

\* \* \* \* \*  
\*  
\*                   STIMULUS ORGANIZING PROCESSES                   \*  
\*           IN STEREOPSIS AND MOTION PERCEPTION                   \*  
\*  
\*                   Peter J. Burt   \*  
\*  
\*                   COINS Technical Report 76-15                   \*  
\*                   (September 1976)                                   \*  
\*  
\* \* \* \* \*

COMPUTER AND INFORMATION SCIENCE

UNIVERSITY OF MASSACHUSETTS AT AMHERST  
AMHERST, MASSACHUSETTS 01002  
U.S.A.

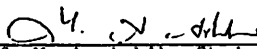
STIMULUS ORGANIZING PROCESSES  
IN STEREOPSIS AND MOTION PERCEPTION

A Dissertation Presented


By

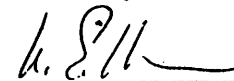
PETER J. BURT

Approved as to style and content by:

  
Prof. M. A. Arbib, Chairman of Committee

  
Prof. E. M. Riseman, Member

  
Prof. D. N. Spinelli, Member

  
Prof. W. Eichelman, Member

  
Prof. R. Graham, Department Head  
Computer and Information Science Department

ACKNOWLEDGEMENTS

I wish to express my gratitude to Dr. M. A. Arbib for his guidance in all aspects of this research, and for his part in building a diverse and stimulating computer science program at the University of Massachusetts. I am also particularly grateful to Dr. Bela Julesz of Bell Telephone Laboratories, Murry Hill, New Jersey, for an opportunity to conduct an experiment in his laboratory, and for very helpful discussions. I would like to thank Drs. Eichelman, Riseman, and Spinelli for their encouragement and for important discussions of various aspects of the research. I am most grateful to Deborah Burt for her contributions, which include typing this dissertation, and to Elliot Soloway for encouragement and insight, and to Michel Poe for help with photography.

This work was supported in part by NIH Grant No. 5 R01 NS09755-06 COM awarded to Dr. M. A. Arbib, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts.

A B S T R A C T  
 STIMULUS ORGANIZING PROCESSES  
 IN STEREOPSIS AND MOTION PERCEPTION  
 (June 1976)

Peter J. Burt, B.A., Harvard University  
 M.S., Ph.D., University of Massachusetts

Directed by: Professor M. A. Arbib

The idea that stimulus organization plays an important role in visual perception is now commonly accepted, principally due to the work of Julesz with random dot stereograms. In order for one to perceive depth in such a stereogram, a low level visual system mechanism must determine which of the many possible matches between individual dots of one half image with dots of the other are appropriate. Furthermore, it must enforce these matches during subsequent image processing. Stimulus organization plays a very similar role in motion perception: some mechanism must match each stimulus point seen at one moment in time with an appropriate point seen at a later moment in order for these points to be perceived as arising from a single moving object. But stimulus organization is important in many perceptual tasks besides stereopsis and motion perception, so in Chapter 1, its function and nature are discussed in fairly general terms, while

later chapters focus on binocular vision and apparent motion.

The essential function of stimulus organization is to segment the image or divide stimulus points into groups which "behave like" objects. This segmentation function is not only a useful but a necessary part of visual information processing when the visual stimulus includes images of more than one object. Stimulus organization resolves local ambiguities in how stimulus points are associated with one another, while it controls the association of individual stimulus points with perceived objects. It is argued that the processes responsible for stimulus organization operate at a low level of the visual system and require little or no high level, semantic control. This point is of particular relevance to computer vision as it indicates that appropriate low level general purpose processes can accomplish important image analysis without input from specialized programs which relate to specific objects.

The nature of binocular rivalry and fusion are studied in Chapter 2. Not only does rivalry provide one of the simplest and most easily examined examples of low level stimulus organization mechanisms, but it provides insight into the way in which image information is coded at the level of the visual system at which binocular information is combined. It is concluded that suppression is the result of inhibitory interactions between two cell populations which code the

visual input to the two eyes separately. A neural model is developed in which monocular images are coded by activity in cells with concentric center-surround receptive fields. Computer simulations of this model show that it is consistent with psychophysical suppression phenomena.

Chapter 3 begins with a comparison and criticisms of existing fusion models for stereopsis. These models all are based on a "projection field" concept. A new projection field model could be developed by combining the good points of these existing models and would account for a very large range of stereopsis phenomena. One difficulty with projection field models is that they incorrectly imply that a fixed relationship should exist between the perceived direction of a binocularly fused image and its binocular disparity. To overcome this difficulty, a modified projection field model is proposed in which there are separate but coupled "half projection fields," one for each eye. Another advantage of this structure is that the stereopsis model becomes a natural extension of the rivalry model in which separate image coding is also postulated. An interesting implication of this model is that stereopsis depends on neural interactions at an earlier stage of visual processing than has been thought. The possible involvement of the lateral geniculate body in stereopsis is explored briefly in Appendix A.

Finally, in Chapter 4, temporal aspects of stimulus organization are considered in the context of a model for

apparent motion. This model is shown, in an informal way, to be consistent with Korte's Laws, and it is discussed in relation to a number of other psychophysical phenomena. A possible combination of the stereopsis and apparent motion models is proposed to account for certain phenomena discovered by Ross which involve both depth and motion perception. Original psychophysical experiments are described in Chapters 2 through 4 which helped motivate the models proposed in these chapters.



## CONTENTS

Abstract	iv
List of Figures	x
Chapter 1: Stimulus Organizing Processes	1
1.1. Why Segmentation?	5
1.2. Segmentation and Theories of Form Perception	17
1.3. Time and Motion	28
1.4. Stimulus Organization	33
1.5. Examples	41
1.6. Organizing Processes	62
1.7. Computer Simulations	73
Chapter 2: Binocular Fusion and Rivalry	92
2.1. Introductory Examples	95
2.2. The Fusion Controversy	101
2.3. The Architecture of Binocular Interaction and Suppression	120
2.4. Is Information Lost?	128
2.5. Information Coding	134
2.6. Scale Factors	142
2.7. Summary	150
2.8. The Model, Part I: Image Code	157
2.9. The Model, Part II: Binocular Combination	171
Chapter 3: A Model for Stereopsis	179
3.1. Projection Field Models for Stereopsis	181
3.2. Problems of Projection Field Models	196

## CONTENTS

Chapter 3	
3.3. Scale Factors	203
3.4. Modified Projection Field Model	221
Chapter 4: A Process Model of Apparent Motion	235
4.1. Apparent Motion Phenomena - The Basic Paradigm and Examples	238
4.2. The Apparent Motion Model	273
4.3. Apparent Motion Phenomena Explained	292
Summary	307
Appendix A: Binocular Interactions in the Lateral Geniculate Body	313
Appendix B: Analysis of Depth Domain Inhibition	341
Bibliography	348

## LIST OF FIGURES

1.1	Flow Diagram of Stimulus Organization and High Level Analysis	4
1.2	Feature Detector Theory for Form Perception	8
1.3	Feature Detector Response to Disorganized and Double Images	10
1.4	Duck - Rabbit	12
1.5	"43" - "LB"	14
1.6	Face - Vase	16
1.7	Occlusion	18
1.8	Schema - Instantiation System	24
1.9	Teddy Bear	27
1.10	Differentiation - Integration System	37
1.11	Communication Between High and Low Level Processes	40
1.12	Stimulus Matching Ambiguity	45
1.13	Random Dot Stereogram	50
1.14	Processor Array	54
1.15	Examples of Organization Problems	57
1.16	The Problem of Motion Segmentation	61
1.17	Necker Cube Interactions	69
1.18	A Two Cell System with Reciprocal Inhibition	76
1.19	Graphs of System Behavior	82
1.20	Two Array Self Organizing System	84
1.21	Four Array System for Figure - Ground Separation	87

1.22	Computer Simulation of Figure - Ground Separation and Reversal	91
2.1	Stereograms Illustrating Binocular Combination Without Suppression of Contours	97
2.2	Stereograms Illustrating Suppression of Contours	102
2.3	Kaufman and Hochberg Stereograms	109
2.4	Stereopsis with Rivalry	113
2.5	Orthogonal Grid Stereogram	117
2.6	Neural Net with Spreading Recurrent Inhibition	118
2.7	Rivalry in Random Dot Patterns	121
2.8	Flow Diagram for Binocular Combination	124
2.9	Face Stereogram	132
2.10	Binocular Feature Detectors	138
2.11	Weighting Functions for Rivalry Interactions	144
2.12	Disk Stereograms	151
2.13	Image Code	159
2.14	Receptive Fields of Code Elements	165
2.15	Two Dimensional Receptive Fields	168
2.16	Array Code of a Bar Stimulus	170
2.17	Three Cell Network for Binocular Combination	173
2.18	Computer Simulation of Crossed Bar Stereogram	175
2.19	Computer Simulation of Kaufman's Stereogram	178
3.1	Projection Field	182

3:2	Ghost Images in the Projection Field	184
3:3	Cell Types in Sperling's Stereopsis Model	191
3:4	Bias Effects in the Projection Field	204
3:5	Disparities and Presentation Times for Two Ambiguous Stereogram Experiment	209
3:6	Sets of Disparity Values Used in Experiment	210
3:7	Experiment Results for Ten Cases	212
3:8	Data Showing the Effect of Presentation Time	214
3:9	Data Showing the Effect of Bias	215
3:10	Disparities Used in Julesz Experiment	220
3:11	Binocular Direction at an Occluding Edge	224
3:12	Suppression in Random Dot Stereograms	225
3:13	Coupled Monocular Projection Field	228
3:14	Domains of Inhibition Associated with a Match Cell	230
3:15	Coupled Projection Field Response to a Random Dot Stereogram	233
4:1	The Time Periods Defined for an Apparent Motion Display	240
4:2	The Simplest Apparent Motion Display	240
4:3	A Sequential Apparent Motion Display	242
4:4	An Ambiguous Periodic Display	242
4:5	Sequence of Patterns Used by Kolers	247
4:6	Beta Motion and "Retrospective Memory Readout"	250

4:7	Hysteresis and Fatigue	250
4:8	Land Square	253
4:9	Transition Data for Ambiguous Apparent Motion Experiment	259
4:10	Sperling's Quality of Perceived Motion	263
4:11	Values of E Which Correspond to Sperling's Results	265
4:12	Binocular Apparent Motion Display	271
4:13	Apparent Motion with $S=d/2$	274
4:14	A Model for Representation of Motion in the Visual System	276
4:15	Active Transient Response	280
4:16	Spatial Distribution of Response	280
4:17	Active Response to Two Stimuli	283
4:18	Active Response to Three Stimuli	286
4:19	Spatial Distribution of Response to Two Stimuli	287
4:20	A Moving Response Pattern	289
4:21	Inhibition and Facilitation Around a Moving Response Pattern	296
4:22	Stimulus and Perception of the Land Square	299
4:23	Stereopsis and Apparent Motion in Ross's Experiment	302
4:24	Depth in Panum's Limiting Case	303
A:1	Anatomy of LGN	319
A:2	LGN Neurons	321

xiv

A:3	Details of LGN Synaptic Patterns	324
A:4	Three Unit Projection Field	336
B:1	Network in Which One Inhibitory Cell Suppresses Activity in All But One Match Cell	

---

## CHAPTER I

### STIMULUS ORGANIZING PROCESSES

#### Introduction

The visual world is a three dimensional space filled with objects, and perception is concerned with the localization and recognition of these objects. Information available to the visual system is in the form of a two dimensional pattern of light on the retina. This image is not structured, as the world is structured, into objects. However, it is reasonable to suppose that the visual system is predisposed towards interpreting the image in terms of objects, and that even before specific objects are recognized, processes within the system are attempting to organize the retinal stimulus into regions which "behave like", and hence may correspond to, individual objects.

These organizing processes may play a major role in many perceptual tasks. The possible nature and function of such processes will be examined in this dissertation from a number of points of view. In the first chapter of the dissertation I shall consider questions such as (1) what are the organizing processes?, (2) how can neural activity organize stimulus information?, (3) are organizing processes necessary?, and (4) how do such processes relate to theories of form perception? Neural models are presented in the remaining chapters

for three specific perceptual phenomena, binocular rivalry, stereopsis and motion perception. Processing in each of these models may be viewed as organizing the visual stimulus.

One function of stimulus organization is image segmentation. The fact that the visual world is composed of objects, and that objects are rigid, compact structures which move in orderly ways according to physical laws, means that the image can be usefully broken down into subregions which also are rigid, compact structures which change in orderly ways over time. One or a small number of these subregions, or image segments, may correspond to single objects in space, so the task of object perception is considerably reduced if the image can be reasonably segmented.

The notion of an image segment is used here in a sense somewhat extended from Dev's model (Dev, 1975). Here an image segment is any area of the image which may be interpreted as a surface in space. All image features within that area are assumed to originate from object features which lie on that surface. Thus segments, like surfaces, may be pieced together to form more complex object representations in space. Two principal suggestions will be made: first, the segment representations constrain and organize subsequent image processing, and second, the segments themselves may be created and assembled largely by processes at a low level of the visual system, in many cases with little or no "high level" guidance. If these suggestions can be substantiated then the study of

organizing processes will be important, not only for understanding natural vision, but for developing computer vision as well.

Figure 1:1 shows a schematic diagram of the visual system with a separate box for stimulus organization. It is assumed that this box corresponds to a layered neural structure, in which each layer is retinotopically organized and different types of information are represented by activity in different layers. Thus, in some layers, patterns of activity represent a segment in terms of the area it covers, while in other layers, activity represents the segment boundaries, segment depth (distance from observer) and segment motion. This pattern of activity constrains the way in which information is processed by the remainder of the visual system, which is represented by the "brain" on the right in the figure. It is imagined that the organization box is chiefly responsible for processing and representing spatial information and that it includes a representation of spatial attributes, such as depth and motion, which are associated with individual features of perceived objects. Semantic information relevant to particular objects is processed in the remainder of the system. (It is not assumed that the neural structures responsible for these two functions are in anatomically distinct brain regions). The motivation and details of this system will be given in the remainder of this chapter.

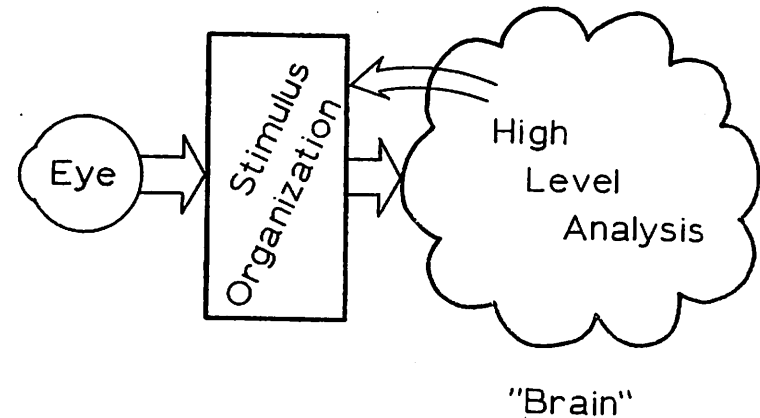


Figure 1:1

This diagram shows the presumed relationship between low level stimulus organizing processes and high level visual analysis. Spatial information, including the area and boundaries of segments, as well as segment depth and motion values, are represented by patterns of neural activity in the organization box, while information associated with the recognition of particular objects in the visual field is represented by activity in "labeled" cells in the "brain" region.

### 1.1. Why Segmentation?

The idea of image segmentation has been introduced as an effective means for reducing the problem of perception in a complex visual world into a number of simpler subproblems. This type of problem reduction is invaluable in computer scene analysis, where the strategy has been to begin analysis by examining subregions of the scene which may contain the image of a single, easily recognized object. Also, piece by piece analysis is particularly appropriate for computers which process information serially. But is this sort of problem reduction necessary in natural vision, where parallel processing is the rule? Another kind of consideration indicates that it is.

The need for image segmentation, or to be more specific, the need for a highly spatial, internal model of the visual world which includes a representation of image segments, is best demonstrated in the context of a feature detection theory of perception. According to proponents of this popular type of form perception theory, the first stage of image processing involves a reduction of the image into elementary features, such as edge segments, line segments, corners, etc. This collection of features becomes the input to the next stage of processing, where groups of elementary features are recognized as being components of particular complex features. At later stages of processing objects are recognized on the basis of appropriate complex or hypercomplex features. This system

is frequently imagined to consist of several arrays of feature detectors, which are arranged in a hierarchy. Each element of the lowest level arrays is "activated" when the specific elementary feature for which it is a detector occurs within its small receptive field. Elements in the next level of the hierarchy are complex feature detectors, and respond when appropriate collections of elementary feature detectors have been activated within their somewhat larger receptive fields. At the highest level, object detectors respond when all necessary complex features have been detected which make up the particular object which they code. An important part of the hierarchy concept is that detectors respond to their specific feature wherever it may occur within a receptive field. This receptive field is small for detectors at the lowest level but becomes progressively larger as one ascends the hierarchy. Thus a "face detector" at a high level responds whenever a face occurs within a receptive field which is large compared to the dimensions of the face itself. In general, there is a decrease in the location specificity of detectors as feature specificity increases. This convergence idea is included in the feature detection theory both to avoid a "combinatorial explosion" and to account for the invariance of form perception with changes in image position on the retina.

The system may be summarized with the type of diagram shown in Figure 1:2. Within the brain are distinct groups of cells which act as detectors for particular objects. We may

draw a circle around those which respond to any object class of interest, such as "face" detecting cells. The visual world is represented as the area inside the large circle in the left half of the figure. The face detector responds whenever a face image occurs within its receptive field, which is the area within the dashed circle in the figure. It is assumed that this face detector responds irrespective of the position, size, and (within limits) the orientation of the face image within the receptive field.

I will now describe several examples which will show that the system as outlined above cannot work. What is missing in this system is a spatially precise, internal representation of individual stimulus features and their association with individual perceived objects. While the descriptions of these examples will assume a feature detection system of form perception, they illustrate some general problems for perception theories.

First, suppose that an image is presented to the system in which the features of the face are rearranged (Fig. 1:3a). If the face detector responds whenever each of its required features occurs somewhere in its receptive field, then this image could be interpreted as a face. Clearly a theory leading to that prediction is incorrect. The spatial arrangement of the features is important and must be preserved through each of the stages of image processing, even if the position of the overall pattern is not. This provision appears to be

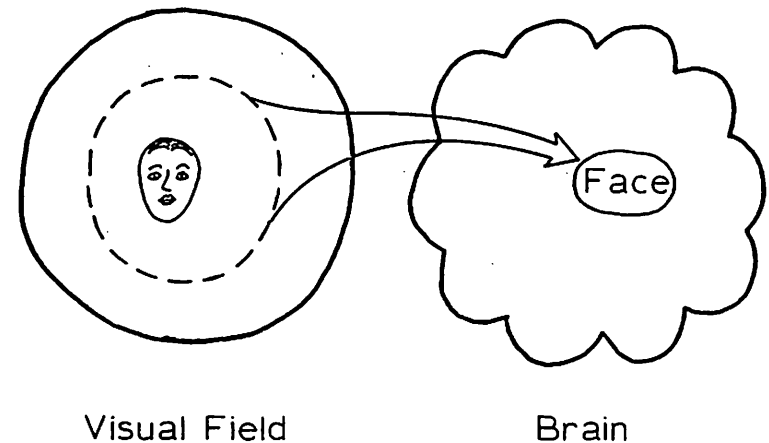


Figure 1:2

The feature detection theory for form perception postulates the existence of cells in the brain which become active whenever the image of a particular object occurs within a subregion of the visual field. This subregion, known as the receptive field of the object detecting cells, is shown here by the dashed line, and it is presumed to be much larger than the object image.



at odds with the assumption that spatial information relating to individual features is lost as information indicating the existence of these features is passed to the next level of the feature detector hierarchy. The simplest solution to the problem is to assume that the general purpose face detector of the present example has inputs from a great many specific face detectors, one for each possible position, size and orientation of the face image. But this is not a satisfactory solution to the problem as it implies a "combinatorial explosion": there must be specific detectors for every conceivable orientation, position and size of every perceivable object.

Next, suppose the system is presented with the images of two faces (Fig. 1:3b). If, as we suppose, the face detector responds whenever a face occurs anywhere within its receptive field, so that it does not distinguish images by position, can it possibly know that there are now two face images in different positions? Again retention of spatial information about each face image seems necessary.

