished even without eye movements. Experiments in which the stimulus is flashed on a tachistoscope bear out this observation. We may recognize Euclidean (or learned) straightness without scanning. This is in direct opposition to our perception of primitive straightness, which, as we saw in Section 2, is crucially tied to eye movement. Any theory of learning in perception will have to account for these facts.

The observation that the elementary properties correspond in some sense to the 'simpler' figures leads to one theoretical attempt at defining

the elementary properties. Attneave [1] suggests that simplicity has something to do with regularity or redundancy, in particular "predictability of the whole from a part". Thus, for example, a Euclidean straight line is so simple because we only need to know two points on it to know all others. And, two parallel lines are so simple because two points on one line and one on the second determine the pair. If a subject is shown a closed curve and is asked to represent it from memory, by, say, ten points, then the points he chooses are those points where regularity is interrupted: corners, sudden bends, and changes in convexity. The reader is referred to the paper by Attneave and to Hochberg [11] for a more detailed discussion.

An alternative theoretical attempt at defining the elementary properties is considerably different from this one. It involves a study of the neural configurations or firing patterns in the cortex, and aims to describe the elementary properties as those corresponding to neural patterns satisfying certain criteria. This neural theory at the same time provides a mechanism for perception of these elementary properties.

Zeeman [18] provides a model of the brain which is particularly useful for the points we shall make. His model of the brain is a triple (C, γ, p), where C is the set of nerve cells in the cortex; γ is the binary relation on C consisting of all pairs of cells (a, b) such that a can fire b; and p is a function $p\colon \gamma \to [0, 1]$ which represents the strength of the connection.[9] We may alternatively think of $p(a, b)$ as the probability that if a fires, then b will fire.

The cortex C consists of three distinct classes of cells: S, T, and R. The cells in S receive sensory inputs; those in T may be fired by other cells in C; and those in R are the 'self-firers'. S and T are disjoint, for otherwise we would confuse sense data and imaginings. A 'thought' or 'perceptual image' is then a 'firing pattern' on C, or more precisely a function $t\colon C \to [0, 1]$, where $t(c)$ represents the 'rate of firing' of $c \in C$.

Zeeman introduces a measure of the 'sharpness of an image', and suggests that straight lines, parallel lines, boundaries, and so forth, produce sharp images. His definition of sharpness $s(t)$ of an image t is:

$$s(t) = \frac{\sum_{c \in C} (t(c))^2}{\sum_{c \in C} t(c)}.$$

Note that $s(t)$ is between 0 and 1, and a sharp image occurs if the cortex

is divided into two parts, one firing rapidly and the other slowly. This, as Zeeman says, is a crude measure. But it is, we feel, on the right track, and at least exemplifies how the notion of elementary property might be explicated by the equation 'elementary property = sharp image'.

3.3. *Learning Elementary Properties*

We propose as suggested above that a constancy is in part innate and in part acquired through the learning of numerous elementary properties and concepts that are invariant under it. In our idealized model, a concept cannot be learned until various relevant elementary properties are learned, and so we shall divide our discussion of learning into two parts, first dealing with elementary properties and then with concepts in Section 3.4.

A mechanism for perception of elementary properties is easy to describe: each elementary figure (or instance of an elementary property, e.g., a straight line) corresponds to a particular firing pattern in the sensory input cortex S (to use the notation of Zeeman's model). *Suppose* that after learning, each steady sensory input coming from a fixed elementary figure always gives rise to the same firing pattern or image t not only on S but after stabilization, on all of C. Then, each elementary property corresponds to one or more such images.

To discuss the mechanism in more detail, let us concentrate on straightness. Thus, for example, a horizontal line at a particular location in **E** will produce a certain familiar image which becomes conditioned to the phrase 'straight line'. Now, although there are infinitely many different horizontal straight lines in physical space, physiological data indicate that there is a relatively small number of corresponding excitation patterns in S, only about 10 or 15 (Hebb [8]). If corresponding to each such pattern there is after stabilization only one image, then, cortically we can distinguish only a small number of horizontal lines. Similarly, we probably can distinguish only a small number of different slopes of lines, and for each slope only a small number of lines of that slope. Thus, we must condition to the words 'straight line' only a small number of firing patterns. In this way, we have a mechanism for perceiving straight lines, or alternatively horizontal lines, and similarly other elementary figures.

The major problem we face with such a model is in justifying the supposition that a certain steady sense input, and hence, a fixed firing pattern on S, always gives rise, under the influence of learning, to the

same image t on C. This is especially doubtful considering all the randomness built into our model. The first thing to do is to modify our demands. It is certainly unreasonable to require that a given steady sensory input coming from an elementary figure always give rise to the exact same image. Instead, we would be happy to have, using the terminology of Zeeman [18], a *tolerance relation* on the class of all perceptual images, so that such an input always gives rise to two images that are 'close' or within tolerance.