The remaining examples will show that if the image presented to the system includes subimages of different objects, then an internal, highly spatial representation of the image is needed to keep track of feature usage. This is essentially a bookkeeping task. There are two types of image features in these cases: those features which belong to particular object images, and those which arise from the coincidental spatial relationship of one object to another. These features must

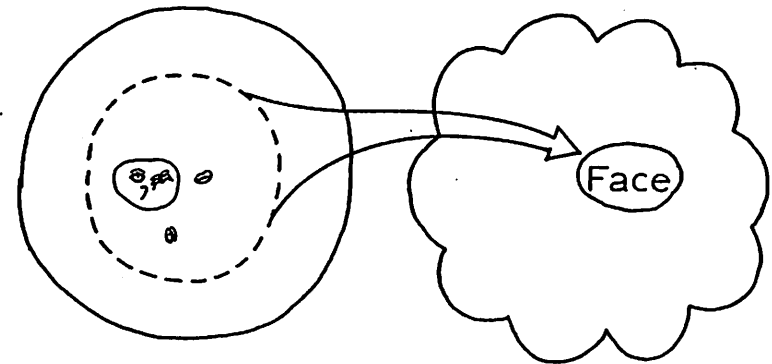


Figure 1:3a

The features of a face appear within the receptive field of a "face detector." How can the detector detect that the features are disorganized?

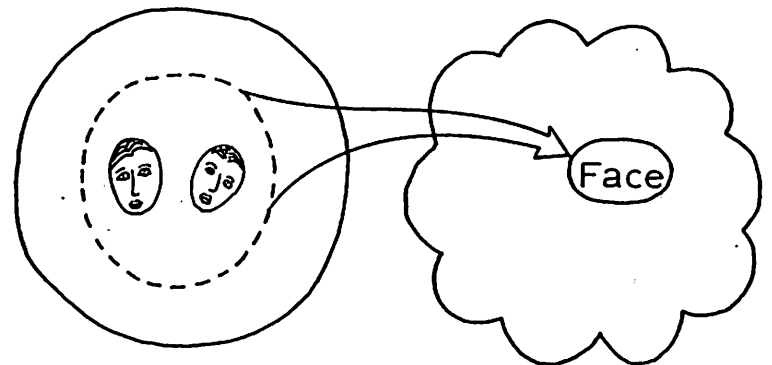


Figure 1:3b

Two face images appear within the receptive field of a "face detector." Can a single detector distinguish the two faces?

be distinguished by the system. All features of the first type must be associated with an object (a "completeness" condition), but no feature should be associated with more than one object (a "consistency" condition), while features of the second type should not be associated with any perceived object.

Figure 1:4a shows an ambiguous image within the visual field which may be seen either as a duck, with its bill to the left, or as a rabbit with its ears to the left.<sup>1</sup> The image is an adequate stimulus for both duck and rabbit detectors, as shown. The curious property of this figure (and other ambiguous figures) is that, at any given moment, one may see either a rabbit or a duck, but not both at once. Reciprocal inhibition might be postulated between duck and rabbit detectors to account for this mutual exclusion property, as shown. However, this supposition leads to an incorrect prediction in the case shown in Figure 1:4b, where separate images of a duck and a rabbit occur within the receptive fields of the detectors. Here as before only the duck or the rabbit should be perceived at a given time. To put this example in real life terms, direct mutual inhibition leads to the prediction that one should not be able to perceive both a duck and a rabbit simultaneously, if they are standing close

<sup>1</sup>This image is adapted from Wittgenstein's (1953) version of Jastrow's original duck-rabbit (Jastrow, 1899).

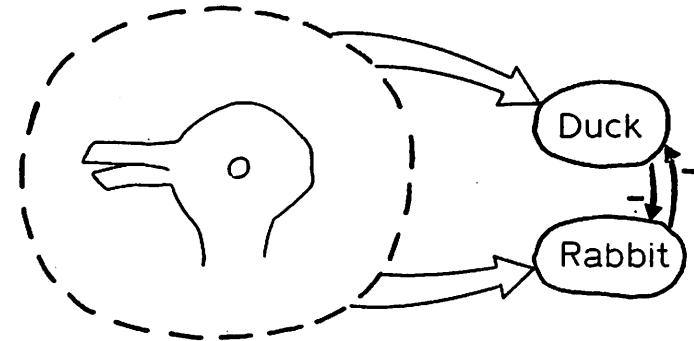


Figure 1:4a

This image may be seen either as a duck or a rabbit, but not both simultaneously. To account for the mutual exclusion property, we may postulate reciprocal inhibition between the duck and rabbit detectors.

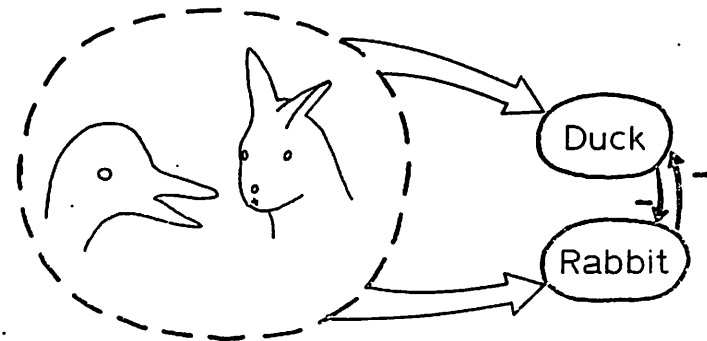


Figure 1:4b

Direct inhibition between detectors leads to the prediction that a duck and a rabbit cannot be simultaneously perceived if they stand close together!

together. This is clearly incorrect. The duck and rabbit should be mutually inhibitory, not because their images fall within the same receptive field, but because the detectors are competing for the same stimulus features.

A related example is shown in Figure 1:5. This ambiguous figure may be seen either as the numbers '4' and '3' or the letters 'L' and 'B', but again, it is not possible to obtain both perceptions simultaneously. Here we might explain mutual exclusion by saying that the detectors for '4' and '3' facilitate one another, as do the detectors for 'L' and 'B', while other combinations are mutually inhibitory, as indicated in the figure. Direct inhibition and facilitation between detectors is inappropriate here, as it was in the previous example, and some kind of competition for image features is implied.

This example includes an additional complication, since under either the '4'-'3' or 'L'-'B' interpretations, the characters partially overlap, so that there are extra features, corners and crosses, which would be detected by elementary feature detectors, but which should not be associated with the perceived characters. What features fall into this "to be ignored" class depend on whether one sees the '4'-'3' or 'L'-'B'. Again, a bookkeeping system is required so that no feature is ignored unless it can be interpreted as resulting from an overlap of object images.

Another example which illustrates the dependence of high level object perception on the allocation of image features

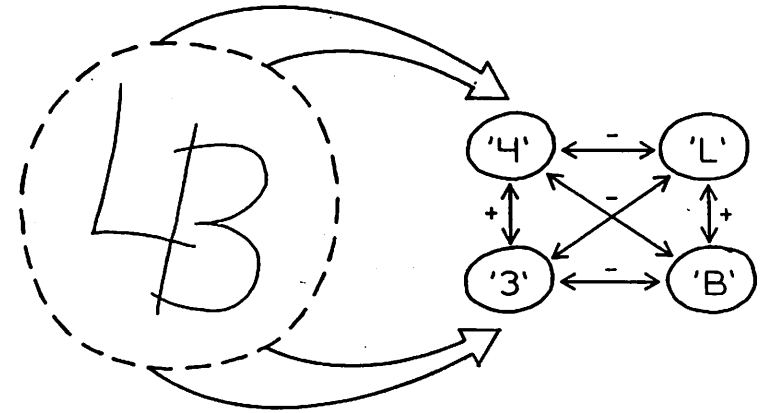


Figure 1:5

The image shown in the receptive field may be seen either as a '4' and '3' or as an 'L' and a 'B'. In either case, the characters overlap, so there are features which will be detected by elementary feature detectors, but which should be ignored by high level detectors. Different image features fall into this "to be ignored" category under the two perceptual interpretations of the image.

is shown in Figure 1:6a. Here the well-known ambiguous image of a vase or two faces is presented to the system within the receptive fields of the corresponding detectors. The fact that one cannot see both the vase and the two faces at once leads to the assumption that there is mutual inhibition between detectors. Again this should be interpreted as competition for image features. Thus when each contour of the image is doubled, as in Figure 1:6b, there are enough image features for both detectors, and the faces and vase can be seen together, even though the dimensions and positions of these images is almost identical to the previous case.

Finally, we should note the importance of a spatially precise internal representation for analysis of image occlusion. An object detector may be activated when not all of its required features occur in the visual image if the missing features are occluded by another object. However missing features can be ignored only if the occluding object is in the right spatial position to account for occlusion; it cannot just be 'somewhere nearby.' Thus the dot pattern shown in Figure 1:7a may be interpreted as vertices of a square, with one dot occluded by the black bar, or as a right triangle. The position of the bar differs slightly in Figure 1:7b and the possibly occluded dot proves missing, so only the triangle interpretation is possible.

To summarize the above discussion, it has been argued that perception of a complex image must include analysis of

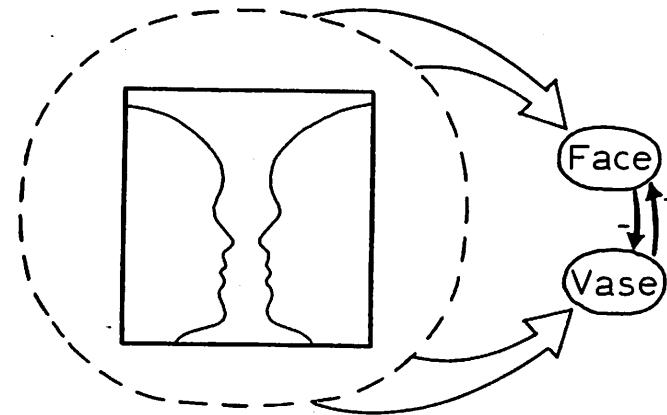


Figure 1:6a

Direct inhibition between face and vase detectors can be proposed to account for the fact that either two faces or a vase, but not both, can be seen in this familiar image.

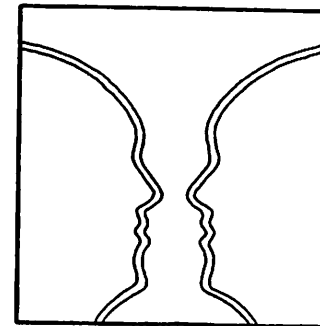


Figure 1:6b

Doubling image contours makes the vase and faces simultaneously visible.

the image into object related subregions, and construction of an internal representation of the image. At least part of this representation must be highly spatial so that it can serve to allocate image features to higher level object detectors. This internal model may also be called a spatial short term memory, SSTM, or a spatial internal model, SIM. Image segmentation is part of the SSTM (SIM).

#### 1.2. Segmentation and Theories of Form Perception

It seems that the principal objective of theories of form perception is to explain how the visual system can identify a particular image as being a member of a class of equivalent images. For example, all images of squares can be said to belong to the equivalence class "square," and we may ask how visual presentation of any member of this class can evoke the same perceptual response: "square." Since members of an equivalence class, in this and other simple examples, differ from one another in spatial respects, such as location, size and orientation, the strategy of many theorists has been to propose model perceptual systems which discard these types of information at an early stage of processing, while retaining space independent "universal" descriptors. However as the examples in the previous section demonstrate, this strategy has to be exercised with extreme caution, since precise spatial information is critical for complex image analysis.

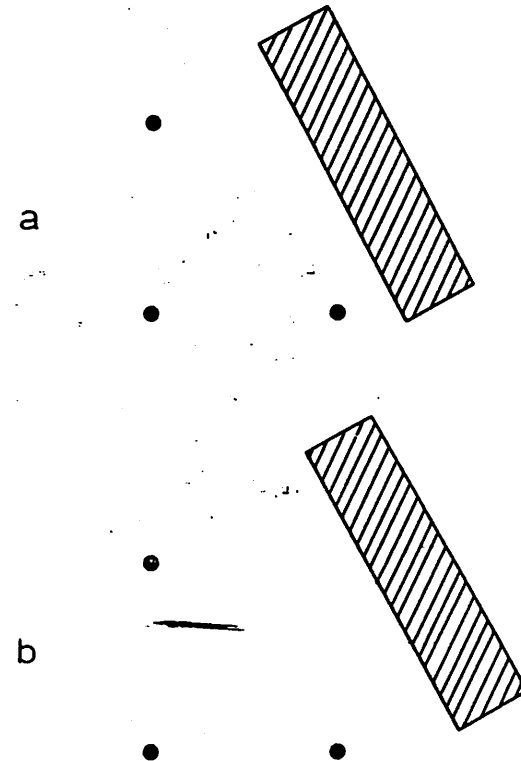


Figure 1:7

The three dots in Figure a may be seen as vertices of a square with the fourth dot occluded, but this is not possible in Figure b. This demonstrates the need to preserve spatial information during visual processing so that the system will know if missing features may be occluded.

In feature detection theories, as outlined above, members of an equivalence class are characterized by a collection of elementary or higher level features, which uniquely define the class. Thus all squares have four right angles and four edges, while triangles have three angles and three edges. These types of features are extracted at a low level of the hierarchical system, and object detectors, or equivalence class detectors, respond if the requisite features are found anywhere within a receptive field, irrespective of their spatial arrangement in that receptive field. This strategy for disposing of spatial information is counterproductive when there are several object images within the field of view, as we have seen. We may conclude from the discussion in the last section that feature detection schemes for form perception can be made to work only if they incorporate a spatial internal model of the visual world, which serves in part to segment the input image into object related subregions.

According to another type of theory, which is now enjoying increasing popularity, the first stage of image processing involves obtaining a spatial fourier transform of the visual image. This idea is attractive (to some) since, if one ignores phase relations, the transform of a given image is independent of its position in the visual field. While it is not independent of image size or orientation, position independence is viewed as an important first step toward reducing an image to a standard form, characteristic of its

equivalence class. However conversion to a frequency domain seems completely inappropriate in view of the need for image segmentation which has been demonstrated here. Segments are contiguous regions of space, so can be parsimoniously represented as areas in a spatial domain. Representation of segments in a frequency domain would be extremely awkward at best, and operations which need information about the spatial relations of two segments would generally have to obtain this information by retransforming the segments back into the spatial domain.

Pitts and McCulloch (1947) have proposed another system for associating an image with its equivalence class. Their idea is to generate all members of the class from the given image, by subjecting it to a group of translation, rotation and scaling transformations. "Universals" which characterize the equivalence class may then be obtained from the characteristics which all class members have in common. This approach to form perception also seems inappropriate if the image is complex and contains subimages of several objects. Unless the image is segmented first, so that each segment may be analyzed separately by the above procedure, many "universals" will be obtained which do not relate to any objects individually, but only to the chance spatial relations of two or more objects. It will be very difficult to sort out the universals which belong to individual objects and those which should be ignored. Even if segmentation is performed prior to transformation,

there will be occlusion situations which will be impossible to sort out in the transform domain.

I conclude from these considerations that the segmentation problem (or feature allocation problem) can only be resolved by a theory which postulates dynamic interactions between a low level highly spatial representation of the visual world, and high level object detecting units. By means of these interactions, the detectors compete for, and "lay claim to" individual image features, so that features claimed by one detector are not accessible to another.

The "schema" theory now being developed by Arbib (1975a) should be mentioned here as it is an example of a theory of form perception which can incorporate dynamic interactions between high and low level processes to account for feature allocation. While Arbib's objective is to develop a single, unified theory which integrates information processing in all sense modalities, as well as motor functions, I shall only mention some aspects of the theory which apply to vision.

The theory may be outlined in terms of the three component structure shown in Figure 1:8. The lowest level component is a retinotopically organized layer which contains the image coded in terms of elementary features. The highest level component is an amorphous structure containing "schemas." These schemas correspond roughly to object detectors in the feature detection theory. However not all schemas represent objects and the schemas interact in complicated ways

which need not be considered here. Also object related schemas are not passive detectors, but active processes. To account for feature allocation we may postulate that one operation performed by an activated schema is to build an "instantiation" of itself within the middle component of the structure in Figure 1:8. This component is a retinotopically organized, general purpose neural medium, in which information is represented in terms of patterns of activity. The instantiations are tailored to match input features, so correspond to specific members of the object class which is represented by the schema. If several examples of the object occur within the input image, there will be several instantiations of the same schema. Each instantiation will occupy an area of the middle layer which corresponds to, or "covers" the area of the bottom layer in which features of the object occur. A region of the instantiation layer which is occupied by one instantiation cannot be invaded by another. Thus image features which are covered by an instantiation are effectively captured by that instantiation and cannot be accessed by others.

If at some time there are no instantiations covering part or all of the instantiation layer, then the image features within the unclaimed area are accessible to all high level schemas. Individual schemas may be activated if one or more appropriate trigger features occur in the set of unclaimed input features. These schemas then attempt to build an

instantiation around the trigger feature. This attempt may be unsuccessful if other required features are not found in the right places, or if other schemas capture these features first. In this way schemas compete for specific stimulus features in a way which is consistent with the examples described in the previous section. We may note that there are also interactions between schemas within the schema layer and that there are inputs from other sensory modalities. These interactions relate to information which is not highly spatial, and their function is to potentiate specific schemas if other related schemas have become active.

Clearly at least part of the information represented in the instantiation layer constitutes a spatial internal model, or spatial short term memory, although other, non-spatial attributes may be associated with individual instantiations as well. I would suggest that the spatial portion of this information includes representations of image segments along with depth and motion values for these segments. This information may be contributed by the active schemas, or it may be derived by processes which are intrinsic to the instantiation layer. The latter processes organize and segment the input in ways consistent with object perception in general, so provide a "syntax" for object perception, while active schemas may modulate these processes in ways peculiar to the perceived objects or the current visual context, so these provide a "semantics" for object perception.

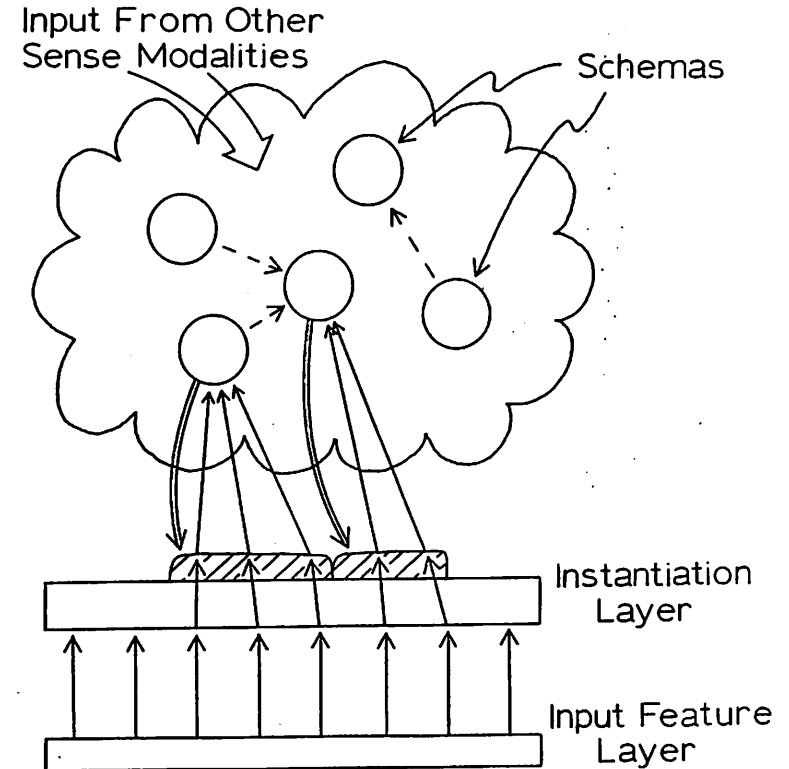


Figure 1:8

This schematic drawing shows how feature allocation mechanisms can be incorporated in the visual portion of Arbib's Schema system. The various interactions between high level schemas and input features by way of intermediate schema "instantiations" are indicated.



If we accept this model for visual perception and the distinction I have proposed between semantic and syntactic contributions to image segmentation, then we may ask what the relative importance of these contributions might be. Here the term "segmentation" means not only dividing the input area into subregions, but representing these regions as surfaces in space (which may be moving). As has been suggested, stimulus organizing processes may be sufficiently powerful in many cases to segment images, in the above sense, without input from high level schemas. This suggestion is supported by empirical data relating to binocular rivalry and stereopsis which will be discussed in Chapters 2 and 3 and by more hypothetical examples in which organizing processes may perform sophisticated analysis. These will be discussed later in this chapter.

Theories which attribute perception to mechanisms which classify a stimulus but which do not involve instantiations do not account for the subjective experience of perception. One's conscious perception includes more than some kind of designation of the equivalence classes to which images belong. When the image of a square is viewed, one sees not just "squareness," but a particular member of the class "square." One is aware of the stimulus itself, with the interpretation "square" somehow superimposed as an additional, organizing characteristic. In the case of the duck-rabbit, one is aware of exactly the same stimulus points when he sees the rabbit

as when he sees the duck; any little marks in the drawing or irregularities in a contour can be examined equally well when the pattern as a whole is a rabbit as when it is a duck. Thus no "ideal form" of an object is evoked when a particular example of the object is presented. Rather, the perception is somehow a transparent interpretation laid over the stimulus pattern, as with an instantiation. It is this non-visible component which changes when one alternates between the perception of a rabbit and a duck, and it is at that level that competition between detectors takes place. The conclusion to draw from subjective experience is that spatial information is not discarded by the visual system, and the interpretation of an image segment is somehow united with the stimulus points it covers.

At this point it is appropriate to qualify the general conclusion that no one feature can be part of two perceived objects simultaneously, and the observation that stimulus organization and object perception do not alter the appearance of individual stimulus features. There may also be evidence that the interpretation of a stimulus image takes place in a hierarchical structure, such as the one outlined in the feature detection model. This is illustrated, for example, when one looks at the image of a teddy bear, in Figure 1.9, which has button eyes and stitched nose and mouth. It seems to be possible to see the buttons both as buttons and eyes of the bear at the same time. If this is the case, the



Figure 1.9

In this image of a teddy bear, it seems to be possible to see the eyes as buttons and eyes simultaneously, suggesting that under some conditions, a set of stimulus points can be interpreted in two ways at once.

implication is that the same stimulus points can be perceptually associated with two objects simultaneously, if one of these objects can be interpreted as a part of the other.

Finally, we should note that perceptual interpretations do seem to modify image features in some cases. In images which show figure-ground separation, the figure region may seem brighter than the ground region when the actual intensities of the two regions are the same. Anomalous contours are another type of perceptual response around which there are apparent changes in image brightness. In the case of binocular vision stimulus points in the combined view may appear displaced relative to their position in either monocular view, and with rivalrous images some points may be "suppressed" in the binocular view. These last cases illustrate fairly major changes in image appearance related to perceptual processing. They will be considered in greater detail in subsequent chapters.

### 1.3. Time and Motion

In the discussion in the last two sections, I have assumed that the visual system is presented with a single image which it then analyzes. No consideration has been given to perception of changing images. This is the point of view taken in computer scene analysis. It has been argued that perception requires the construction of an internal

representation of the visual world, and at least part of this representation is low level and highly spatial. The low level portion of the internal model includes the representation of image segments and other types of spatial information. A parsimonious assumption is that this information is represented by a pattern of neural activity in a retinotopically organized layered neural structure.

This low level representation is necessary to perform an organizing function in keeping track of feature allocation. Segmentation has also been suggested as a means for reducing the perceptual task, as it seems likely that analysis of a complex scene will be possible only if subregions of the scene can be cordoned off and separately recognized as images of familiar objects.

The fact that the visual system must operate continuously and in a changing environment adds an important dimension to these considerations. It is clear that perception cannot involve a "one way" flow of information, as processing of some image features is always contingent upon the results of processing of others. There must be feedback from high level processes to low level processes, so that particular image features may be claimed by activated object detectors. Processing which relies on feedback necessarily takes time. Thus if the visual system analyzed a continuously changing image as a sequence of more or less independent picture frames, there would be a maximum rate at which analysis can

be accomplished. This system would necessarily be very inefficient. Reanalysis of the scene would have to be sufficiently frequent to detect small changes which may be of survival value, which means that, in most cases, very little would have changed in the scene from the time of one analysis to the next, and processing is terribly redundant.

A third function of the spatial internal representation is to make image analysis efficient over time. Once the representation has been developed, it may be compared to the subsequent patterned input information, so that areas of image change can be detected. Reprocessing is directed only at these regions of the image.

We should think of segments as having a temporal dimension as well as a spatial dimension. In physiological terms, the pattern of neural activity which represents a segment is stable in time. All stimulus points falling within the area covered by the segment, and within a period of time, are perceptually integrated and associated with a single object. It is this temporal dimension of the spatial internal representation which accounts for perceptual stability: when one views an object image for several moments, he perceives one object which is continuously present over this time, rather than one object which is suddenly replaced by another identical object. Perceptual stability is also made apparent by ambiguous images, which are perceived in one or another of their states for prolonged periods of time.

The above considerations apply to moving images as well as to stationary images. As long as motion is continuous and predictable, it should not be necessary to regenerate the internal representation of the moving segment at every moment in time. Since this representation is a retinotopically arranged pattern of activity in a layered neural structure, the representation can be maintained if the pattern moves within the structure as images move on the retina. I have shown elsewhere (Burt, 1974, 1975) that activity patterns can move within neural structures without change in shape.

There are several types of orderly motion the system should anticipate. These will be listed here because they show the importance of representing various types of spatial information with the image segments.

1) Image motion due to eye rotation. This is the simplest type of motion, since all features of the image move together at a velocity which is equal to the velocity of eye rotation. Information about eye rotation should be provided to the visual system in the form of a "corollary discharge" from the oculomotor system.

2) Image motion due to constant object motion should be anticipated. In this case, image segmentation is necessary since different objects move at different velocities. A perceived object velocity is associated with each segment.

3) The system should anticipate relative motion of

object images due to the parallax effect as the observer moves through his environment. Depth information must be associated with image segments in this case, along with a complicated corollary discharge from the motor system. (Actually this corollary discharge may not be necessary, as the visual system may deduce observer motion from perceived image parallax motion).

4) The system should anticipate the orderly disappearance of image features as one object moves in front of and occludes another. This requires that depth values be associated with individual segments, so that the system can know which segment will be occluded as two move towards each other.

Several types of empirical evidence may be cited in support of the suggestion that the spatial internal representation is updated in anticipation of image motion. First, we should note that the perception of ambiguous images, such as the Necker cube remains stable even when the image is moving on the retina. Careful study might reveal that motion changes the reversal rate somewhat.

This stability seems to occur also with ambiguous random dot stereograms, even in cases where eye movements cause an ambiguous region of the stereogram which is seen at one depth to fall on a portion of the retina which prior to the eye movement had been exposed to an area of the stereogram which had been seen at a different depth.

The most interesting type of motion mentioned above is

parallax motion. That the visual system can anticipate this type of motion is implied by the fact that parallax motion is an effective depth cue. The same neural mechanisms which interpret this cue can update the internal representation in anticipation of changes due to parallax motion.

#### 1.4. Stimulus Organization

Thus far, I have discussed the need for a low level, highly spatial internal model, and have suggested that the representation is in the form of a stable, possibly moving pattern of activity within a retinotopically organized neural structure. In this section, I consider how this pattern of activity may "organize" the incoming information, and in a later section, I consider processes which organize the activity patterns themselves.

Again, one critical function of the spatial internal model is to represent image segments, which may correspond to individual objects. These dynamically defined regions must be tied together by integrative processes, while being isolated from one another by segregation processes. Dev's segmentation model (Dev, 1975) gives us a way to begin to think about the neural structures which might support these integration and segregation processes as they apply to the spatial dimension of the visual input. However, as has been said, integration occurs in time as well as space, so that if an

image falling on the retina at one time closely resembles an image falling on the retina at a just previous moment, these are perceptually treated as images of a single object.

Since the spatial short term memory includes information which has a high degree of spatial and temporal resolution, it must be represented by activity at a low level of the visual system, a level at which the requisite spatial and temporal information is still available. On the other hand, we know from recent physiological studies that the very low level neurons of the visual system tend to be differentiators rather than integrators. Thus, there are cells in the retina and striate cortex which respond maximally to stimulus intensity variations in the spatial dimension, the edge detectors, others which respond to temporal variations, the "on" and "off" detectors, and still others which combine these temporal and spatial differentiation properties and are moving boundary detectors.

That the initial stage of visual processing should be differentiation seems reasonable, since that process acts as an initial information filter, drawing attention to changes in the visual field. Thus sophisticated vision systems should include populations of cells which act as differentiators, and populations which act as integrators. These populations are shown schematically in Figure 1:10, where for simplicity, cells of the two types are separated into distinct neural structures. In an actual visual system, the two

types of cells might be physically mixed together, or both types of functions might be performed, to some extent, by individual neurons. The assumed function of the various structures is as follows:

Retina: The obvious functions are assumed here; receptors respond to optically formed images.

D<sub>1</sub>: This is the first differentiation layer, and includes ganglion cells and cells of visual cortex which act as edge detectors, on-off cells, and moving boundary detectors.

I<sub>1</sub>: This is the first integration structure, and it is here that patterns of neural activity represent the spatial short term memory. This structure corresponds to the "stimulus organization" box of Figure 1:1. The various models described in later chapters will be concerned with processing within this structure. The structure is assumed to include several interconnected layers which are retinotopically organized, and different sorts of information are represented by the activity in different layers, including image segments and perceived distance and object motion associated with each segment. These are integrated types of information, since there is no direct image stimulus for them. Again it should be stressed that these "layers" may be distinct cell populations retinotopically distributed within a single anatomical layer.

D<sub>2</sub>: The second differentiation layer is similar to the first; it contains edge, change and motion sensitive elements.

But now the stimulus for these elements is not light on the retina, but activity in the I<sub>1</sub> structure. Thus, object edges and object motion may be directly sensed at this level. These types of features augment the retinal image, so that subsequent processing centers "see" not only the image features, but a (tentative) assignment of the features to surfaces in depth.

I<sub>2</sub>: This structure represents the rest of the brain, or the high level processes of Figure 1:1. We suppose that specialized object detectors and schemas (in Arbib's model) are located here. I am not concerned, in the present discussion, with processing in I<sub>2</sub>, but communication between I<sub>1</sub> and I<sub>2</sub> will be considered.

There are two ways in which activity in I<sub>1</sub> can organize the incoming stimulus information. One function is to constrain processing at higher visual centers in the ways which have been described in earlier sections: for example, by controlling the assignment of image features to object detectors. A second is to constrain processing within the I<sub>1</sub> structure itself. With respect to the first function, we may ask how a low level, highly spatial vision center can communicate with specific high level non-spatial centers. It might seem that direct paths between the centers would have to be dynamically established, which link the information about object kind with object position. This type of directed communication is difficult to envision, and I would suggest that it is not

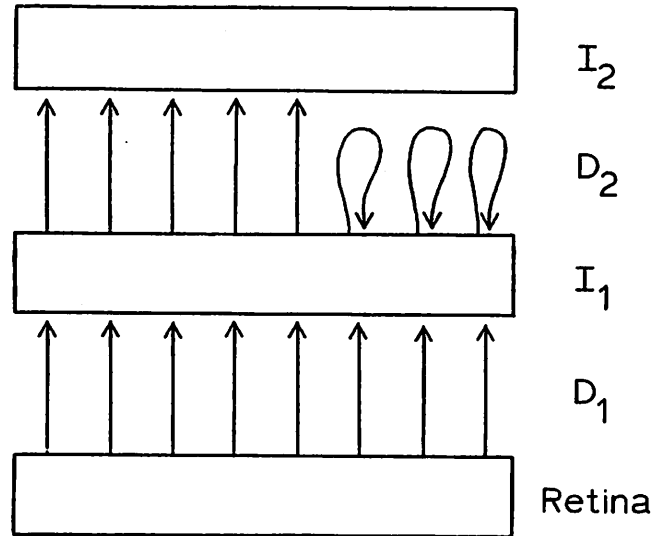


Figure 1:10

Processing of the visual system is depicted as alternating stages of differentiation, in  $D_1$  and  $D_2$ , and integration, in  $I_1$  and  $I_2$ . Layer  $I_1$  corresponds to the stimulus organization box of figure 1:1.

necessary. Any two processing centers of the system can communicate with each other via a common "data bus" if messages are "tagged" in some way. To clarify this idea, suppose initially that all image features are available to all object detectors. Physiologically this means that all afferent information projects to all the groups of cells which code specific object types. The situation is shown schematically in Figure 1:11 which is a reorganized rendition of Figure 1:1 showing lines of communication between the organizing box and detectors within the high level visual region. As image information passes through the organization box, certain perceptual attributes become associated with individual features. These include depth and motion values and perhaps an average position value which characterizes all features within a segment. The feature information, augmented with perceptual attributes, then passes to all object detectors via the data bus. A given detector can respond if the appropriate features occur in the common input, but only if the attributes associated with these features all have nearly the same values. The features which are organized into a single segment are all assigned nearly the same attribute values, while features in different segments will generally differ significantly in one or more attributes. Thus association of attributes with features may be the principal mechanism for constraining the use of features to single object detectors.

The reverse communication, from object detectors to low level representations, can also make use of a common bus. This efferent projection cannot be spatially precise and directed at activity in a specific location of the organizing box, since we assume detailed position information is not used by the object detectors. Thus, an object detector must communicate with the corresponding segments in the organizing box via a diffuse projection. However, again we assume the messages are "tagged" so that they can be appropriately sorted out in the receiving area. Tagging in this case could be with the same attribute values which characterized the features which activated the detector. Harth (1976) proposes a similar principle in his model for communication between cortex and the lateral geniculate body.

A second mechanism by which patterns of activity in the stimulus organizing box may organize stimulus information is by modulating information processing within the box itself. For example, the patterns of activity which represent segment boundaries control intrinsic integrative processes, so that integration over any spatial domain does not cross a perceived object boundary. Thus, whatever stimulus attribute is being integrated is associated with a single segment. In addition, the boundary representations control association of image features which are detected at boundaries with the appropriate neighboring region. (Note: boundary features, such as motion or depth, should be associated with only one

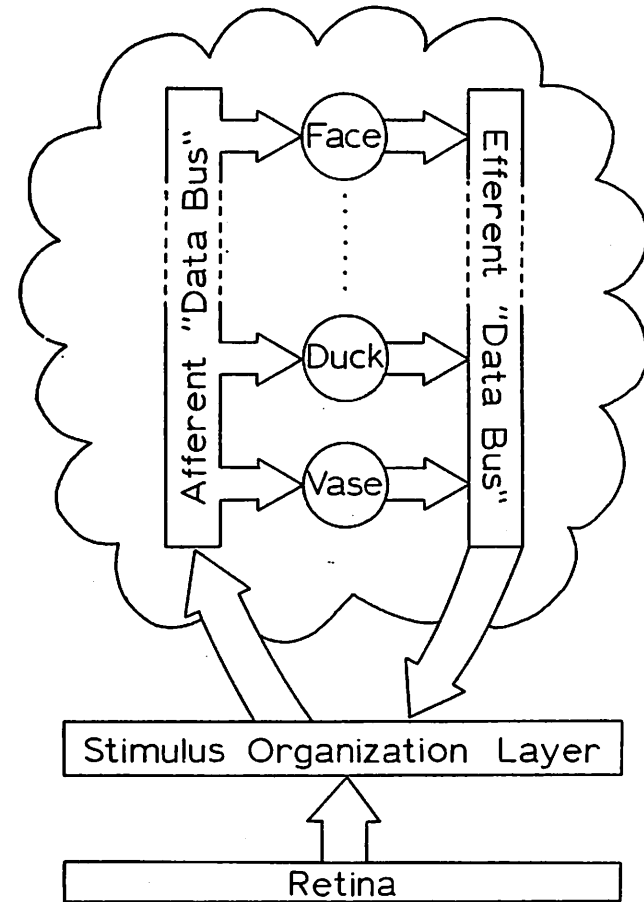


Figure 1:11

This diagram is an elaboration of Figure 1:1 which shows a possible means of communication between the low level spatial information in the stimulus organizing box and high level object detectors.



of the two areas which the boundary separates). The nature of these interactions will be discussed in more detail later.

### 1.5. Examples

Several examples will now be described, both to illustrate the structure outlined above and to motivate a discussion of organizing processes which will follow. The first examples have to do with stereopsis and motion perception. A global organization of the stimulus pattern in each of these cases is proposed as a mechanism for resolving a local stimulus matching ambiguity.

Stereopsis and motion perception involve very similar computational tasks. In both cases, the desired information is not contained in a single brief visual image, but must be obtained through a comparison of two or more images. In stereopsis, depth information is obtained by finding the difference in the positions of objects or features in the images presented to the two eyes. This binocular disparity is determined at a given moment in time. For motion perception, the same disparity information must be obtained, but in this case, the comparison is between images to the same eye, (or to both eyes together - the "cyclopean eye") at slightly different moments in time. This computational similarity suggests that similar neural mechanisms may be responsible for both stereopsis and motion perception, and that the respective

computations may be performed at about the same stage of visual processing.<sup>2</sup>

Not surprisingly, theories of stereopsis and motion perception have had much the same history. In particular, in both cases, an old view, that disparity information is computed at a late stage of visual processing has now been replaced by a conviction that computation must occur at the initial stages of processing. According to the old view, each image, i.e. the stimulus presented to a particular eye and at a particular moment, is processed more or less independently to the stage at which objects are identified and localized. Then objects are matched between images and the difference in their positions computed. According to the new view, individual points or small features are matched between the two images before image segmentation or object identification.

The old high level view was appealing on several accounts. Since it is clear the the visual system can and does reduce a visual image to a number of perceived objects with associated directions in the visual field, it is a simple matter to proceed from there to a direct comparison of perceived directions for objects in two or more images.

---

<sup>2</sup>In fact there is evidence of a direct interaction between processes. For example: 1) Random process stereograms, in which there is time delay between stimulus presentation to the two eyes, but no spatial disparity may yield both depth and motion (Ross, 1974); 2) Motion parallax yields a depth perception; 3) Interactions exist between apparent motion and depth (Julesz, 1971; Kolers, 1972).

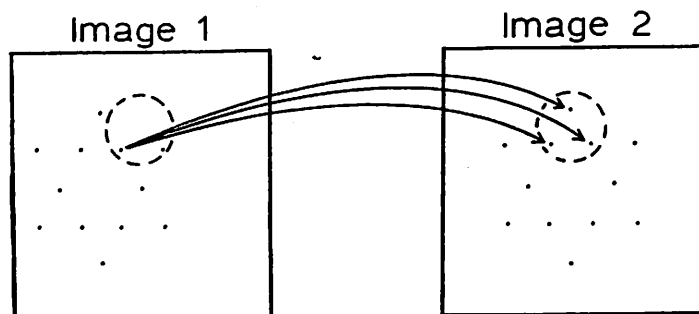
Furthermore, there is a significant computational difficulty associated with the alternate idea that disparities are computed at an early stage of visual processing between image points or features. In a typical pair of images, a small feature in one image can be matched with any one of a number of similar features in the other image. How is a low level, local feature matching mechanism going to decide which of the possible matches is appropriate without bringing in more global information about the overall pattern of features within the images, as in the high level matching scheme?

This local matching ambiguity may be illustrated in a couple of ways. Suppose first that the two images consist of a star-shaped dot pattern, Figure 1:12a, and that a mechanism exists which matches single dots in one image with single dots in the other. Each image 1 dot should be matched with one and only one image 2 dot. Notice that the star pattern is shifted in one image relative to the other. but by hypothesis our point matching mechanism cannot know this. The matching mechanism can look only at local regions in the same position of both images (as shown by dashed circles) and any match within the window is permissible. The matching problem in this example is like that which exists in a Julesz random dot stereogram, or in an apparent motion display in which a pattern of stimulus points is stroboscopically presented.

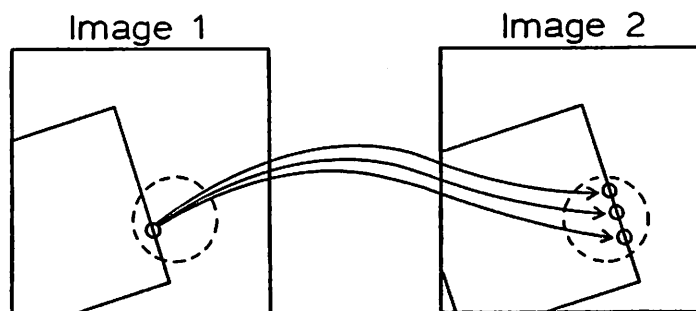
As a second example, suppose each image is an outline drawing of a geometric figure, Figure 1:12b. In this case

the elementary features which must be matched between images might be short line segments, as are shown in the small circles. The matching procedure can pair similar features falling within a larger image region, as indicated by the dashed lines. Again many matches are possible, and each one implies a different image displacement, and hence, if these are pictures taken at different moments in time, a different velocity.