Hebb [8] suggests a method by which this can come about, and this method is easily understood in the framework of Zeeman's model of the brain. Recall that for Zeeman the brain is a triple (C, γ, p), where C and γ are essentially unchanging physiological constants. Thus, our learning must involve change of p. This change occurs, both Zeeman and Hebb suggest, through the process of *facilitation*: if cell a fires cell b, then in the future it is slightly easier for a to fire b.[10] The physiological process of facilitation has been observed, and appears to be basically chemical in nature. It is easy enough to suggest a neural model for facilitation which might be compared with physiological data.

Suppose for simplicity that firing patterns over a course of time are governed by a distribution so that for every pair of neurons a and b at each unit time t_n, the probability that a and b both fire is u, the probability that only one of the two fires is v, and the probability that neither fires is $1-u-v$. Suppose that $p_n(a, b)=p_n$ represents the 'strength' of connection at time t_n, and that

$$r_n(a, b)=r_n=\begin{cases} 1 \text{ if } a \text{ and } b \text{ both fire between time } t_n \text{ and time } t_{n+1}, \\ 0 \text{ if only one fires in this time interval,} \\ p_n \text{ if neither fires in this time interval.} \end{cases}$$

Finally, if θ is a constant between 0 and 1, one possible learning procedure is to modify p_n according to the equation

$$p_{n+1} = (1 - \theta)\, p_n + \theta r_n .$$

A simple computation shows that $\lim_{n \to \infty} E(p_n)=u/(u+v)$. If now v is small, i.e., if, frequently, when one of these neurons fires, then so does the other one; and if u is not too small, then $u/(u+v)\approx 1$. Hence, the expected value of the strength of connection between a and b, or of the probability that if a fires then b does, approaches one as time passes. This is inde-

pendent of the original strength of connection. (More complicated models may be developed for more complicated distributions.)

Hebb's suggestion may now be formulated in this framework. There are in the cortex certain neurons arranged in loops or cycles, i.e., groups $a_1, a_2, \ldots, a_n$ such that $(a_1, a_2) \in \gamma$, $(a_2, a_3) \in \gamma$, $\ldots$, $(a_{n-1}, a_n) \in \gamma$, $(a_n, a_1) \in \gamma$. Let $A_1, A_2 \ldots, A_n$ be sensory input neurons so that A_i can fire a_i and so that the connections A_i to a_i are strong in the sense that $p(A_i, a_i)$ is close to 1. If we indicate the relation x *can fire* y by an arrow, the picture is as shown in Figure 4.

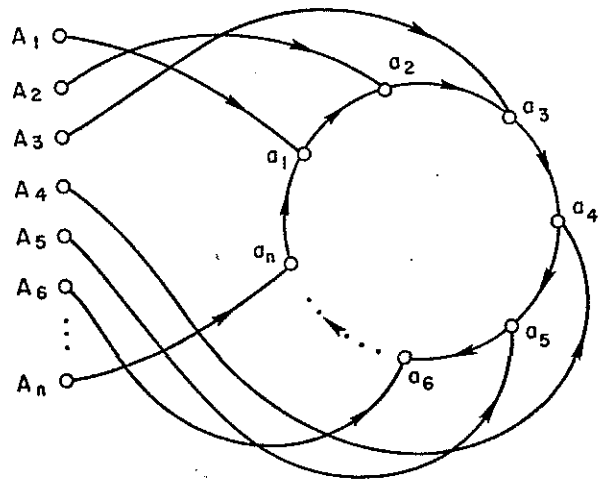

Fig. 4.

Suppose now that $A_1, A_2, \ldots, A_n$ correspond to the sensory input in the cortex when, say, a given straight line (or other elementary figure) is perceived. Then, $A_1, \ldots, A_n$ are often stimulated, and hence fired, together; or more importantly, usually when one fires, then all fire. It follows that in general, since the connections $A_i \to a_i$ are strong, that some time later the neurons $a_1, a_2, \ldots, a_n$ will all fire. Thus, the elementary figure or in particular the straight line in question, after learning, almost always gives rise to an image within tolerance of the image t such that $t(a_i) = 1$, for all $i = 1, 2, \ldots, n$ and $t \equiv 0$ otherwise. Such images then become conditioned to the phrase 'straight line'.