Despite these apparent difficulties in resolving local matching ambiguities, the current conviction is that stereopsis and motion must be computed at a low level of the visual system and without access to overall pattern information. This shift in point of view was the result principally of certain recently discovered psychophysical phenomena. In the case of stereopsis, Julesz (1960) has elegantly demonstrated with random dot stereograms that image points can be matched and depth perceived without prior perception of objects. Similarly experiments with apparent motion have shown the existence of "objectless motion" ( $\phi$  motion) in which motion is sensed between two apparently stationary stimuli (Wertheimer, 1912). Also, when the stimulus for apparent motion is a periodic pattern, the pattern may not appear to move as a whole (i.e., as a moving global pattern), but may break down into regions of motion in different directions. This will be described in Chapter 4.



(a)



(b)

Figure 1:12

Two examples are shown here of the local stimulus matching ambiguity. It is assumed that any elementary feature within the local region encircled by the dashed line in image 1 can be matched by the system with any similar feature in the corresponding local region of image 2. Each match implies a different depth or motion.

The idea that stereopsis and motion are processed at a low level of the visual system is appealing for other compelling though not deciding reasons. For example, the recent microelectrode recording from individual neurons in visual cortex have shown that many of these cells, which must play a role in the initial stages of processing, are sensitive to motion and binocular disparity. Also, from a computational point of view, low level extraction of motion and depth information seems much more elegant than high level processing, provided the problems of matching ambiguity can be resolved in a satisfactory way. It should not be necessary to repeatedly go through the very complex processing required to analyze images into objects in order to obtain the much simpler disparity information for depth and motion perception.

The local matching ambiguity can be resolved at a low level of the visual system by stimulus organizing processes. The idea is simply that, in addition to local feature matching processes, there should be local constraints on how neighboring features may be matched. The matching processes are then restricted to local matches satisfying these neighborhood constraints. The pattern of point to point matchings between two images is said to be globally organized when local constraints are satisfied over the entire image. When the local constraints are properly defined, the global organizations should correspond fairly well with a matching between images on the basis of high level pattern or object

information.

Roughly speaking, the local constraint appropriate for stereopsis and motion perception is that neighboring points of one image should be matched to neighboring points of another so that the disparity between point pairs is the same. This constraint, when applied to images containing a random dot stereogram, results in extended areas over which points are matched at a single disparity. The corresponding perception is consistent with psychophysics - extended dense planes seen in depth. The same constraints applied to motion computation would result in extended areas of uniform motion, moving objects.

The idea that local constraints may account for global organization in stereopsis was first proposed and modeled by Julesz (1971) and has since been incorporated in other models, including the model which will be described in Chapter 3. To my knowledge, no models of motion perception have dealt with the problem of matching ambiguity, and none make use of the ideas of local organizing processes and constraints.

The relation of stimulus organization in stereopsis to information processing in the structure shown in Figure 1:10 is this. Instead of one retina, there are two. Afferent cell activity is projected from the retinae to layers in the stimulus organizing structure,  $I_1$ . There information from

the two eyes is retinotopically organized, but local features seen by one eye may be matched with any features seen by the other eye within a range of disparities. Details of this matching procedure will be left for Chapter 3. Here we only want to make several points of general relevance to stimulus organization.

The neural activity which codes afferent image information resides in a separate population of cells from that which codes image segmentation and controls feature matching. This second type of activity has several components. Local activity in "segmentation layers" represents surfaces in depth and controls feature matching in the corresponding local regions of the image representation layers. Local activity in another layer represents the location of segment boundaries. As mentioned earlier, an edge is associated with only one of the segments it separated, and this is the segment which is perceived as nearer to the observer. This fact has several important consequences. First, integration processes which respond to various surface features, such as color, local depth and motion stimuli, etc., should not integrate across segment boundaries, as this would lead to a confusion of features of one object with features of another. Thus, one function of the activity which represents segment boundaries is to appropriately constrain these integration processes. Second, many depth and motion stimuli are not "area" stimuli, but occur only along the boundaries of object

images. Since the activity in  $I_1$  which represents segment boundaries is associated with only one of the segments it separates, this activity may also cause edge stimuli to be integrated with the appropriate segment. Finally, when adjoining segments move relative to one another, the representation of the boundary between these segments will indicate which surface is moving to occlude the other, and on which side of the boundary features will disappear due to occlusion.

Now suppose that the input image is a random dot stereogram, such as the one shown in Figure 1:13. This stereogram is constructed so that, when binocularly viewed, all dots within a central square shaped area appear at a second depth. A curious fact about this effect is that individual dots are not perceived as points suspended in space, but as points lying on a surface. This surface seems smooth and continuous over both the central square and the surrounding regions, but there is a sharp boundary dividing the two regions. Since there are no specific edge or surface stimuli in the random dot images, these perceptions are anomalous. However, they may be the psychophysical correlates of neural activity which has been postulated in  $I_1$  of Figure 1:10 to represent segment edges and segment areas respectively.

Our supposition that image features and image segments are represented by activity in separate cell populations is consistent with observations made by Ross (1976) with random process stereograms. These stereograms are similar to the

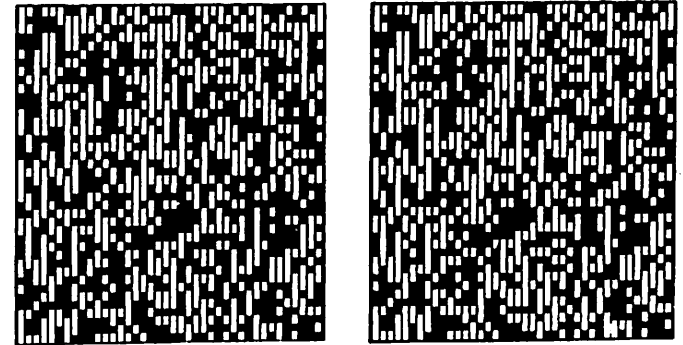


Figure 1:13

This Julesz type random dot stereogram is constructed so that, when stereoscopically viewed, all dots within a central square shaped area appear nearer the observer than do dots in the surrounding area.

more familiar random dot stereograms of Julesz, but the dot pattern is presented on a CRT screen and is continually changing. The stereogram is constructed so that dots presented within the central square area always appear at one depth, while those in the surround appear at another. Again binocular combination of this type of stereogram results in the perception of solid surfaces in depth which are separated by a sharp boundary. These anomalous perceptions are stable while the real stimulus pattern continually changes, like a "snow storm."

It is interesting to note that the amount of time needed to clearly see surfaces in depth in a stereogram like that shown in Figure 1:13 is about 50 msec. (Julesz, 1964). We may interpret this as the amount of time required for stimulus organization to be achieved and for segment representations to begin to emerge in  $I_1$ . In a random process stereogram, the stimulus pattern may be completely changed in a small fraction of this time, and yet stereopsis is easily achieved and maintained. Again our interpretation is that once an organization has been achieved and is represented by activity within one population of cells, it can continue to match points in the rapidly changing stimulus pattern which is represented by activity within another population of cells.

Ross has also found that if there is a small time delay, say 50 msec., between the time each stimulus point of a random process stereogram is presented to one eye and the time

it is presented to the other, one may perceive both depth and apparent motion. He reports that apparent motion never causes dots to cross the boundary between the central and surrounding areas. This is consistent with the functional significance attributed here to the boundary representations in  $I_1$ .

The study of stereopsis has revealed a number of phenomena which are best characterized in terms of stimulus organization. Other examples will be mentioned later in this chapter and some of these phenomena will be studied in Chapter 3. The above examples were described to illustrate the supposed roles of the area and edge components of the segment representations. It was also suggested that these representations correspond to the anomalous surfaces and contours experienced in stereopsis. A similar interpretation for anomalous contours has been proposed by Frisby and Julesz (1975).

We now turn from the examples of stereopsis, where there is much evidence for stimulus organization, to examples involving other perceptual functions, where less evidence is available, but where, for computational reasons, such processes seem appropriate. Here we will consider the nature of stimulus organization in a somewhat more abstract, but perhaps more explicit way.

Suppose the visual system consists simply of two two-dimensional arrays as in Figure 1:14. The first array contains the input stimulus, which might correspond to a

digitized photograph, or might contain some other stimulus type, such as local motion or depth cues. The second array is an array of "processors." All processors are identical, and each may be in one of a number of possible internal states at a given moment. In addition, there are a number of local consistency constraints, which are rules indicating when neighboring processors are in mutually consistent states, and when the state of a given processor is consistent with the value in the corresponding input array element. The state of the system is simply the combined states of all the individual processors, and the values of the stimulus array. The array is globally organized if it is in a state such that all local consistency constraints are satisfied over the entire array. Depending on what local constraints are defined, there may be many global organizations, only one, or none. The task of organizing processes is to put the array into a globally organized state by appropriately changing the states of individual processors. Some examples will clarify these definitions.

Figure-ground. Suppose that the input image is an outline drawing of the face-vase figure, as in Figure 15a. Each processor may be in one of two states, "on" or "off," and initially states may be randomly assigned. There is only one consistency constraint: for local consistency, two neighboring processors may be in different states if and only if they are separated by a boundary line in the input image. It is

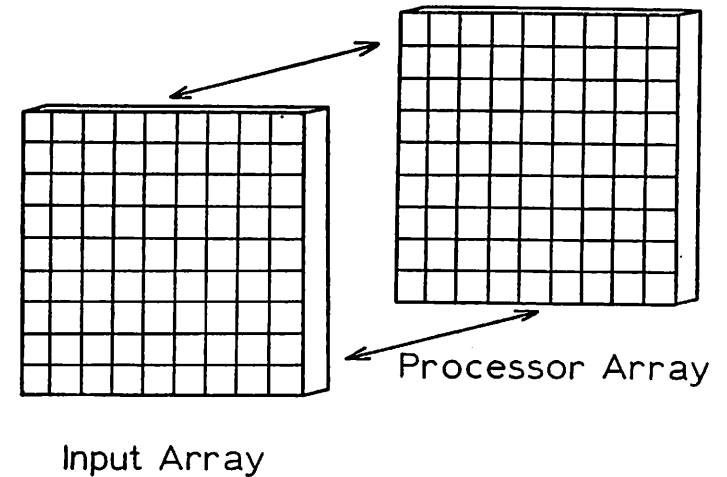


Figure 1:14

Here the visual system is modeled as a two array system. The first array contains a representation of a stimulus image in terms of elementary features. The second array is made up of locally interconnected processors. The internal state of each processor represents the perceptual state associated with a local region of the input image. The processor array is self-organizing.

clear that there are only two global organizations consistent with this local constraint, and these correspond to the two perceptual states of two faces or vase. A process capable of finding these global organizations might solve this rendition of the figure-ground problem without semantic information or intervention of face and vase detectors. Psychologists will point out that nonsense figures such as Figure 1:15b will be resolved into figure-ground relationships, which means semantic information cannot be critical to the resolution process. Of course recognition of objects as faces or vase must involve high level semantic information, but this may follow figure-ground separation.

Necker cube. Suppose that the input image depicts a Necker cube, Figure 1:15c. In this case, suppose the state of each processor is given by five numbers, A, B, C, D and E, where

A may be  $\begin{cases} 0 & \text{(interpret as a ground point)} \\ 1 & \text{(interpret as a line segment)} \end{cases}$

B may have any positive value, but  $B = \infty$  if  $A = 0$ .  
(interpret B as depth)

C, D, E is a vector of numbers, each having a value between -1 and +1 (interpret this vector as the orientation of a line segment in three-dimensional space).

Thus the state of a processor codes the depth and orientation of a line segment in space. The local consistency constraints are:

1. The state of a processor is consistent with the

stimulus input when its value of A is 1 if a line segment is projected to that processor, and 0 otherwise.

2. The states of neighboring processors in which  $A = 1$  are consistent if the line segment indicated by the state of the first meets the line segment indicated by the other at an angle of 90 or 180 degrees in three-dimensional space.

Again I think one can see that when these two local constraints are applied to the Necker cube image, two global organizations of the array will be allowed, organizations which correspond to the two perceptually stable states of the cube. And again, an organizing process capable of finding these global organizations would "solve" the Necker cube problem without reference to information contained in cube detectors.

To make this example more realistic, we might relax constraint number 2 somewhat to say that neighboring line segments should meet at 180 or roughly 90 degrees in three-dimensional space. This would reflect a tendency to see the angle between two lines as too large, if it is acute, or too small if it is obtuse. These perceptual tendencies compensate for distortions which naturally occur due to foreshortening when the lines bound a surface which is tilted away from the observer. Stimulus organization subject to this modified constraint would still tend to induce depth into the cube drawing, and drawings of other crystal shapes. This depth might not correspond exactly to a cube unless the



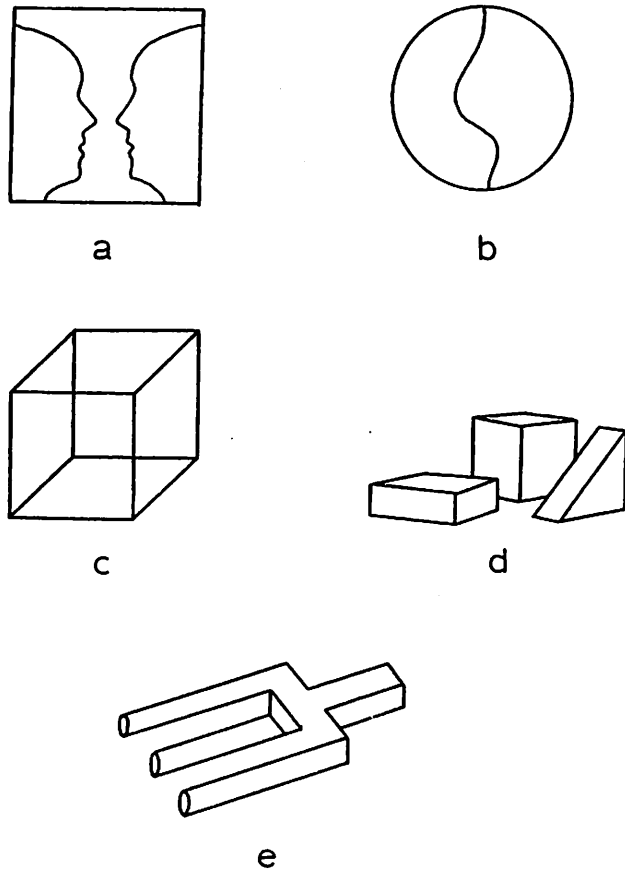


Figure 1:15

Images which can be "analyzed" by the processor array.

observer is familiar with cubes and the organization representation is "tuned" in accord with learned information.

In a similar way, one could define an array which interprets line drawings as solid objects, rather than "wire" figures as in the Necker cube example. This array could be applied to figures such as Guzman's stack of blocks, Figure 1:15d. The organizational principles embodied in this array would be very similar to those studied in detail by Waltz (1975). Note that when the array is presented with a paradoxical image, such as Figure 1:15e, there would be two conflicting partial organizations, but no global organization. (See Huffman, 1971, for a similar interpretation).

The final example has to do with perception of object motion. To introduce this example, consider what happens when an outline drawing of a box is moved in front of an observer whose visual perception of motion is based on the output of oriented moving bar detectors, Figure 1:16a. Such detectors respond when a properly oriented bar is moved through their receptive field. Within any small area of the visual field, there will be detectors for motion in all directions, but only those cells with motion specificity well matched to the direction of motion of a local stimulus will be activated. The receptive fields of two such detectors are shown in the figure. Notice that if the ends of a straight bar stimulus extend outside the receptive field of a detector, that detector cannot respond to the component of bar motion

which is parallel to the bar's orientation. Thus, the detected component of motion is that which is perpendicular to the bar's orientation, as is indicated by the short arrows in the drawing. In a natural visual system there might be detectors with much larger receptive fields, but these would not contribute significantly to perception of motion with this box image, since the outline contours would be very small compared to the size of the receptive fields. Also, if there were other differently moving stimulus points outside the box, this would interfere more frequently with large receptive field detectors than with small receptive field detectors. We may thus presume that the receptive fields of all elements which respond well to this moving image and contribute to the perception of a moving object, have receptive fields which are small compared to the dimensions of the box. It follows that the local motion stimulus will be for motion in different directions along differently oriented edges of the image, and neither of these detected motions matches the motion of the box. On the other hand, the observer actually perceives the box outline moving as a ridged unit, and he sees the area within the box outline moving with the contours. The question therefore is what neural mechanism integrates disparate, locally sensed motions into a single, perceived object motion?

We may restate this problem in the form of the previous examples. Referring again to the system composed of input

and processor arrays, suppose the input is the array of locally detected velocity vectors associated with small segments of the moving rectangle, as in Figure 1:16b. Now suppose that the state of each processor is expressed as the value of a two-dimensional vector  $V$ , to be interpreted as velocity, and a binary number,  $A$ , which is '1' if the input element corresponding to that processor contains a stimulus vector, and '0' otherwise. The local consistency constraints are that 1) neighboring elements with equal  $A$  values must have equal  $V$  values, 2) the  $V$  value of each processor in which  $A=1$  must equal the  $V$  value of at least one of its neighbors in which  $A=0$ , and 3) in processors in which  $A=1$ , the component of  $V$  which is parallel to the input velocity vector in the corresponding input cell must equal that vector in magnitude.

Global organizations satisfying these constraints are of three types: 1) All points within the rectangular region bounded by the stimulus may be represented as moving together to the right, while points in the exterior region are either stationary or moving together at some other velocity. 2) All points in the exterior region may be represented as moving together to the right while points in the interior region are stationary or moving at some other velocity. (In this case, the figure is seen as a hole in a moving surface). 3) All points in the whole array move at the same velocity to the right. (In this case, the figure is seen as surface features

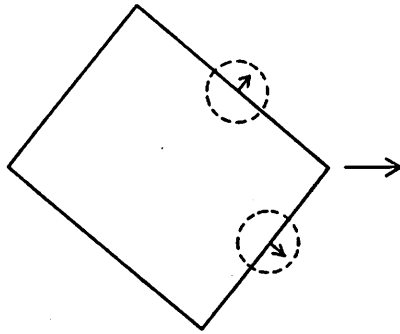


Figure 1:16a. A box-shaped image moves to the right at the velocity indicated by the large arrow. This generates local motion stimuli which are perpendicular to boundary orientation, as is shown in the dashed circles.

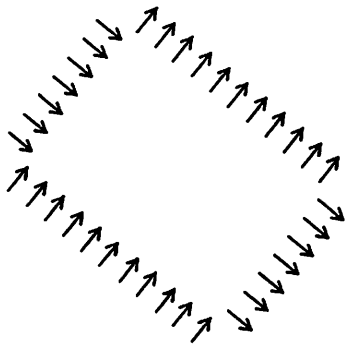


Figure 1:16b. This shows the patterns of local motion stimuli associated with the moving box image. These stimuli must be integrated in some way in order to obtain the perception of a single ridged image moving to the right.

rather than boundaries of the moving object).

To make this system more complete, the stimulus organization processes described here should be embedded in a structure which represents motion as a hierarchy of relative motions rather than as absolute motion on the retina. The hierarchical nature of motion perception has been studied extensively by Gunnar Johansson (see for example Johansson, 1975), while a neural mechanism capable of two level hierarchical representation has been proposed by Burt (1975). The latter representation separates object from observer components of retinal motion.

Any organizing process capable of finding these global organizations could operate at a low level of the visual system to segment the image on the basis of local motion cues. Since the best operational definition of a segment is an image area in which all stimulus points move together, as if attached to a single object, segmentation on the basis of motion should be a particularly valuable capacity for the visual system. Nothing has been said in this section about the organizing processes themselves, and that is the subject of the next section.

#### 1.6. Organizing Processes

The idea of an organization has now been proposed for two visual system functions. In the first, a spatial

component of the visual short term memory, or internal world model, has served to organize stimulus features and control their use by higher level object detectors. In the second case, the spatial short term memory has itself been viewed as a self-organizing pattern of neural activity, the organization being modified by the stimulus input even as it organizes the stimulus. It has been suggested that the SSTM controls the use of stimulus features by associating additional features with them, such as the distance and orientation in depth of the surface on which the stimulus features are assumed to lie. These added, non-stimulus features may be viewed as the state vector of local processors. But how do these processors become organized?

A process may be viewed as a sequence of steps, or a program, which a processor executes. The state of the process at a given time is its stage in the execution of these steps. A change of state or sequence of state changes may be triggered by the arrival of a new stimulus, and the new state which is entered is determined both by the previous state and by the input stimulus. In this way a given input may affect the process in different ways depending on the current processor state. On the other hand, a new stimulus frequently does not cause a change of state or alter a sequence of changes which is already in progress. This is the case when the input is consistent with the current processor

state. A processor state is to be interpreted as a perceptual state in some small region of the visual field. Stable perceptual states, which occur when the input is everywhere consistent with the perception, correspond to stable processor states.

The problem to be considered now relates not to single processes, which for the present purposes may be quite simple, but to arrays of processors which are locally interconnected and operate in parallel. It is assumed that all processors are identical and that each receives an external stimulus input as well as inputs from each of its neighboring processors. Following the definitions of the previous section, the array is said to be globally organized when each processor is in a stable state. The local consistency constraints used to define global organization in the previous section are now assumed to be built into the state change rules for each processor. Also implicit in these rules is a strategy designed to cause states to change from inconsistent to consistent, stable states.

We would like to know if there is any state change strategy which may be built into individual processors, so is only locally followed, but which will lead the array as a whole to a global organization. The difficulty in this problem arises from the fact that processors are coupled. When the array is in an unorganized state, some individual

processors may not be able to enter any state which is consistent with the current state of their neighbors, while other processors may be in locally consistent states which will not be consistent with any global organization.

Two strategies will be outlined here. The first of these is guaranteed to find a global solution of any organization problem if such a solution exists. This strategy is one which causes the array to cycle through all its states in a prescribed order until a global organization is found. The strategy may be built into the state change rules of individual processors, so that no external control is required. However the solution is uninteresting on two accounts. First, there are too many array states, so the strategy will take prohibitively long to execute. For example, if the array dimensions are  $N$  by  $N$ , and each processor may be in one of  $M$  states, then the number of array states is  $M^{N^2}$ . The second problem is that once the array is in a globally organized state, if there is a small change in the input pattern which requires a small local change in the array state, this strategy will cause the system to cycle through all array states in order to find a new global organization.

A second strategy may be proposed which is more heuristic in nature and which (in its present formulation) is not guaranteed to find a global organization. However, if it

does find such an organization, it is likely to find it much more quickly than the exhaustive search strategy. The first rule of this strategy (this rule will be qualified shortly) is that no processor should change states if it is in a state which is locally consistent. This rule is intuitively reasonable and is clearly appropriate when the array is in a state which is almost globally organized. The second rule of the strategy is that processors which are not in locally consistent states should sometimes change states, but not too frequently. This rule means that any processor which is not in a locally consistent state will eventually change state. However, a processor may enter a consistent state without changing states, if its neighboring processors change states appropriately. This is the reason why a processor should not immediately change states when it is in a locally inconsistent state: it first waits to see if the inconsistency is resolved by the neighbors.

The above strategy will be illustrated in computer simulations in the next section. There, individual processors will correspond to model neurons. The timing mechanism which regulates the rate of state changes is based on fatigue of active cells. Any state the processor (or cell) enters is semi-stable, and the processor will stay in that state until its fatigue level reaches a threshold, at which time it enters a new unfatigued state. On the other hand, if a state

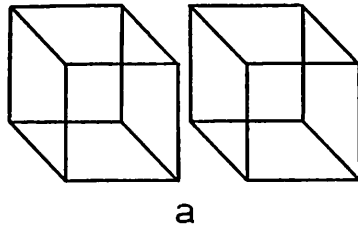
is locally consistent, the threshold is higher and never reached.

This strategy yields several interesting network properties which have psychophysical interpretations. If the system is initialized in a completely disorganized state, processors will begin to organize themselves locally, so that the array can be divided into many small regions, each of which is internally organized, while processors along the boundaries between regions are in states which are inconsistent with one region or the other. Some of these regions will then grow and spread over other regions as processors along the boundaries change allegiance. In this way, the number of subregions will gradually be decreased until only one remains and this will represent a global organization of the array. Thus one characteristic property of the system is that regions of organization spread across the array by recruitment along boundaries, just as crystals grow in a solution. This type of organizing process is implied in psychophysical phenomena where perceptual interpretations seem to spread across the spatial dimension. This seems to be true of stereopsis with complex stereograms: fusion is difficult to achieve at first, but when it is achieved in one region of the stereogram, it seems to spread gradually to neighboring regions. It also seems to be true of ambiguous images. In the case of the Necker cube, for example, it frequently seems

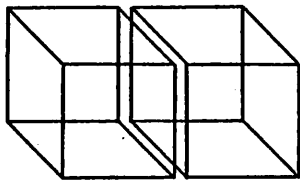
that a change of state is not abrupt but progressive, as part of the cube changes state first, and then other pieces which are neighbors of the part follow.

Another consequence of the local interactions postulated in the array is that an organization in one part of the array can affect organization in another only if the organization extends over the space in between. Thus, apparently long range interactions between distant processors are possible, but depend on the construction of a "bridge" of organization between the processors. This also has psychophysical interpretations. Suppose one views an image composed of several ambiguous figures, as in Figure 1:17. If locally connected organizing processes are responsible for the perceptual state of individual images, the state of one may be expected to determine the state of another if they are touching, as in Figure 1:17c, but not if they are separated by an "unbridgeable" space, as in Figure 1:17a. My subjective experience is consistent with these predictions.

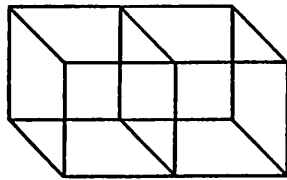
Another interesting property of the array results from the proposed timing mechanism. States which are not locally consistent have been described as semi-stable: they exist for awhile, but then change in accord with a fatigue process. States which are locally consistent may also be semi-stable, rather than stable, if parameters of the system are appropriately chosen, or if the stimulus input is "weak." Then proc-



a



b



c

Figure 1:17

This figure illustrates spatial interactions between two Necker cubes. The strength of the interaction depends on the separation of the cubes, and is of little importance in Figure a, but is sufficiently strong to always cause both cubes to be in the same state in Figure c.

processors which are in locally consistent states may also change state due to fatigue, but the time required for this is much longer than for states which are not locally consistent. When a processor which is in a consistent state does change state, the change of states spreads rapidly to neighboring processors, since these are also in fatigued states. Psychophysically, this leads to the phenomenon of spontaneous reversal of ambiguous figures. It should be noted that fatigue is postulated here not to explain such phenomena, but because it is a necessary part of the strategy for finding global organizations on the basis of strictly local interactions.

Finally, from a computational point of view, it is important to note that when information is represented in the form of an organized pattern of activity within a general purpose neural structure, rather than by unpatterned activity in specialized neural structures, the competition between alternative organizations has an inherently serial nature. Because any region of the array can be organized in only one way at a time, the existence of one organization completely excludes competitors within the region. Several competing organizations may coexist within the array, but these must be in spatially separate subregions. Within each region, the existing organization may match the input and satisfy local consistency constraints, while these constraints are violated along boundaries between regions. Competition will take

place along region boundaries as individual processors change allegiance from one organization to the other. (Consensus, of course, is reflected in the satisfaction of constraints within the boundaries). Each organized region will try to spread and overrun the others. If it happens that an established organization is inconsistent with the stimulus input at some point, the organization will break down at that point and another will begin to form. A point of mismatch between one organization and the stimulus becomes a "seed" for the formation of another organization.

On the other hand, before regions of organization become established in the array, processing in the array as a whole is parallel. In this condition, there are many small groups of processors which are independently trying to match the stimulus within corresponding small areas of the array (or visual field). Of course, each of these processor groups can be in only one local state at a time. Thus we may say that initial processing is locally serial but globally parallel.

Stimulus organization has been formulated here as a problem in constraint satisfaction. Procedures for obtaining solutions to such problems have been proposed by Waltz (1975) and Rosenfeld, et al. (1976) and have been discussed by Arbib (1975b). Several similarities and differences between these procedures and the procedure I have outlined should be noted. In the present terminology, Rosenfeld's relaxation procedure

finds a set of candidate states for each processor such that each candidate state of one processor is consistent with its stimulus input and at least one candidate state of each neighboring processor. On the basis of a pairwise examination of consistent processor states, the procedure eliminates many states which cannot possibly be part of a global array organization. However, the procedure does not actually find global organizations. For this, a systematic search through combinations of candidate states is proposed, which is based on a tree search procedure of Waltz.

The array self organization procedure which has been described here performs the same function as Waltz's tree search, but follows a rather different strategy. In initial stages of the search for a global organization the present scheme is parallel: processors become locally organized in many small regions of the array, then regions of organization compete and spread. The tree search scheme is essentially sequential, although many branches emanating from each node can be trimmed using the parallel relaxation procedure. The timing mechanism which controls the rate at which individual processors change states in the present procedure causes groups of processors to cycle through possible local organizations in a way which is roughly equivalent to following different branches in Waltz's tree. However, the present scheme does not have the degree of control that tree search



has, so while it may find a global solution faster in some cases due to its parallel nature, it is not guaranteed to find such a solution.

The importance of the relaxation procedure in reducing subsequent tree search depends on the nature of the constraints and the stimulus pattern. In examples such as the figure-ground separation (which will be simulated in the next section) relaxation would have no value since all processor states are candidate states. In other more complex cases, the procedure is quite effective. In any event, the procedure is essentially embodied in the present scheme if whenever a processor changes state, it changes to a state which is consistent with the current state of neighboring processors.

### 1.7. Computer Simulations

In this section, computer simulations of two systems are described which illustrate some of the ideas about organizing processes which were expressed in previous sections. The first example is a simple two element bistable system, which is of interest for two reasons. First, it constitutes a minimal model of rivalry and alternation between two competing perceptions, and can be used to demonstrate several interesting phenomena. For example, the dependence of the rates of alternation on stimulus strength with this model is in qualitative agreement with empirical data obtained by

Levelt (1965) for binocular rivalry. Second, this two element system will become the basic processing unit in a two-dimensional array of processors. The processor array system is also simulated to illustrate the transition from an initial unorganized state to global organization and spontaneous alternation and may be interpreted as an organization model for figure-ground separation and figure-ground reversal.

The two element system is shown in Figure 1:18.<sup>3</sup> We may suppose that activity in these cells represents two competing interpretations,  $P_1$  and  $P_2$ , of an image. These interpretations might be the alternative interpretations of an ambiguous image, or the two images of a rivalrous stereogram. Each cell receives an external input,  $x$ , which is proportional to the stimulus strength for its interpretation. These inputs are assumed to be constant in time in the present analysis. The outputs of the cells,  $y_1$  and  $y_2$ , give the "activity level" of each interpretation and this corresponds to the strength, or vividness of the perception. The values of  $y$  may vary in time, but are always positive. An

<sup>3</sup>The simple two element network with reciprocal inhibition and fatigue, which is described here, has been studied by Reiss (1962) and related networks have been studied by Wilton (1964). These authors are interested in rhythmic behavior of the net as a possible model for nervous control of stepping or flying activities. The analysis is repeated here, in somewhat different terms, because the application to perception is novel, and because an understanding of the two element net is important for understanding the organizing properties of the processor array.

interpretation is not perceived when the corresponding  $y$  is zero.

The two cells inhibit each other reciprocally, as shown. The cell output is a function of the difference between its stimulus input and inhibition from the other cell:

$$(Eq. 1) \quad y_i(t) = \begin{cases} (x_i - y_j(t)) \cdot G \cdot F_i(t), & \text{if } x_i \geq y_j(t) \\ 0, & \text{if } x_i < y_j(t), i \neq j. \end{cases}$$

Here the gain factor  $G$  is a constant and the same for both cells, while the fatigue of the cells is given by the variables  $F_1(t)$  and  $F_2(t)$ . It is assumed that  $F(t)$  decreases at a rate which is proportional to the activity level,  $y(t)$ , and that it simultaneously recovers at a rate which is proportional to the difference between the present state of fatigue and an unfatigued state where  $F=1$ . Thus the fatigue values are described by the differential equation:

$$(Eq. 2) \quad \frac{dF_i(t)}{dt} = \beta(1 - F_i(t)) - \alpha y_i(t).$$

Depending on the values of  $x_1$  and  $x_2$ , this system will show one of three behaviors: stable dominance, alternating dominance, or no dominance. Given appropriate  $x$  values, one cell will be dominant when its output,  $y_i$ , is greater than the

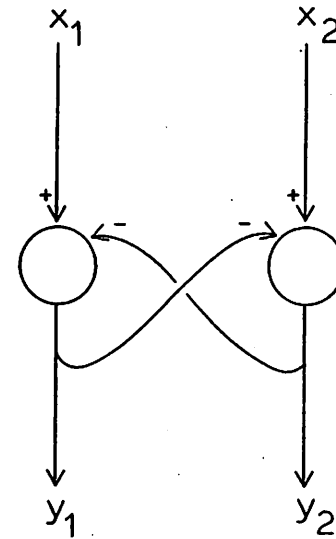


Figure 1:18

This two cell system may be used to illustrate three types of behavior: stable dominance, alternation and no dominance (see Figure 1:19). The output of each cell is equal to the sum of the inputs times a gain factor and a fatigue factor. The cells reciprocally inhibit one another.

input of the other,  $x_j$ , as this will mean the output of the non-dominant cell is zero. Once a cell is dominant, it will remain dominant until its  $y$  value drops below the other cell's input, due to gradual fatigue. When this happens, dominance quickly switches to the other cell, since it is less fatigued. The same sequence of events will lead to repeated alternation. For other values of  $x_1$  and  $x_2$ , dominance may not occur at all, and in still other cases, dominance will occur but will not alternate between cells.

In order to determine the conditions necessary for each of these three types of behavior, we may consider several special cases. First, suppose  $x_2 = 0$ , and at time  $t = 0$ ,  $F_1(0) = 1$ . Then  $y_1(0) = x_1 G$ . However  $y_1$  will decrease rapidly from this initial value due to decreases in  $F_1$ . The rate at which  $y_1$  decreases will itself decrease as  $y_1$  asymptotically approaches a minimal value  $\hat{y}_1$ , which is the level of cell activity at which fatigue and recovery from fatigue balance, so  $dF/dt = 0$ :

$$(Eq. 3) \quad \hat{y}_1 = \frac{\beta/\alpha \cdot x_1 G}{\beta/\alpha + x_1 G}$$

If we now suppose that  $x_2$  is not zero but that  $\hat{y}_1 > x_2$ , then it is clear that if cell 1 ever becomes dominant, it will remain dominant. Of course, it may be the case that

$\hat{y}_2 > x_1$  as well, in which case once either cell becomes dominant, its dominance will be stable.

Now suppose  $x_1 = x_2 = x$ . The condition for stable dominance is that  $x \leq \hat{y}_1, \hat{y}_2$ . Let  $\hat{x}$  be the largest value of  $x$  for which this is true, then  $\hat{x} = \hat{y}$  and from equation 3 we find:

$$(Eq. 4) \quad \hat{x} = \frac{\beta}{\alpha} \left(1 - \frac{1}{G}\right)$$

Thus when the two stimulus strengths are equal to  $x$ , and  $0 \leq x \leq \hat{x}$ , the system will exhibit stable dominance behavior.

Now we want to determine the conditions under which alternation may occur. In an alternation situation, one cell will be dominant, and then the other, in a regular repeating cycle. Also, there will be transition periods during which neither cell is dominant. These periods will generally be very short compared to time either cell is dominant during a cycle, and it will be convenient to ignore the transition periods in the following derivation.

A necessary condition for alternation is that a cell recover from fatigue during the time the other cell is dominant to the same level of excitability as it reached during the previous cycle. If we suppose that  $x_1 = x_2 = x$ , then the lengths of time each cell is dominant during a cycle will be equal. The average rate of recovery from fatigue during the

time a cell is suppressed must then equal the average rate at which it became fatigued while it was dominant. If the rate of recovery is less than the rate of fatigue, the length of time a cell is dominant each cycle will decrease until these rates become equal. If with arbitrarily short dominance periods the rate of recovery still is less than the rate of fatigue, then the cell will never become dominant. The limiting case for alternation is therefore a case in which the cycle time is very short. Under these conditions, the value of the dominant cell never differs significantly from the threshold value, and  $y \cong x$ . We may then compute the rate of fatigue during dominance,  $F_D$ , and the rate of recovery from fatigue,  $F_S$ , when the other cell is dominant, from equations 1 and 2:

When cell i is dominant:

$$y_i \cong x, y_j = 0$$

$$F_D = \frac{y_i}{Gx_i} = \frac{1}{G}$$

$$F_S = \rho \left(1 - \frac{1}{G}\right) - \alpha x$$

When cell i is suppressed:

$$y_i = 0, y_j \cong x$$

$$F_S \cong F_D \cong \frac{1}{G}$$

$$F_S \cong \rho \left(1 - \frac{1}{G}\right)$$

Let  $\hat{x}$  be the limiting value for alternation. Then  $x = \hat{x}$  when  $F_D = F_S$ , and

$$\hat{x} = 2 \frac{\rho}{\alpha} \left(1 - \frac{1}{G}\right).$$

This value for  $\hat{x}$  is approximate since it ignores the transition periods during which neither cell is dominant. However it is a useful approximation, as has been borne out in computer simulations. We conclude from the above derivation that when  $x_1 = x_2 = x$ , the dependence of system behavior on the value of  $x$  is as follows:

stable dominance when  $0 < x < \hat{x}$

alternation when  $\hat{x} < x < \hat{x}$

no dominance when  $\hat{x} < x$ .

These three behaviors are shown in Figures 1:19a, b and c.

In relating these results to the perception of rivalrous or ambiguous images, it is interesting to note that alternation occurs in an intermediate range of stimulus values. When the stimulus is weak, one or the other perceptual interpretations remains dominant, when the stimulus is strong, both are perceived together. I know of no evidence for stable dominance with weak stimuli, but this is an interesting possibility.

Of course, the strength of a stimulus in the case of an ambiguous image is not equivalent to image contrast, but depends on figural qualities of the image itself which are impossible to decipher. It may be possible to test this prediction with binocular rivalry where stimulus strength is related to image contrast.

I also know of no examples of ambiguous images where two different interpretations can be seen at once, as is predicted when the stimulus strengths are sufficiently large. Again it is difficult to judge which ambiguous images have large stimulus values. On the other hand, it is possible to make the stimulus strengths of the two images of a rivalrous stereogram very large by using high contrast and rapidly repeated presentation. When Kaufman (1963) presented a stereogram composed of orthogonal grids stroboscopically, both images were seen at once. Under normal viewing conditions, this type of stereogram is very rivalrous; there is a strong tendency for one image to dominate and for dominance to alternate between eyes.

Levelt (1965) has observed that the rate of alternation with rivalrous images increases as the stimulus strength of both images is increased. Levelt also found that if the stimulus strength of only one image is increased, then the length of time that image is dominant each alternation cycle remains unchanged, while the length of time the other image

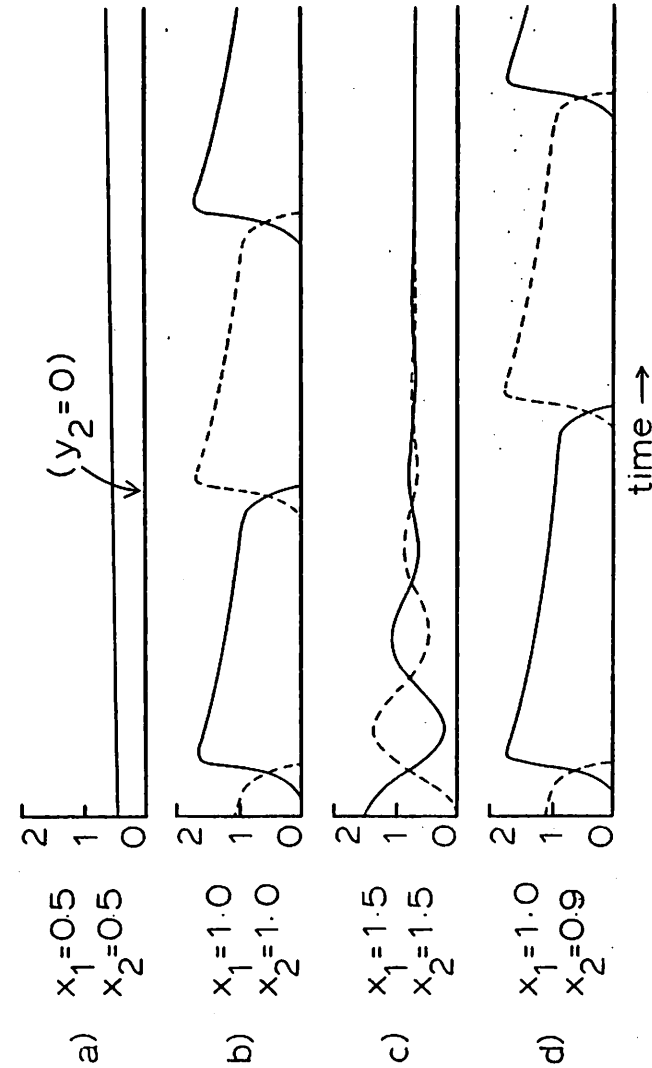


Figure 1.19. Cell outputs are shown as a function of time for four input values.  $y_1$  is solid line.  $y_2$  is dashed line. Here  $\beta/\alpha = 0.8$ , and  $G=6.0$ .

is dominant decreases. Both of these observations may also be made of the present system. Within the range of  $x$  values where alternation occurs, the rate of alternation increases with increased  $x$ . Figure 1:19d shows a simulation in which  $x_1=1$  and  $x_2=.9$ . When this is compared to Figure 1:19b, in which  $x_1=x_2=1.$ , we see that the effect of decreasing  $x_2$  is to increase the time cell 1 is dominant.