To test this model, let us return to the observation that elementary properties, once learned, can be perceived without eye movement or scanning. Does the model have sufficient structure to account for this phenomenon? The answer is affirmative. Suppose the elementary figure gives rise to firing of the neurons $A_1, \ldots, A_n$ as above. Now, because

$a_1, \ldots, a_n$ often fire together, the connections $p(a_i, a_{i+1})$ become strong through facilitation. Thus, after learning, if the single sensory neuron A_i or any subclass of the class $A_1, \ldots, A_n$ is fired by a sensory input, it follows that all the neurons $a_1, \ldots, a_n$ are likely to fire. Hence, without waiting for more 'input', we immediately 'reach the same conclusion' or 'have the same image' t, as if $A_1, \ldots, A_n$ all had fired. Without any scanning, we instantaneously infer 'straight line'.

3.4. *Learning Concepts*

We turn now to the problem of providing mechanisms that will account for the learning of concepts that are the invariants of a given group of transformations. This problem is formidable, and we do not pretend to offer a detailed theory here. We would like to sketch at least one approach that seems promising enough to be outlined. This approach builds on the ideas of stimulus-response learning theories, particularly stimulus-sampling theory. The psychological processes of stimulus sampling and conditioning are central to the theory, but some additional aspects are needed to account for the phenomena at hand.

Let us begin by considering a concrete problem – that of recognizing regular polygons under rotation. A square is visually a square, no matter what angle its base forms with the horizontal. How do we recognize it in different positions of orientation? We may think of the stimulus elements here as the elementary properties of squares whose bases are horizontal. Typical properties are these: four sides (f), a horizontal segment (h), a vertical segment (v), parallel sides (p), interior angles that are right angles (r), no curves but only segments ($\bar{c}$), all segments of equal length (l), intersections only at end points of segments (i), and so forth. We are not trying to give an exhaustive list. We may describe the set enumerated as $\{f, h, v, p, r, \bar{c}, l, i\}$. It is perhaps important to point out that the step from the firing of individual neurons or the activation of individual receptor cells to recognition of these elementary properties is a large one conceptually. We are assuming only that it has been made already, hopefully along the lines suggested in Section 3.3.

Presented with a square whose base is horizontal, the individual can sample all the elementary properties listed and condition them to the concept of a square, or, to be more concrete, to the word *square*. He is now asked to pick out squares from a number of plane figures presented

to him. For simplicity, let us assume that the individual stores an ordered list of elementary properties, all of which are highly salient. He thus converts the unordered set $\{f, h, v, p, r, \bar{c}, l, i\}$ into the ordered set $(f, h, v, p, r, \bar{c}, l, i)$. Presented with a figure he then checks off the appropriate presence or absence of an elementary property. Asked if a triangle is a square, he can say 'no' immediately because the triangle has property $\bar{f}$, the negation of f.

Suppose the subject in our hypothetical experiment is now presented with a square whose base is at a 45° angle to the horizontal. Our subject will respond that this figure is not a square, because it has elementary properties $\bar{h}$ and $\bar{v}$. He is corrected and told that it is a square. At this point learning and conditioning enter. With probability e he eliminates each elementary property that has varied, i.e., that is not an invariant. To eliminate here means, in the formal representation, to replace h or v by o, to indicate neutrality, not to replace h by $\bar{h}$, of course, or v by $\bar{v}$. Thus with probability e^2, h and v are eliminated, with probability $e(1-e)$ only h is, again with probability $e(1-e)$, only v is, and with probability $(1-e)^2$ neither is. More realistically we would probably want to introduce a different elimination parameter e for each elementary property with the intention that e varies directly with the saliency of the property. Once both h and v are eliminated, the tilted square will be recognized as a square.

What we have just described is the approach to using positive instances of the concept to obtain information about the concept. If figures are presented on a randomized schedule and the subject has no choice of what is the next figure, then little information can be obtained from negative instances, at least little information at the elementary level. Indeed, if only certain elementary inference mechanisms are assumed it is easy to show that under randomized presentation schedules negative instances add no information whatsoever. It is true that more complicated hypotheses can be settled by inspection of negative instances, but in the present account we shall limit ourselves to the use of elementary steps of inference to pass from elementary properties to concepts and so learning will apply only to positive instances.