We may now construct a system, made up of cells like those described above, which will be self-organizing. The cells are arranged into two two-dimensional arrays, as shown (in one dimension) in Figure 1:20. Every cell has an external stimulus input,  $x_{i,j,k}$  where  $i$  and  $j$  are the coordinates of the cell in an array and  $k$  designates the array (1 or 2). In addition, corresponding cells of the two arrays reciprocally inhibit one another. Thus, pairs of corresponding cells form a subsystem which is similar to the two cell system studied above, except that there is spread of inhibition to neighboring cells. These inhibitory inputs are equal to the output,  $y_{i,j,k}$  of the inhibiting cell times a weighting factor, which is  $w$  for the neighbors of the corresponding cell, and  $1-4w$  for the corresponding cell itself,  $0 \leq w \leq 1/5$ . Thus the total weight given to the five inhibitory inputs of any cell is 1. The outputs and fatigue factors of each cell are given by:

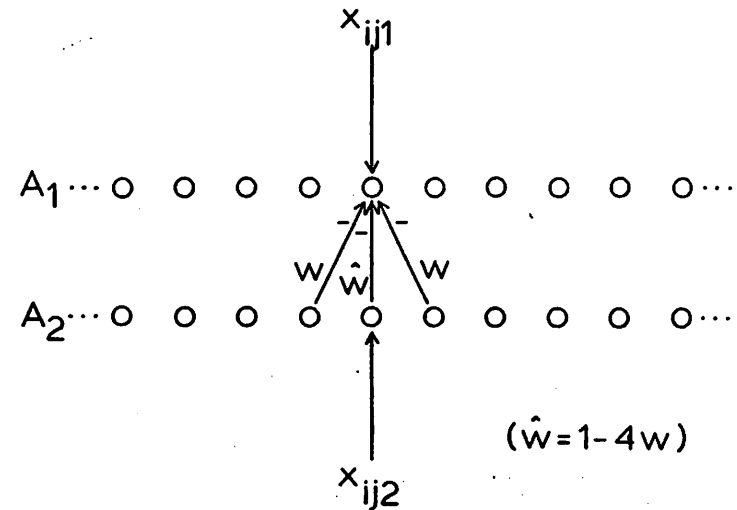


Figure 1:20

This figure shows the two array system of cells which may be used to illustrate self organizing processes. All inputs are shown for cell  $i,j$  of array 1. The same pattern of inputs is repeated for all cells in the system.

$$y_{i,j,k} = (x_{i,j,k} - (1-4w) \cdot y_{i,j,l} - w \cdot (y_{i+1,j,l} + y_{i,j+1,l} + y_{i-1,j,l} + y_{i,j-1,l})) \cdot G \cdot F_{i,j,k}$$

$$\frac{dF_{i,j,k}}{dt} = \beta(1 - F_{i,j,k}) - \alpha y_{i,j,k}$$

where it is understood that  $y$  and  $F$  are time dependent variables and that if  $k = 1, 2$ , then  $l = 2, 1$ .

It follows from these definitions that if at some point in time all the inputs to one array of cells equal  $x_1$ , and all inputs to the other equal  $x_2$ , and if the fatigue factors are the same within each array of cells, then the behavior of this array system is exactly the same as the behavior of the two cell system of Figure 1:18, so that the analysis of that system will serve as a lumped analysis of the present system.

Following definitions given in the last section, when a cell is in the same dominance state as its four neighbors, it is in a locally consistent state, and when all the cells in one array are dominant at the same time, then the array system is globally organized. Now it should be noticed that if a dominant cell is in a locally consistent state, the inhibition from cells of the other array will be zero, but if it is not in a locally consistent state inhibition will be non-zero.

Thus when a cell changes state, it will remain in the new state for a length of time which depends on the number of neighbors which are in the same state, so that cells in consistent states will remain in those states for a long time compared to cells in non-consistent states. With appropriately chosen values for  $x_1$  and  $x_2$ , cells in consistent states may be completely stable, so they only change states if one of their neighbors changes states first. These cell properties are the same as those described as appropriate for arrays of self-organizing processors in the previous section.

Two of these coupled array systems can be incorporated into a model for figure-ground separation in the following way. Suppose that dominant cells in one array,  $A_1$  of Figure 1:21, represent figure points, while dominant cells in a second array,  $A_2$ , represent ground points. The stimulus input to all these cells is the same. Cell states in the other two arrays code the relationship of boundary elements to neighboring area segments. These cells have zero stimulus input unless they occur in array positions which correspond to the positions of line elements in an input image. For simplicity, it is assumed that these line elements are all vertically oriented. (An additional array pair would be needed to code horizontally oriented line elements). A cell in array  $B_1$  is dominant if it has a stimulus input when the boundary element it codes is perceptually associated with the region to its

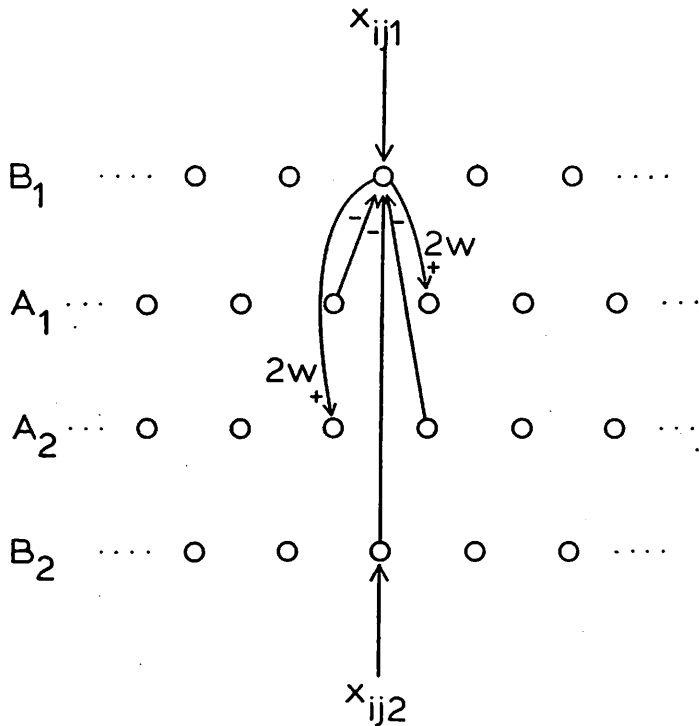


Figure 1:21

Two two-array systems like that which is shown in Figure 1:20, are coupled here to illustrate figure-ground separation and reversal. In addition to the interconnections between cells within the two array pairs shown in Figure 1:20, there are interconnections between cells of different pairs. These are shown here for one  $B_1$  cell.

right. A cell in  $B_2$  is dominant when the boundary element is perceptually associated with the region to the left. In either case, the boundary should be associated with the figure area, i.e. dominant regions in array  $A_1$ . In order to establish this association, the edge and area coding arrays must be coupled. The pattern of these interconnections is shown for several cells in the figure. A cell of layer  $B_1$  is reciprocally inhibited by the corresponding cell of  $B_2$  as well as by a cell to its right in layer  $A_2$  and a cell to its left in  $A_1$ . In return, the  $B_1$  layer cell excites a cell to its left in  $A_2$  and right in  $A_1$ . The weights on these excitatory projections are  $2w$ . Now suppose the  $A_1$  cells to the right of a dominant  $B_1$  cell are also dominant, while  $A_2$  cells are dominant to its left. Study of these interconnections will show that this is a locally stable state. The inhibition between cells in opposite states along a boundary is balanced by the facilitation of the boundary representing cell.

Now suppose the input stimulus for this system consists of two vertical lines, as in Figure 1:22a. The system will be globally organized when all  $A_1$  array cells within the central region defined by the input stimulus are in one state, either representing figure or ground, while all  $A_1$  cells in the remainder of the array are in the opposite state.

Figures 2:22b to f show a computer simulation of this system. The dimensions of the arrays are ten by ten. The



activity levels of  $A_1$  cells is shown as intensity in arrays of light spots at several moments in time. We should interpret array elements as representing figure points when the corresponding spot is fairly bright.

Figure 1:22b shows the randomly assigned states of these cells at an initial point in time. The remaining figures show array states at a sequence of later times. The number below each figure is the number of computer iterations and is proportional to elapsed time, which we may assume is in arbitrary units. We see that as time progresses, the array begins to become organized. Figure 1:22c shows the array completely organized with the central region represented as ground. The remaining figures show a sequence of states in a spontaneous figure-ground reversal. Thus several cells in Figure 1:22d are changing states, and these cells serve as the "seed" for a reorganization which can be seen spreading over the net in figures d to f. The reversal is nearly complete in Figure f. In this simulation, an additional 180 steps passed before there was another figure-ground reversal.

It should be mentioned that this system does not always converge on a global organization. In some cases, two competing regions of organization will continually spread over each other along different portions of their common boundary. Thus it is clear that the present strategy for obtaining global organization from strictly local interactions is not

guaranteed to find the global organization. This should not be considered a flaw in the system as proposed, since if these organizing processes were embedded in an actual visual system, they would be further constrained by interactions with many other perceptual processes, which could prevent unproductive, cyclic competition between organizations.

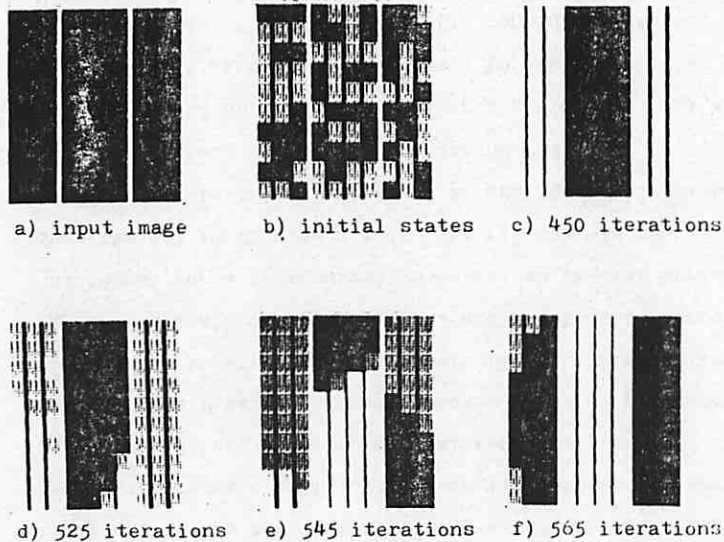


Figure 1:22

This is a simulation of net self organization and figure-ground separation and reversal. Figure a shows the input image, which is a simplified face-vase image where the two contours are vertical lines. Figure b shows array elements in their initial, randomly assigned, states. Figure c shows the array after 450 computer iterations when the image is first resolved into figure areas on the sides and ground in the middle. Figures d to f show figure-ground reversal. Bright areas correspond to figure.

CHAPTER II  
A MODEL FOR  
BINOCULAR FUSION AND RIVALRY

Introduction

When different images are presented to the two eyes, one's conscious experience is not of two distinct and separate images, but of a single image which may be identical to one of the monocular "half images," or may be a montage pieced together from parts of the two half images. Frequently, parts of the monocular images do not appear in the consciously perceived "combined image," and are said to be "suppressed." Phenomena associated with binocular combination and suppression have several interesting aspects. For example, what happens to visual information which is suppressed? Is that information lost or ignored by the visual system, or is it excluded from conscious perception, whatever that may be, but not from subconscious perception? What neural mechanisms mediate suppression, and what stimulus or visual system factors determine the stimulus points which will be suppressed?

These questions will be examined carefully in the first half of this chapter. However, my principal interest is not to study the suppression aspect of binocular combination for

its own sake, but to learn about neural structures and processes which are involved in another aspect of binocular vision, namely stereopsis. Stereopsis and suppression phenomena seem to be tightly coupled, and indeed, these phenomena may best be regarded as two aspects of a single perceptual process: suppression occurring when images presented to the two eyes are different, stereopsis, when they are nearly the same. Partial suppression also seems to play a role in stereopsis, but discussion of that close association of what otherwise seems to be complementary phenomena will be postponed until the next chapter.

A neural model for suppression phenomena will be developed in the second half of this chapter. Again my ultimate objective, in this and the next chapter, is to propose and motivate a neural model for stereopsis, but for several reasons it is appropriate to begin by developing a suppression model. First, the model for suppression is simpler; we do not need to be concerned yet about how the visual system handles small disparities between like image features presented to the two eyes, and the suppression model will become part of the stereopsis model. But it is also appropriate to study suppression before stereopsis because this study affords better insight into certain aspects of binocular image combination, such as the neural coding of image information at the level of the visual system where combination

occurs, and the interocular mechanisms which control the combination processes.

Binocular suppression may also be studied as a competition-cooperation phenomenon. When two unlike features appear in the same region of the visual field but in different eyes, they compete for dominance in the combined image, the non-dominant image being suppressed. Depending on the nature of the competing stimuli, dominance may be stable or it may alternate between eyes over time. In the latter case, the stimuli are said to be "rivalrous."<sup>1</sup> The cooperative nature of these phenomena is shown by the fact that regions of dominance (or, conversely, regions of suppression) tend to grow and spread in the visual field.

Processes responsible for binocular combination may be said to "organize" the stimulus pattern. In the case of stereopsis, these processes determine how individual stimulus points will be paired with points of the other eye. In the case of suppression, the processes do not so much determine how particular stimulus points will be perceptually interpreted, as what points will be available for interpretation.

---

<sup>1</sup>Properly used, the term "binocular rivalry" refers to continually changing patterns of ocular dominance. However, I shall frequently use this term more loosely; a rivalrous stereogram will be any which results in partial suppression of either half image when binocularly viewed, whether suppression is stable or unstable.

and what points will be suppressed.<sup>2</sup>

This chapter is divided into nine sections. In the first seven sections, I review psychophysical data relating to the suppression phenomena, including results of a number of original experiments. A set of hypotheses is proposed to characterize suppression phenomena, and each hypothesis is supported by psychophysical evidence. These hypotheses, which are summarized in Section 7, form the basis of the neural model for suppression which is developed in the remaining sections of the chapter. There, computer simulations of the model are described which show it to be in substantial agreement with the psychophysical data. Physiological data relevant to this and the subsequent stereopsis model are examined in Appendix A.

### 2.1. Introductory Examples

A few examples will be described in this section which will serve to introduce binocular suppression phenomena to the reader. The simplest example of the binocular combination

---

<sup>2</sup>The binocular combination system offers an ideal system for studying competition between alternative stimulus organizations. In other systems, competing organizations correspond to alternative perceptual interpretations of a single ambiguous image. It is difficult to study systematically the effects of changes in the stimulus pattern because such changes tend to affect all interpretations. In the case of binocular dominance and suppression, the competing images are separate and are presented to different eyes. One image can easily be changed independently of the other.

of dissimilar images is shown in Figure 2:1a. A contour image is presented to one eye, while a uniform field is presented to the other. The contour invariably appears in the combined image, as shown. If contours appear in both half images, but in separated regions of the visual field, then all contours appear in the combined image, as in Figure 2:1b.

When dissimilar contours are presented in the same region of the two half images, the combined image may be assembled from the half images in two ways. The first possibility is illustrated in Figure 2:1c, which is based on an example of Helmholtz (Helmholtz, 1962; Sperling, 1970). Here black, orthogonally oriented bars are presented to either eye. A cross is perceived in the combined image which has a curious pattern of gray level shading. The ends of the bars appear black, but towards the central square area where the bars cross, the black changes gradually to lighter shades of gray, so that the central square, which is black, is surrounded by a halo of white. Thus all contours which outlined the bars in the half images appear in the combined image as transitions between appropriate black and white regions. However, the contrast of these contours may decrease near the central square.

These examples, and particularly the stereogram of Figure 2:1c, suggest a couple of types of mechanisms which may underlie binocular combination. The first of these is a

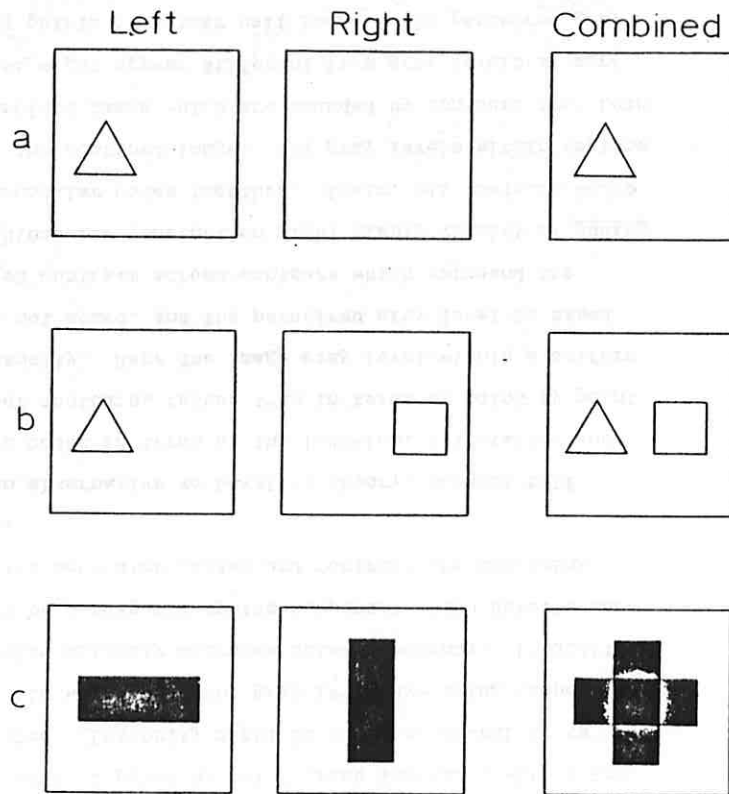


Figure 2:1

Three stereograms illustrating binocular combination without suppression of contours.

two component system proposed by Levelt (1965). One component of Levelt's system performs a weighted gray level averaging of the two half images to obtain the apparent gray level of the combined image. Thus for a particular point,  $\theta$ , in the visual field, at which the image brightnesses of the left and right half images are, respectively,  $B_L(\theta)$  and  $B_R(\theta)$ , the brightness of the combined image is

$$B_C(\theta) = w_L(\theta)B_L(\theta) + w_R(\theta)B_R(\theta).$$

The second component of Levelt's system is contour dependent and controls the values of the weights,  $w_L(\theta)$ ,  $w_R(\theta)$ , by the following rule: Suppose  $\theta$  is a point of the visual field which does not fall on a contour in either half image, and let  $d_L(\theta)$  and  $d_R(\theta)$  be the distances from the point  $\theta$  to the nearest contour in the left and right half images. Then the values given to the weights are such that  $w_L(\theta) + w_R(\theta) = 1$ , and the ratio  $w_L(\theta)/w_R(\theta)$  is monotonically decreasing with  $d_L(\theta)/d_R(\theta)$ . When  $\theta$  is equidistant from contours in each half image,  $w_L(\theta) = w_R(\theta) = .5$ , and when  $\theta$  falls on a contour in one half image, but not the other, the corresponding weight reaches its maximum value of one. Thus all contours from both half images will appear in the combined image, and points which fall near a contour in either half image will have about the same gray level in the combined

image. Combined images generated by this system are clearly consistent with the examples of Figure 2:1.

Levelt's theory implies that the two half images are coded in terms of point by point image intensity within the visual system. Intensity might be coded as neural spike frequency, in which case the gray level averaging component of the system actually averages spike frequency. In addition, there must be a separate system component which detects contours in the monocular images and controls the averaging processes.

As an alternative to Levelt's theory, suppose half images are coded in terms of the location, orientation and contrast of contours, rather than in terms of point by point image intensity. Here the image gray level within a uniform region is not coded, and the perceived gray level is based entirely on contrast across contours which surround the region. Binocular combination might simply consist of adding the two monocular codes together. Again, all contours would appear in the combined image. The gray levels within regions of the combined image which are bounded by contours from both half images might appear different from gray levels at corresponding points of either half image. The perceived gray level in these cases is based upon contrast over a different set of contours in the combined image than the monocular images.

This second theory provides a somewhat simpler explan-

ation of binocular combination and gray level averaging than does the first. However, according to this theory, none of the examples in Figure 2:1 involve binocular suppression: binocular combination is simply a matter of summing the monocular codes. Several stereograms will now be considered in which suppression clearly does occur. Neither of the theories as stated above can account for combined images obtained with these examples.

Several stereograms are shown in Figure 2:2 in which image contours are suppressed. Figure 2:2a, which is based on stereograms by Dodwell (1970), demonstrates that a single contour in the left half image can suppress a textured region in the right half image. This texture is itself composed of many contours. The stereogram of Figure 2:2b is similar to that of Figure 2:1c, except that the bars are narrower. This stereogram produces binocular rivalry: one's perception alternates between the two combined images shown. In each case, a portion of one or the other monocular bars is suppressed in the region where they overlap in the binocular view.

The stereogram of Figure 2:2c consists of two identical but orthogonally oriented grid patterns. When this stereogram is binocularly viewed, one may just see one half image, or he may see any combination of regions of one half image with non-overlapping regions of the other. Several examples are shown. The combined image is very unstable, so the

dominance pattern continually changes to produce an interesting and dynamic visual experience.

### 2.2. The Fusion Controversy

In the last section, I considered several examples of stereograms in which differences between images presented to the two eyes resulted in suppression of some image information in the combined view. The questions are these: how different must the half images be for suppression to occur, or must they be different at all?

When identical images are presented to the two eyes, only a single image is perceived. Single vision is also obtained when the monocular images are identical except for some slight disparities which may lead to a perception of depth. The binocular image has very nearly the same "brightness" as either half image in these cases. This means that doubling the stimulus energy by presenting it to two eyes rather than one, does not result in a noticeable change in the brightness or contrast of the single combined image.

Two opposing theories have been proposed to account for binocular "single vision" in an historically important debate which remains unresolved today. The first of these is the so-called "fusion" theory (Boring, 1933; Dodwell and Engle, 1963; Sperling, 1970; Julesz, 1971 and others). According to this theory, when very similar image features occur in

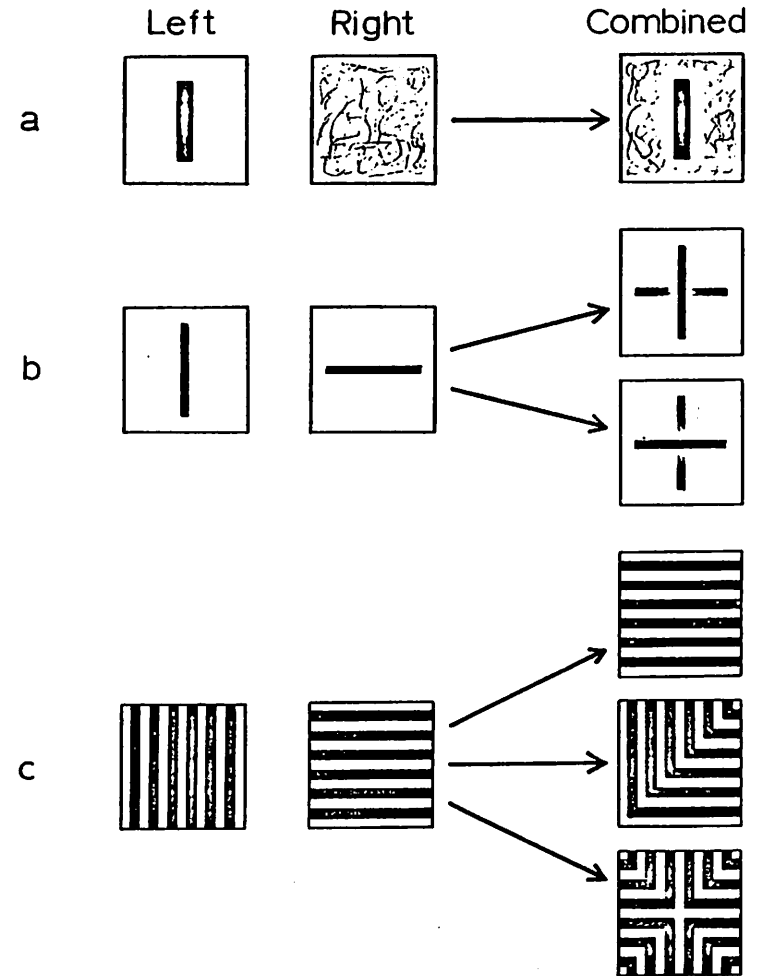


Figure 2:2

Stereograms illustrating suppression of contours and rivalry.

nearly the same positions of the two half images, the corresponding feature codes "fuse" somehow in the visual system to form a single binocular code. This state of "sensory fusion" is distinct from states of rivalry and suppression in that the two half images contribute in the same way to the binocularly combined image.

According to the alternative "suppression" theory (Asher, 1953; Hochberg, 1964; Levelt, 1965 and others), the visual processing of similar images is the same as processing of dissimilar images: suppression occurs in both cases and it is suppression which accounts for singleness of vision when images are alike. One is not aware of suppression when the images are alike, since the dominant information is exactly equivalent to the suppressed information.

Needless to say, fusion and suppression theories explain binocular combination, and particularly stereopsis, in different ways. In this section I shall review the principal arguments which have been advanced in favor of both theories and propose a new, and I think quite strong, argument in support of the fusion theory.

According to the suppression theory, any stimulus presented to one eye causes inhibition of all stimuli falling on corresponding and nearby points in the other eye. The strength of the inhibition depends on the character of the stimulus. For example, a contour causes stronger inhibition

than a uniform field, and a long, high-contrast contour causes stronger inhibition than a short, low-contrast contour (Wilde, 1938, summarized in Levelt, 1965; Crovitz and Lockhead, 1967). Of two images falling on corresponding points of the two eyes, the one generating greater inhibition will dominate, while the other will be suppressed. When the images generate roughly equal inhibition, one will still gain an upper hand and dominate, but dominance will tend to alternate between the two, over time.

The suppression theory is attractive for several reasons. First, from a theoretical point of view, the neural mechanisms required for binocular combination, according to this view, are far simpler than those implied by the fusion theory. The fusion theory requires that the two monocular images be partially analyzed in the visual system prior to binocular combination so a decision can be made as to whether particular image points should be fused or suppressed. In the suppression theory, dominance depends only on relative stimulus strength: no comparison need be made between stimulus patterns themselves.

Experimental evidence in favor of suppression is also much easier to obtain than is evidence for fusion. Suppression may be readily and unequivocally demonstrated with stereograms made of rivalrous half images. Fusion, on the other hand, can not be directly demonstrated. Fusion is presumed



to occur only with identical images, and, as noted previously, it is not possible to determine from direct observation whether a single combined image is composed of one or both of two identical half images. Any argument in favor of either suppression or fusion of identical half images must be based on indirect evidence. Three types of evidence will be considered here.

Consider first an argument in favor of the suppression theory of single vision. It has been observed that if different stimuli are presented to the two eyes, the strength of mutual inhibition is greatest when the images are presented to corresponding points of the two eyes, and the strength decreases rapidly as the stimuli are moved apart. (This inverse relationship between strength of inhibition and distance is implied, for example, by the experiment of Kaufman, which will be described below). Furthermore, even similarly oriented contours presented in slightly different positions of the two half images may partially suppress one another. (See below for Hochberg's evidence). If slightly disparate but similar features tend to suppress one another, just as do non-similar features, should we not expect that tendency to increase as the disparity is decreased between similar images as it does with non-similar images? Complete suppression and single vision should be anticipated when similar features are sufficiently close together. According to

suppression theory, there is a critical range within which complete suppression occurs even with similar features. This critical range corresponds to the so-called Panum's "fusional" area.

Levelt (1965) and Kaufman (1974) have pointed out that support for this interpretation of Panum's fusional area comes from an experiment by Kaufman (1963), in which he measured the spread of suppression around a line contour. Kaufman presented subjects with the stereogram shown in Figure 2:3a, so that one eye saw two vertical lines while the other saw a single, horizontal line. When binocularly viewed, the lines bisect one another, and if the separation of the vertical lines is sufficiently small, the section of horizontal line which falls between the vertical lines in the combined view will tend to be suppressed. Thus by varying the separation of these lines, a maximum range of suppression could be determined. This was found to be about 14 minutes of arc. When the experiment was repeated with two horizontal lines and one vertical, the estimate obtained for maximum vertical separation was about half this value or 7 minutes of arc. Two contours contribute to the suppression effect in this experiment. If we suppose each contour accounts for half the suppression at maximum separation, then the range of suppression associated with a single contour becomes 7' horizontally and  $3\frac{1}{2}'$  vertically. The interesting observation

is that these ranges are the same as the dimensions of Panum's fusional area (Ogle, 1950).

The above argument must be re-evaluated in the light of the recent discovery by Fender and Julesz (1967) that the dimensions of the fusional area depend on the nature of the stimulus pattern, and for appropriately presented random dot patterns, "fusion" can be maintained with up to two degrees disparity. The range of suppression around a contour also depends on factors such as contrast (Crovitiz & Lockhead, 1967) and bar width. I have found that I can get complete suppression of one grid by another, orthogonal grid (as in Figure 2:2c), even when the width of the bars is 2 degrees. In view of this variability of range of fusion and suppression, it seems that it may only be fortuitous that suppression ranges obtained by Kaufman match the classical dimensions of Panum's fusional area. In any event, it has yet to be shown that stimulus factors which affect the spread of suppression also affect the size of the fusional area in the same way, or vice versa.

One other point should be made with respect to the above argument in favor of the suppression theory. The argument was based, in part, on an observation by Hochberg that non-intersecting, similarly oriented contours, which are presented with slight displacement to the two eyes, partially suppress one another. The stereogram he used to demonstrate this

point is shown in Figure 2:3b. Each half image is composed of three horizontal bars which are arranged so that the lower two superimpose in the binocular view, while the upper bars from each half image appear separately, one above the other. The upper bars tend to suppress one another, as indicated. While this shows mutual inhibition of non-intersecting contours, i.e. inhibition at a distance, it is not clear that it illustrates mutual inhibition of similar features. In fact, the suppressed zones occur along edges, suggesting that it is the edges which are antagonistic. While these edges are alike in orientation and nearby in position, they are very different features, since they represent transitions from black to white which occur in opposite directions.

Arguments in favor of the fusion theory frequently make use of the phenomenon of stereopsis. Supporters of this theory argue that stereoscopic depth could not be computed by the visual system if one or the other monocular images has been suppressed. On the other hand, suppression theorists suggest that stereopsis and suppression take place in separate visual channels, so that information which appears lost in the consciously experienced binocular image is not, in fact, lost in the channel which computes depth. Proponents of both views have tried to prove their points by devising stereograms which show the presence or absence of stereopsis in rivalrous stereograms.

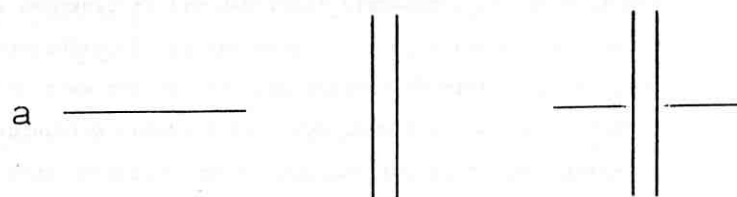


Figure 2:3

Stereograms used in experiments by Kaufman and Hochberg.

A very interesting example of a stereogram combining rivalry and depth was devised by Helmholtz (1962). Starting with an ordinary stereogram, which depicted an outline drawing of a geometric figure in depth, he replaced one half image with its negative (white points are made black and black, white). Such a stereogram seems to be rivalrous everywhere, and yet depth is perceived. A simple example of this type of stereogram is shown in Figure 2:4a.

Kaufman also has constructed a stereogram (similar to the one shown in Figure 2:4b), in which a central square-shaped region may be seen in depth, despite the fact that both this region and its surround are represented by orthogonal, rivalrous grid patterns. Again, rivalry does not prevent stereopsis.

On the other hand, Sperling (1970) describes a stereogram (which was originated by Kaufman and Pitblado) in which suppression of one half image does prevent the perception of depth. (See Hochberg, 1964b for other examples). Sperling's stereogram (similar to Figure 2:4c) consists of two concentric circles in each half image, superimposed on a grid pattern. One circle is slightly displaced in one half image relative to the other so that when the circles are binocularly combined, they appear at different depths. The grid pattern in one half image is orthogonal to that of the other, so that the stereogram is rivalrous. The binocularly combined image

is usually composed of subregions of either half image, but the pattern of dominance continually changes. The circles appear in depth as long as the combined image is composed of parts of both half images. However, occasionally one half image will be entirely dominant, while the other is completely suppressed. When this occurs, stereopsis is lost.

It is my opinion that failure of stereopsis in stereogram 2:4c is strong evidence in support of the fusion theory, while the co-existence of rivalry and stereopsis in the other two examples is less compelling evidence against that theory. In fact, I think the perception of depth in these examples can be explained by the fusion theory in the following ways.

First we consider the possibility of fusion in the negative correlation stereogram of Figure 2:4a. Helmholtz suggested that this stereogram supports the hypothesis that images are coded by boundaries and it is these boundaries that are associated with one another when one perceives depth. This explanation will work for Figure 2:4a, if we assume that the boundary code does not include information about the direction of intensity change across the boundary. However, the explanation does not account for failure of stereopsis in two related stereograms. If one half image of a random dot stereogram is replaced by its negative, stereopsis is not possible (Julesz, 1960). If the disparity of the stereogram is represented by displaced regions of uniform intensity, as in Figure 2:4d, rather than by displaced contours of an outline

drawing, as in Figure 2:4a, then stereopsis again is impossible. (This stereogram is based on examples by Treisman, 1962).

There is a better explanation of depth perception in negatively correlated stereograms. The above examples suggest that stereopsis results only when the stereo image is represented by an outline drawing. Thus a black line on a white ground must be perceptually associated with a corresponding white line on a black ground, in the other eye. Consider a case in which these lines are vertically oriented and one is displaced horizontally with respect to the other. Each line may be analyzed into two edge contours which represent the boundaries between white and black regions. Suppose now that the line in the left half image is black against a white ground. Thus its left edge is a boundary between a white region on its left, and a black region on its right. The same is true for the right edge of the corresponding white line in the right half image. That these two contours are identical and may fuse in the normal way, provided the disparity between them is not too great. This is the case in figure 2:4a, but not in Figure 2:4d. My subjective experience with the stereogram in Figure 2:4a is that depth is perceived, not when corresponding black and white lines appear superimposed, but when they appear side by side, consistent with this explanation.

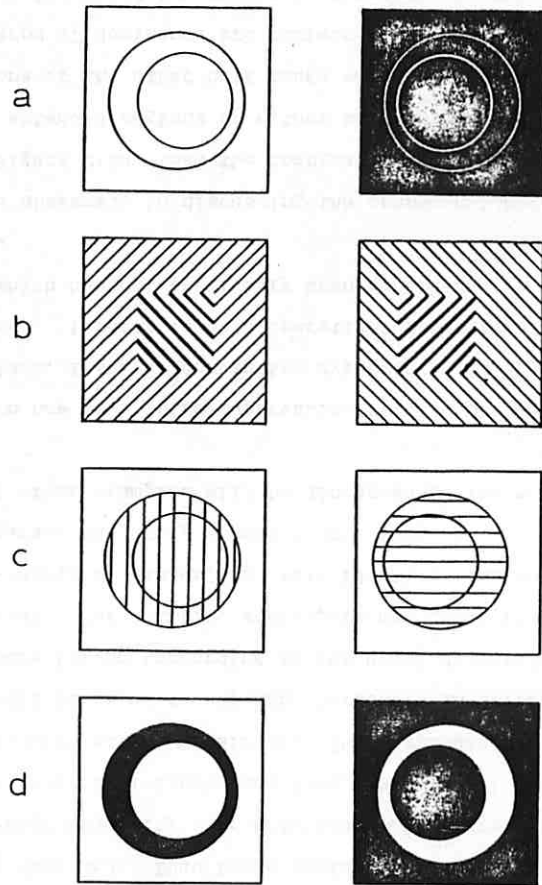


Figure 2:4

These stereograms incorporate rivalrous and stereopsis stimuli. Depth and rivalry seem to coexist in a and b, but not in c and d.

An explanation of depth perception with the stereogram of Figure 2:4b in terms of a fusion theory will follow from an extension of what we mean by the term "fusion." This extended concept of fusion will be briefly described here, and more carefully defined in Section 6 of this chapter and in the next chapter on stereopsis.

Let us suppose that images are coded in terms of local features (which need not be defined here) and that many of these code elements differ only in scale: some will be small (high resolution) and others will represent the same image feature, but at a larger scale (low resolution). A given image pattern will be coded by features of all sizes. We may say that total fusion occurs when the codes for two images are identical and corresponding elements of each code become somehow associated. On the other hand, partial fusion occurs when some, but not all, code elements match and are associated with one another. It now follows that two types of partial fusion may occur: one in which total fusion occurs over subregions of the two images, and the other in which one group of features fuse within a given image region but not another group. The second type of partial fusion occurs, for example, when large scale features fuse while small scale features do not. We assume that partial fusion is sufficient for stereopsis.

Partial fusion may occur in the stereogram of Figure 2:4b

because the inner square region is drawn with darker lines than is the surround. Thus large scale features which respond to average intensity over extended regions, but which do not resolve individual lines, may fuse, while small scale features are unfused and rivalrous. This expanded definition of fusion will allow us to explain stereopsis in several other examples where fusion (according to the usual definition) does not occur. For example, stereopsis may occur with disparate non-identical images, and with identical images which are so disparate that they appear double (diplopia). These and several other examples will be discussed in the next chapter.

We turn now to a third suppression related phenomenon which provides, I think, convincing evidence in favor of the fusion theory. I refer to a "cooperative" property of suppression, which has not previously been considered in the literature.

It was observed, in discussing the orthogonal grid stereogram of Figure 2:2c, that the combined image tended to be made up of extended regions of either half image, and on occasion, one or the other half image would dominate completely. These patterns of dominance are curious in view of the dominance pattern described for the crossed bar stereogram of Figure 2:1c. The crossed grid stimulus is like a repeated pattern of crossed bar stimuli, so one might expect that the

resulting combined image would be like that for the crossed bars, but repeated, as shown in Figure 2:5. Local patches of dominance are not observed around each contour as suggested here. Instead, patches of dominance extend over much larger regions. This observation suggests that dominance over one region of the visual field tends to spread to nearby regions by way of some sort of lateral interactions. This effect can cause dominance by one eye to spread over regions where the stimulus would otherwise favor dominance by the opposite eye. A possible neural mechanism for suppression which would account for the region growing behavior is shown in Figure 2:6. Image information is represented by activity in separate sets of cells for the two eyes. Each cell inhibits activity in the corresponding cell for the opposite eye. This inhibition is recurrent and spreads laterally to nearby cells. Thus when one cell becomes dominant, it reduces the output of the corresponding cell in the opposite eye to zero. This in turn means a release of inhibition to the dominant cell and to cells which neighbor that cell. The neighbors will then have a greater chance of becoming dominant as well.

The critical point to be made here is that, if single cells of vision are to be explained by suppression, then this phenomenon of spread of dominance should occur when the two images of a stereopair are identical, just as it occurs when they are very different. This property can be tested experimentally.