We now make the theoretical approach just sketched more precise and detailed. We let the set S be the set of elementary properties or stimulus elements. There are various ways of talking about these elementary properties. Here we shall simply treat the properties extensionally, so

that each elementary property s_i is a finite partition of F, where F is the set of geometric figures whose invariance under the group G of transformations is being learned. We assume that F is closed under G, i.e., if $f \in F$ and $T \in G$, then $Tf \in F$, where Tf is the figure that results from applying T to f. We also assume that the set S is finite, and enumerate its elements in the order $s_1, \ldots, s_N$. Thus, the basic situation facing an organism is described in the present theory by a triple (S, F, G). In the psychological literature it is also common to call what we have termed properties, *dimensions,* and then to talk about the values of the dimensions, corresponding to the elements in the partition. For property s_i we shall use the notation s_{ij} to refer to the *jth* value (*jth* partition element) of the dimension or property. The simplest case would be a two-element partition, e.g., s_{11} = at least one line segment (in the figure), $s_{12} = \bar{s}_{11}$ = no line segment. A slightly more complicated example would be: s_{11} = exactly one segment, s_{12} = exactly two segments, s_{13} = more than two segments, s_{14} = no segment.

Following another terminology, which is increasingly used in the literature of concept learning, we may say that a concept is then formally represented by a *template* which is an N-tuple $(t_1, \ldots, t_N)$ such that each t_i is some s_{ij} or the whole set F. The meaning of the last alternative is that no restriction is placed with respect to property s_i on the figures exemplifying the concept.

It is clear that the number of concepts that can be defined in terms of S is large, even for an S of modest size. For example, if we restrict ourselves to five elementary properties each of which has five values, then the number of extensionally different concepts, given that each value of each property is exemplified in the set F, is 5^6. (The exponent is 6 rather than 5 because we include the possible value F.)

A concept C immediately defines a two-element partition of F, namely, the partition of F into those figures that possess the property defined by the concept and those that do not. With a slight abuse of language we shall say that figures are elements or members of C, even though C is an N-tuple and not a subset of F. The meaning is clear: for every f in F, if $C = (t_1, \ldots, t_N)$, then $f \in C$ if and only if for $1 \leq i \leq N, f \in t_i$.

Following standard lines, we shall say that a concept C is *invariant* with respect to a group G if for every figure f in F and every transformation T in G, $f \in C$ if and only if $Tf \in C$. We shall restrict ourselves for the following discussion to concepts invariant under a given G.

For the present context, assumptions about sampling or observing elementary properties will be highly simplified by assuming that all elementary properties are observed or sampled whenever a figure is presented. Almost certainly this assumption will not be satisfied in actual experiments, but the use of it here affects only slightly the central ideas. Each figure f in F presented possesses some degree or value s_{ij} of each elementary property s_i. Thus each figure f may be described by an ordered N-tuple $(u_1, \ldots, u_N)$ where for $1 \leq i \leq N$, u_i is always s_{ij} for some j. This N-tuple will be called the *elementary pattern* of the figure f. Again, for the sake of conceptual simplicity, we are assuming that no perceptual errors occur in deciding whether a figure has property s_{ij}.

We now assume that learning a concept C involves learning the appropriate template $T=(t_1, \ldots, t_N)$. This is accomplished gradually. On each trial n, the subject has a template $T_n=(t_{1,n}, \ldots, t_{N,n})$ associated with the concept C. Through conditioning, T_n is modified until it eventually becomes T.

Prior to stating any specific axioms we need to define the notion of the elementary pattern of a figure matching the template of a concept.

DEFINITION: *Let $U=(u_1, \ldots, u_N)$ be the elementary pattern of figure f in F, and let $T_n=(t_{1,n}, \ldots, t_{N,n})$ be the template of concept C on trial n. Then U matches T_n if and only if for every $1 \leq i \leq n$, $u_i \subseteq t_{i,n}$.*

Following familiar treatments of stimulus-sampling theory, we now state axioms divided into the three categories of sampling axioms, response axioms, and conditioning axioms. (For such a treatment of stimulus-sampling theory see Suppes and Atkinson [17].)

Sampling Axiom

S1. *On every trial the elementary pattern $(u_1, \ldots, u_N)$ of the presented figure f in F is completely sampled.*

Response Axioms

R1. *Until a positive instance of the concept is presented, the subject does not have a template of the concept, and each response is made in terms of a fixed guessing probability ρ.*

R2. *On trial n, if the subject has a template T_n for C, then figure f in F is classified as an instance of C if the elementary pattern of f matches the template, otherwise not.*

Conditioning Axioms

C1. *On every trial, the subject has at most one template for the concept.*

C2. *The initial template is the elementary pattern of the first positive instance of the concept.*

C3. *Let* $(u_1, \ldots, u_N)$ *be the elementary pattern of the figure* f *presented on trial* n.

(i) *If* $f \in C$, *then for each* i *such that* $u_i \subseteq t_{i,n}$, $t_{i,n+1}$ *becomes* F *on trial* $n+1$ *with probability* e, *and remains* $t_{i,n}$ *with probability* $1-e$.

(ii) *If* $f \in C$ *and* $u_i \subseteq t_{i,n}$, *then* $t_{i,n+1}$ *remains* $t_{i,n}$.

(iii) *If* $f \notin C$, *then* $t_{i,n+1} = t_{i,n}$ *for all* $1 \leq i \leq N$.

It is easy to prove that if the probability e of conditioning is not zero, then for a wide range of presentation schedules of figures, a concept C definable in terms of elements of S and invariant under a group G will be learned with probability 1. We hasten to add that somewhat more complicated axioms are needed to give an empirically realistic account of classification responses in the early trials of learning. The most unrealistic aspect of the present axioms is that they predict only one kind of error after the first positive instance of the concept appears, namely, misclassification of positive instances, with classification of negative instances always being correct. Secondly, it is unrealistic to build a completely specific template on the basis of the first positive instance of the concept, as called for in Axiom C2. A closely related, but more complicated scheme calls for the probabilistic construction of a complete template, in terms of noticing or sampling all the elementary properties, over a number of trials. The modifications required also naturally lead to the prediction of errors in classifying negative instances.

Other shortcomings of the theory formulated in the axioms just given are not hard to find, but we shall not pursue the analysis further here. All that we have intended is to illustrate the kind of theory that seems promising for giving an account of learning perceptual invariants. More detailed theoretical developments and the consideration of quantitative empirical data we leave to another time and place.

APPENDIX: PROOF OF LAMB'S THEOREM

We shall attempt to define the terminology, explicitly state the assumptions, and then sketch the arguments needed for a proof of Lamb's Theorem, thus explicating his intuitive presentation.

1. Let us begin by assuming a fixed coordinate system in Euclidean 3-space and letting O denote the origin. Let the eyeball (in primary position, say) be denoted by E. As in Section 2.2 we make the following assumption:
Assumption 1: E is a rigid body centered at O.

Let T be a sphere centered at O and containing E. We describe the movement of E by the movement of T. We shall mean by 'external space' that collection of points in our fixed 3-space which is external to T and within a fixed angular distance of the primary point A. We shall in general denote points of external space by upper-case letters and points of T by lower-case letters, except that the point O lies in both.

For each point P of external space, there is by Donders' Law a corresponding position $T(P)$ of T. Let us identify T with $T(A)$ and then think of $T(P)$ as the image of T in our fixed 3-space which results from mapping each point b in T into the point $b(P)$ occupied by the point b when T is in the position $T(P)$. In this way we may describe the movements of the sphere T.

2. It is probably reasonable to assume that there is an idealized center of the retina, say a point $c \in E \subseteq T$. Moreover, there is a point $f \in E \subseteq T$ which is the focal point of the visual system. To carefully explicate Lamb's argument, we make the following additional assumptions:

Assumption 2: If P is the point of fixation in external space, then the point P is imaged in the point c of the retina.

Assumption 3: The retinal image of a point Q lies on the line determined by f and Q.

Assumption 4: The points c, f and O are collinear.

Note that by Assumptions 2 and 3, the points c, f and P are collinear when P is the point of fixation. Hence, Assumption 4 is equivalent to the assumption that the point f always lies on (the extension of) the visual axis.

For convenience, we introduce the point $a \in T$ which is the point of intersection of the surface of the sphere T with the visual axis OA in this primary position (see Figure 5). Summarizing all of the above remarks, we have

LEMMA 1: *If the point P in external space is the point of fixation, then the points c, f, a, O and P are collinear. To use alternative notation, the points $c(P)$, $f(P)$, $a(P)$, O and P are always collinear* (see Figure 6).

Remark: To determine new position $T(P)$ from point P and primary position $T(A)$, note that the new position of axis Oa is completely determined: it must coïncide with the line OP. Thus, by Assumption 1, the only new degree of freedom is a rotation about the axis OP. This will be crucial in the proof of Lamb's Theorem.

LEMMA 2: *If the point of fixation P lies on the line segment MN in external*

space, then the image on the retina of line segment MN in this situation lies in the plane O, M, N.

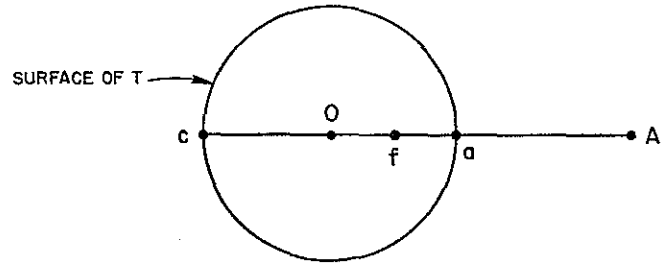

Fig. 5.

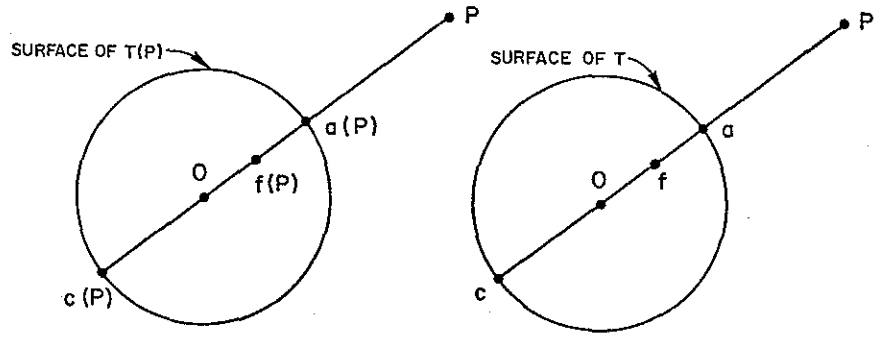

Fig. 6.

Proof: Since $f(P)$, O and P are collinear by Lemma 1, $f(P)$ lies in the plane O, M, N. Thus, for every point R on MN, the line determined by R and $f(P)$ also lies in this plane. Finally, by Assumption 3, so does the image of R (see Figure 7). Q.E.D.

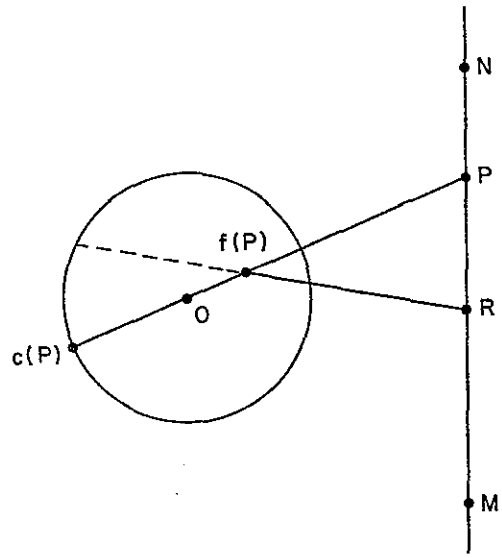

Fig. 7.

3. Let us introduce, as in Section 2.2, a spherical surface S centered at O and completely surrounding T. As before, each point of external space corre-

sponds to one point of S under projection from O. A given line segment MN in external space corresponds to a great circle arc $M'N'$ on the spherical surface S (see Figure 8). The crucial result in the proof of Lamb's theorem may now

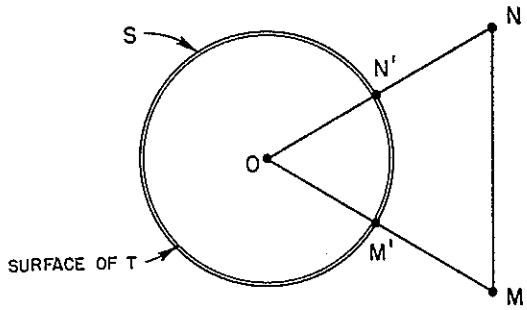

Fig. 8.

be formulated as follows:

LEMMA 3: *If a straight line segment MN in external space is seen as straight, then as the fixation point moves from M to N along MN, the eye rotates along the corresponding great circle arc M'N'. (More precisely, if R and R' are any two points on MN then the position T(R') is reached from the position T(R) by a rotation τ about an axis perpendicular to plane O, M, N, and through an angle ROR' as shown in* Figure 9).

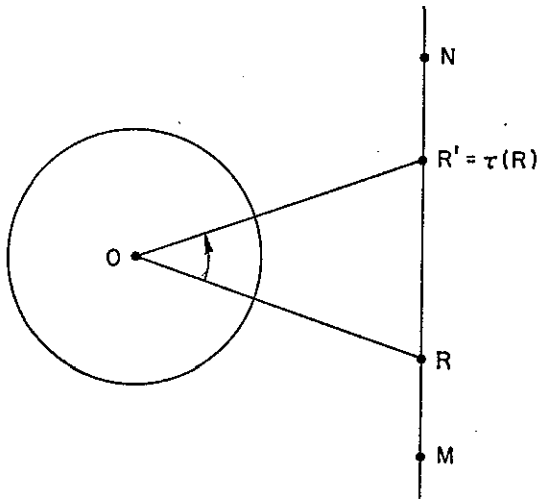

Fig. 9.

It is this lemma which Lamb sloughs over in one line and which is at the very heart of his result.

Proof: Suppose R and R' are two points on the line segment MN. As the point of fixation moves from R to R' along MN, Lemma 1 implies that the axis Oa rotates in the plane O, M, N as shown in Figure 10. Let B be the point of intersection of the surface of sphere $T(R)$ with a line in the plane O, M, N perpendicular to line $Oa(R)$ (see Figure 11). Suppose $B = b(R)$ for some point

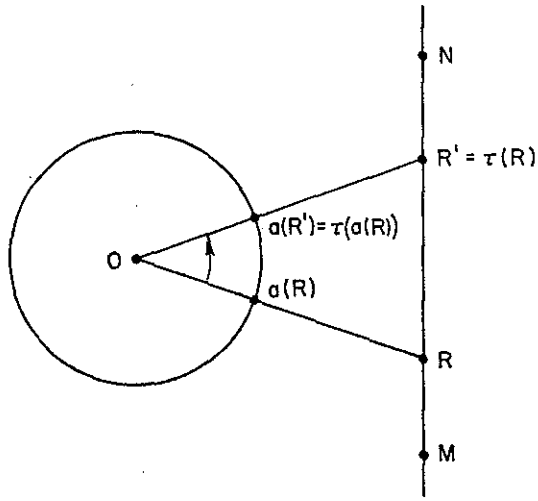

Fig. 10.

b in T. (Thus, the lines Ob and Oa are perpendicular in T.) Using the remark after Lemma 1, the new position $T(R')$ may be found from $T(R)$ by describing the angle θ between the lines $O \cdot \tau(b(R))$ and $Ob(R')$ (see Figure 12).

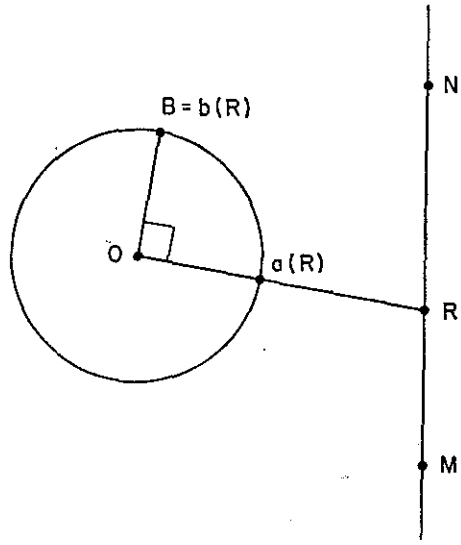

Fig. 11.

By Lemma 2, if the point of fixation is R, then the image of a portion of line segment MN centered at R will lie in plane O, M, N = plane $O, b(R), a(R)$ = plane O, b, a of T. If the point of fixation is moved to R', the image of a portion of line segment MN centered at R' will by Lemma 2 again lie in plane O, M, N. But since line MN is seen as straight, it follows by definition that the image will again hit the same elements of the retina and so lie again in plane O, b, a of T. Thus plane $O, \tau(b(R)), \tau(a(R))$ = plane O, M, N = plane $O, b(R), a(R)$ = plane O, b, a of T = plane $O, b(R'), a(R')$ and hence the angle θ must be O.

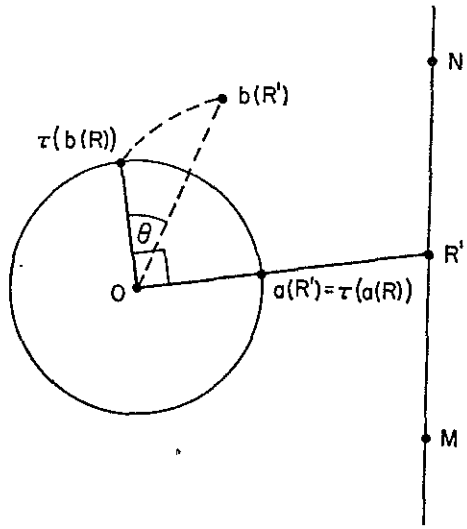

Fig. 12.

Thus, the position $T(R')$ is reached from $T(R)$ by a rotation τ along the great circle arc corresponding to RR'. Q.E.D.

It should be noted that we have not used anywhere near the full strength of the seen-as-straight assumption in this proof. This is why Lamb can say that the proof is really independent of the shape of the retina.

4. The rest of the argument is simple but ingenious. Here, we follow Lamb closely. Since a straight line in external space corresponds to a great circle arc on S, a triangle PQR in external space corresponds to a spherical triangle $P'Q'R'$. By Donders' Law, as we scan the perimeter of the triangle starting with point P (and eyeball in position $E(P)$ or $T(P)$) and eventually return to P, the eye returns to its original position $T(P)$. Suppose now that each straight line in external space is seen as straight. Then by Lemma 3, as the eye scans the perimeter of the triangle starting with point P, the eyeball rotates in order along the great circle arcs $P'Q'$, $Q'R'$, $R'P'$. But by a classical theorem of Hamilton [7], the resultant motion of the eye is a rotation about the line OP' through an angle equal to the spherical excess of the triangle $P'Q'R'$. This in general is not the same as returning to the original position. We conclude that not every straight line in external space can be seen as straight. Q.E.D.

Stanford University

BIBLIOGRAPHY

[1] F. Attneave, 'Some Informational Aspects of Visual Perception', *Psychological Review* **61** (1954) 183–193.

[2] R. Beals, D. H. Krantz, and A. Tversky, 'Foundations of Multidimensional Scaling: Metric Models Based on Ordering of Pairs', *Psychometric Society Meetings*, 1966.

[3] A. A. Blank, 'Axiomatics of Binocular Vision: The Foundations of Metric Geometry in Relation to Space Perception', *Journal of the Optical Society of America* **48** (1958) 328–333.
[4] A. A. Blank, 'Analysis of Experiments in Binocular Space Perception', *Journal of the Optical Society of America* **48** (1958) 911–925.
[5] D. Fender, 'Control Mechanisms of the Eye', *Scientific American* **211** (1964) 24–32.
[6] D. Fender, 'The Eye Movement Control System: Evolution of a Model', in *Neural Theory and Modelling, Proceedings of the 1962 Ojai Symposium* (ed. by R. F. Reiss), Stanford University Press, Stanford, Calif., 1964, pp. 306–324.
[7] W. Hamilton, *Lectures on Quaternions*, 1853.
[8] D. O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949.
[9] H. Helmholtz, *Wissenschaftliche Abhandlungen*, Vol. 2, J. A. Barth, Leipzig, 1883.
[10] H. Helmholtz, *Handbuch der physiologischen Optik*, L. Voss, Hamburg and Leipzig, 1896. (English translation: *Treatise on Physiological Optics* (ed. by J. P. C. Southall), The Optical Society of America, Menasha, Wisc., 1925.)
[11] J. E. Hochberg, *Perception*, Prentice Hall, Englewood Cliffs, N. J., 1964.
[12] H. Lamb, 'The Kinematics of the Eye', *Philosophical Magazine* **38** (1919) 685–695.
[13] R. K. Luneburg, 'The Metric of Binocular Visual Space', *Journal of the Optical Society of America* **40** (1950) 627–642.
[14] W. H. Pitts and W. S. McCulloch, 'How We Know Universals – the Perception of Auditory and Visual Forms', *Bulletin of Mathematical Biophysics* **9** (1947) 124–147.
[15] J. R. Platt, 'How We See Straight Lines', *Scientific American* **202** (1960) 121–129.
[16] R. M. Pritchard, 'Stabilized Images on the Retina', *Scientific American* **204** (1961) 72–78.
[17] P. Suppes and R. C. Atkinson, *Markov Learning Models for Multiperson Interactions*, Stanford University Press, Stanford, Calif., 1960.
[18] E. C. Zeeman, 'The Topology of the Brain and Visual Perception', in *The Topology of Three Manifolds* (ed. by M. K. Fort), Prentice Hall, Englewood Cliffs, N. J., 1962, pp. 240–256.

REFERENCES

[1] This is an assumption which should be studied more closely if the kinematic restrictions are weakened.
[2] This notion is more precisely defined below.
[3] Some such assumption is necessary if we are to make any sense out of our visual sensations. For, if the eye can fixate on the same scene in different positions, then different retinal images result.
[4] It should be noted that the relation between $E(P)$ and $E(Q)$, which can be derived from Listing's Law, it not quite so simple if $Q \neq A$.
[5] We are using 'group' in the technical algebraic sense. It might be more reasonable to think of a small set of allowable transformations and G as the group generated by them.
[6] Cf., for example, the discussion in Pitts & McCulloch [14].
[7] This is probably a little too simple-minded, for size constancy is observed only out to a reasonable distance from the observer.

[8] For example, the target may be fitted directly to the eye by means of a contact lens mounted with a tiny projector. See Pritchard [16].

[9] $[0, 1]$ is the set of all real numbers x with $0 \leq x \leq 1$.

[10] For simplicity we shall in the following discussion disregard rates of firing and identify firing with firing at (or near) full strength. If this simplification is dropped, the discussion is easily modified.