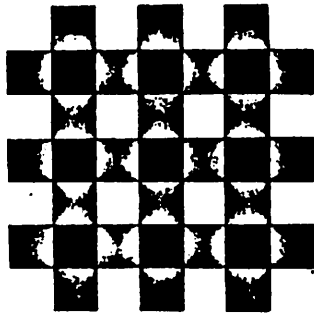


Figure 2.5

The combined image with orthogonal grids which would be anticipated from the combined image with orthogonal bars, as in figure 2.2c.

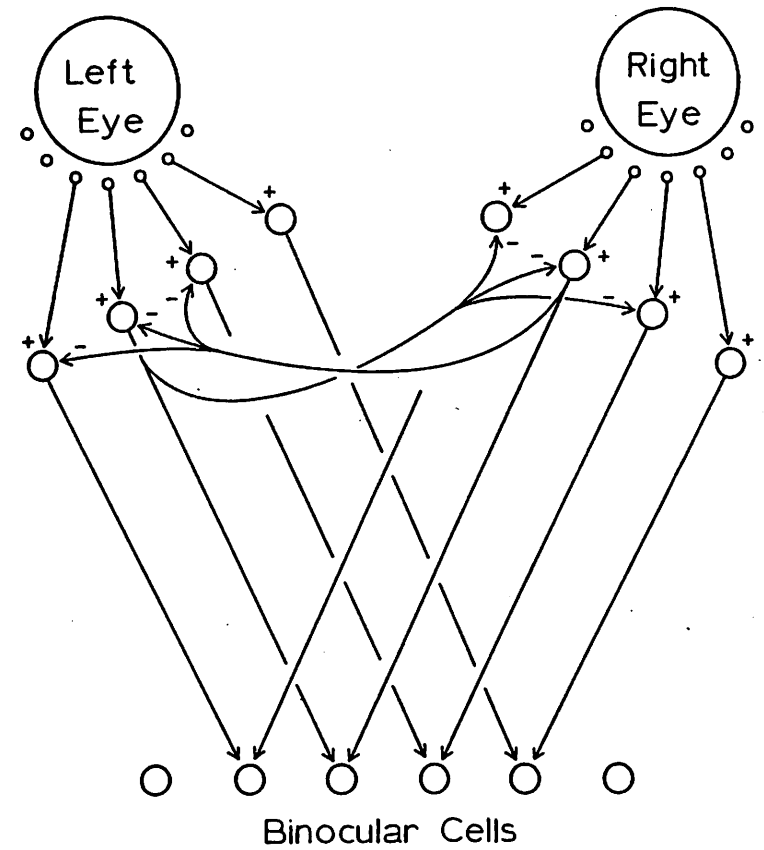


Figure 2.6

Spreading recurrent inhibition between monocular cells may account for spread of ocular dominance.

Consider the stereogram of Figure 2:7a, in which the half images are different (uncorrelated) random dot patterns with a dot density of about .5. When uncorrelated random dot patterns such as these are binocularly superimposed, the resulting combined image seems to have the same dot density as the half images (Tyler, 1975). If we accept a fusion hypothesis, about half the dots would be expected to fuse, and of the rest, 50% must be suppressed. According to the suppression hypothesis, half the total number of dots must be suppressed. Now suppose an extra dot is added to both half images so that these dots are smaller than the dots of the random patterns, and are positioned so that they should appear near one another, but not superimposed in the combined image. These marks provide easily discriminated "monocular flags" and may be used to detect regions of suppression. When the stereogram is binocularly viewed, there is a strong tendency for one or the other monocular flag not to appear in the combined image. Furthermore, if several such flags are added to each half image, there will be a strong tendency for all flags from one eye to be suppressed while all flags from the other remain visible in the combined image. These facts are consistent with the observation that eye dominance tends to spread over extended regions of the visual field. Where the principal dots which make up the random dot pattern are suppressed, the monocular flag points are suppressed as well.

Now suppose we repeat this experiment with identical (correlated) random dot patterns, Figure 2:7b. Again, if suppression occurs, we expect it will occur in extended patches and monocular flag points will be suppressed as well. However, I observe that whenever the two images appear to be "fused," all monocular flags from both half images are clearly visible. It follows that binocular combination of correlated images is qualitatively quite different from the binocular combination of uncorrelated images and this difference corresponds to the occurrence of fusion in the first case and suppression in the second.

The weight of the evidence in the above examples is in favor of the fusion, rather than suppression theory of binocular single vision, and I shall assume for the remainder of the discussion that fusion does occur as a distinct perceptual state. It remains for us to characterize this fused state, and to suggest neural mechanisms which might be responsible for determining which stimuli should be fused and which should be rivalrous.

### 2.3. The Architecture of Binocular Interaction and Suppression

I shall be concerned, in the next several sections, with neural structures and pathways which may mediate binocular interactions and suppression. That is, I shall try to deter-



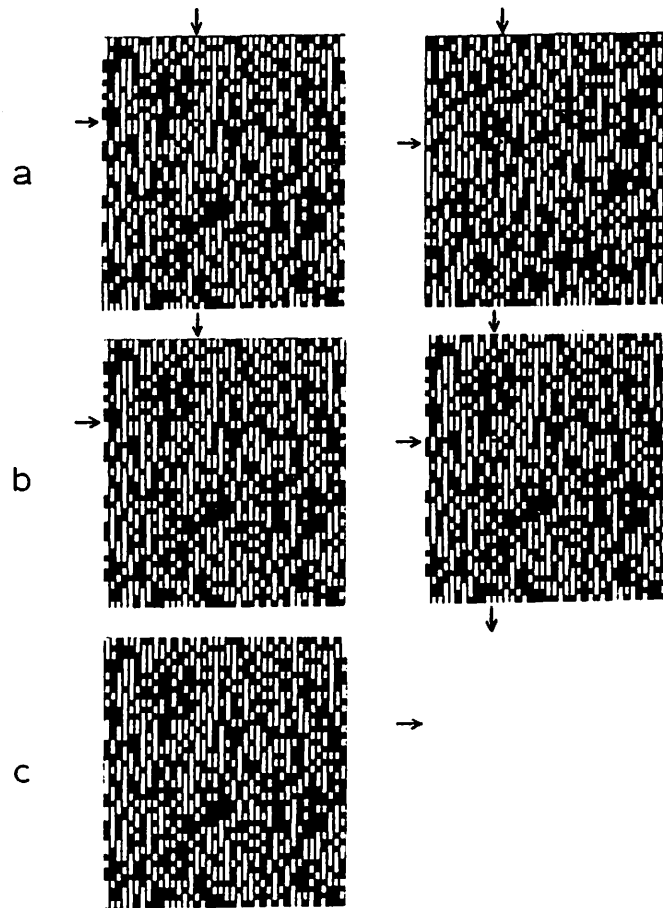


Figure 2:7

Fusion and rivalry are shown to be distinct perceptual states in the uncorrelated and correlated random dot stereograms a and b. In c the ocular dominance is shown to depend on context. Arrows show the positions of monocular flags.

mine whether the processing responsible for these functions resides in the lateral geniculate body, the visual cortex, or some structure involved in later stages of visual processing. And I shall try to determine whether control of this processing is afferent, efferent, or intrinsic to the neural structure. In order that this study may seem reasonably organized and systematic, I begin in this section by presenting a very schematic flow diagram for binocular information processing within the visual system, and identify the various stages at which the binocular interactions of interest might take place. Then, in subsequent sections, I will consider each of these possible stages in turn, in order to eliminate those which seem inconsistent with available psychophysical data. Related physiological are considered in Appendix A.

Figure 2:8 shows a flow diagram for processing within the visual system. Processing is divided into three stages: A) monocular, B) low level binocular, and C) high level binocular. Principal projection (afferent) pathways are shown by solid arrows, while possible efferent and intrinsic control pathways are shown as dashed lines.

It is assumed that in box A all information from the two eyes is represented and processed separately. Anatomical structures of the geniculo-striate visual system which are subsumed by this box include the LGB and monocular cells of visual cortex. It is further assumed that these cells have

center-surround receptive fields as in the LGB. The interocular control pathway, P1, might be mediated by interneurons which cross laminar boundaries in the LGB, while the monocular control pathways, P2, are mediated by interneurons within single LGB lamina.

Binocular combination of afferent information from the two eyes takes place in box B. I include in this box those cells of the visual system which receive afferent stimulation from both eyes, i.e. the binocular cells of visual cortex. To make this idea precise, assume: 1. individual cells in box B code the same type of information for both eyes, so that information indicating the eye of origin of a stimulus is lost at this stage, and 2. cells in this box have "simple" or "complex" receptive fields, as defined by Hubel and Wiesel (1962), so may be said to code the image in terms of "elementary" line shaped features. Interneurons in visual cortex which modulate activity of the box B principal cells are represented by control pathway P4.

Box C represents the remainder of the visual system, or all visual processing after binocular combination. This box is relevant to the present investigation of mechanisms of binocular combination principally because there is a possibility that combination is under efferent control, via P5. Thus, semantic information may play a role in determining regions of dominance and suppression by way of this pathway. Similarly, binocular combination may be controlled by Box B

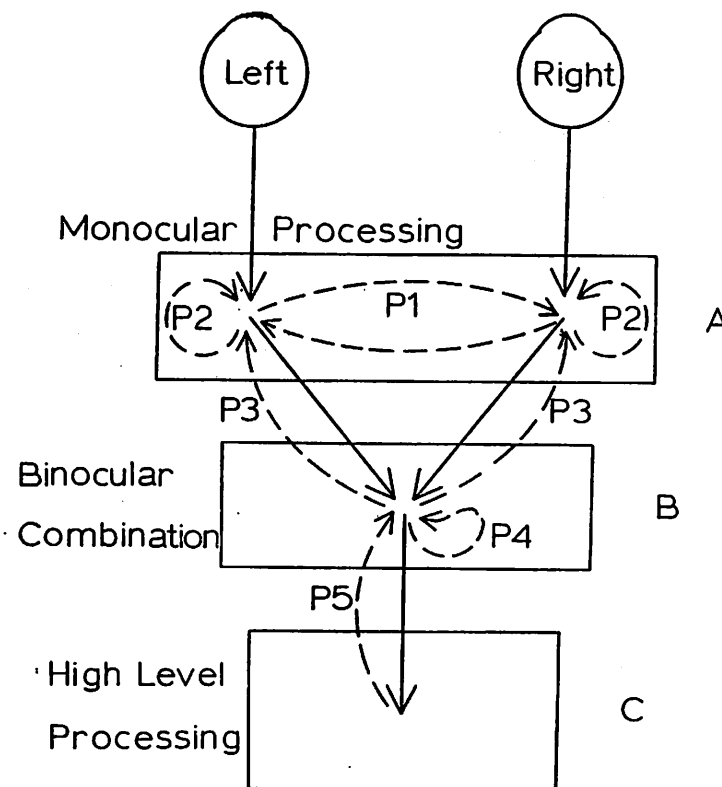


Figure 2.8

In this flow diagram for binocular combination, projection (afferent) pathways are shown as solid lines and control (efferent and intrinsic) pathways as dashed lines.

indirectly through Box A by way of efferent pathway P3.

With each control path, we may associate a possible model system or "architecture" for binocular suppression. There are two places where suppression might occur - in boxes A or B, and for each of these, there are two possible control paths, one intrinsic and one efferent. In the subsequent discussion, I will evaluate these four possible architectures in the light of experimental evidence. Here I will mention the principal functional differences between architectures and formulate the critical questions which may guide evaluation of each model.

If suppression takes place in box A under intrinsic control (P1,P2), then suppression must correspond to the inhibition of monocular afferent activity from one eye by afferent activity from the opposite eye. A possible neural structure for this interaction has already been proposed, Figure 2:6. The critical question with this architecture is: can it account for both fusion and suppression? If monocular codes are center-surround, as assumed, is it possible to determine which image features in the two eyes are sufficiently alike to warrant fusion or sufficiently different to warrant suppression?

This difficulty is avoided in architecture 2: suppression in box A under efferent control via P3. Here it is supposed that image comparison is accomplished in box B on the basis

of low level feature analysis of the images. However there is a potential difficulty with a system in which information suppression occurs a stage prior to image comparison: once a rivalrous stimulus is detected and suppression is accomplished via an efferent signal, information getting to box B will no longer be rivalrous - and mismatch information needed for directing suppression will be lost. The efferent control scheme could work if there exist two processing channels which operate in parallel. One of these channels would be the primary projection pathway to higher visual centers, while the second channel would control binocular combination in the first. Both channels initially carry the same afferent information. But when a mismatch is detected in the control channel, at the level of box B, then the suppression may occur in the primary channel, at the level of box A, under efferent control. Thus, the mismatched information, which is needed to maintain suppression, is retained in the control channel. (These two channels might correspond, for example, to the X and Y projection systems described by Enroth-Cugell and Robson, 1966).

In the third architecture for binocular combination, suppression occurs in box B, under intrinsic control via pathway P4. Since it is assumed that information about the eye from which a stimulus originates is lost at this stage of processing, the inhibition responsible for suppression

must be directed between rivalrous features after binocular combination. A potential difficulty here will be to explain why extended regions of suppression should occur in which only features seen by one eye are suppressed.

In the fourth architecture, suppression occurs in box B under efferent control from higher visual processes, via pathway P5. The decision about which features will dominate, and which will be suppressed will depend upon semantic factors in this case: a feature will be suppressed if it does not make sense in the context of other features, and does not contribute to the recognition of objects in the current visual scene. Again, there is the problem associated with efferent control mentioned with respect to architecture 2, so it must be assumed that separate primary information and control channels exist. In this case, we might suppose that information in the primary channel corresponds to that information which is consciously perceived. Thus, information which is suppressed in terms of conscious perception (this is how suppression is defined in the first place) still exists in the visual system and is accessible to other "subconscious" visual processes, including those which control binocular combination. To evaluate this architecture, we need to determine whether suppressed information is actually lost and unavailable to all high level visual processes, or whether it is only inaccessible to conscious inspection.

#### 2.4. Is Information Lost?

Helmholz believed that afferent information from the two eyes is never "organically" fused in the visual system; that is, projection neurons of the two eyes never converge onto common binocular neurons, such as we hypothesize in box B of Figure 2:8. Instead, associations between features coded in the two pathways may be set up at a "psychic" level; associated features are perceived as single because the observer has learned that they are images of a single object in space. When image features presented to the two eyes are different and rivalrous, the system "attends" to one image or the other, and while the attended features are consciously perceived, the unattended features seem to be suppressed.

In some respects, Helmholz's theory is clearly wrong. For example, it is now known, from physiological studies, that at least some binocular information is organically fused. On the other hand, attention may still play a critical role in binocular combination as he suggested. This is the case with "architecture 4" outlined in the previous section. Presumably the mechanisms which direct one's attention must exist at an unconscious level, and must have access to the information which seems suppressed at the conscious level. The function served by these mechanisms is to direct attention to a subset of image features, some taken from either half image, which is semantically meaningful. According to this view, suppressed

information is not lost, and differs from dominant, unsuppressed information only in its status with respect to conscious perception.<sup>3</sup>

I will consider two experimental paradigms for investigating unconscious control of suppression. In the first, complex rivalrous patterns are presented to either eye and we determine whether the consciously perceived, combined view is pieced together from the two half images in a way which results in a semantically meaningful pattern or a nonsensical pattern. Again, one would expect the former to occur if suppression is controlled by high level object detection processes. The second paradigm involves visual search: one attempts to locate and "make visible" particular stimulus points which are initially suppressed. Again, on the assumption that both suppressed and unsuppressed stimulus information is available to the attention and other subconscious

---

<sup>3</sup>This question, "is suppressed information lost," has relevance to perceptual processes other than those responsible for binocular combination. For example, in Arbib's (1975b) model for perception, a stimulus pattern activates a number of internal "schema." A subset of the active schema becomes dominant, and it is this array which determines conscious perception. Other schema remain active but not sufficiently active to gain the status of conscious perception. Associated with each object-related schema is an array of stimulus points, an image segment, as described in the previous chapter. We may ask, therefore, whether individual schema can influence the dominance-suppression status of points within their segments, and whether schema have access to suppressed points. If the answers to these questions are "no," then suppression mechanisms impose considerable constraints on subsequent visual processing.

mechanisms, one would expect that suppression should not make a stimulus inaccessible to visual search.

As an example of the first paradigm, consider what happens when one attempts to double his reading rate by reading two books simultaneously, one with each eye. A number of problems are encountered, both mechanical (can one train his eyes to move independently) and attentional (can one comprehend two word streams simultaneously). But there is also a problem of rivalry and suppression. As one reads the word fixated by one eye, the word which is simultaneously fixated by the other eye tends to disappear. If type fonts are similar, as with two pages from the same book, one encounters an additional difficulty: the "words" one fixates may be made up of letters seen by both eyes, while other letters from both eyes are suppressed. Such words will generally not make sense. In some cases, letters will be combined from pieces of letters from both eyes and will also be nonsense. It may be concluded that semantic information processing routines, such as may be involved in word comprehension as well as character recognition, do not control the rivalry and suppression mechanisms.

Another example of this paradigm which is perhaps more adapted to visual thinking is shown in Figure 2:9. The two half images contain different features of a face, along with a non-face feature, a short vertical line in the right half

image which is rivalrous with the "eye" feature in the left half image. The combined view may be a completed face, in which features from either half image are assembled into a new semantic unit with the vertical bar suppressed, or the combined view may be of a face with one eye replaced by a vertical bar. My experience is that both configurations occur with roughly equal probability, and that one sees repeated alternation between the two with continued viewing. Again, the conclusion is that semantic, efferent control is not critical in suppression phenomena.

One argument advanced by Helmholtz in support of the high level suppression control hypothesis makes use of the second paradigm: search for a suppressed feature. Helmholtz reports that when one views the orthogonal grid stereogram (Fig. 2:2c) he may willfully cause one or the other grid to become dominant by appropriately directing his attention. However, this demonstration is not convincing because eye movements may play a critical role. A moving stimulus on one retina will tend to dominate and suppress a stationary stimulus on the other retina. Thus, motion is a factor affecting the afferent control of suppression. If, when one views an orthogonal grid stereogram he moves his eyes parallel to one grid and perpendicular to the second, the latter will be a far stronger stimulus and will tend to become dominant. Thus, an alternative explanation of one's ability to purposefully alternate

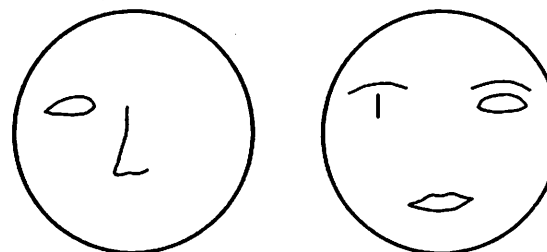


Figure 2:9

Binocular combination of this stereogram results in either a face, with the vertical line suppressed, or a face with one eye replaced by the vertical line.

between grids is that he has, perhaps "unconsciously", learned to move his eye in one direction to see one grid and in an orthogonal direction to see the other grid. In this case, control of ocular dominance is indirect and dependent on changes in the afferent stimulus. Strong support for this second explanation of Helmholtz's example is obtained when grids are replaced by uncorrelated random dot stereograms (Fig. 2:7a). In this case, direction of eye movements does not differentially enhance stimulus strength of the two half images. In my experience, it is very difficult to willfully change ocular dominance with this type of stereogram. If one adds small monocular "flags" to either half image and gives himself the task of locating the suppressed flags, the search may take several seconds. Even then the dots seem to become visible only after an uncontrolled change in ocular dominance. Since search for flags in the dominant half image takes a small fraction of this time, we may conclude that suppressed information is not available at an unconscious level for examination by attention directing and search mechanisms. A more parsimonious assumption is that suppressed information is actually lost.

To summarize this section, evidence is given to show that suppression during binocular combination is not under high level efferent control. In terms of the flow diagram in Figure 2:8, it may be concluded that control pathway 5,

and hence information processed in box C, does not play a major role in control of suppression.

## 2.5. Information Coding

How is visual information coded at the level of the visual system at which rivalry and suppression are mediated? In view of the fact that the minimal rivalrous stimuli are contours of different orientations<sup>4</sup>, it is tempting to suppose that this coding is largely in terms of elementary contour segments. These elementary features of the human visual system may correspond to the "simple cells" found in the visual cortex of cats and monkeys by Hubel and Wiesel (1962, 1968). The code might also include some compound features, composed of several contours, as in Hubel and Wiesel's complex and hypercomplex cells. If image coding of this kind does exist, we might account for rivalry and suppression by postulating reciprocal inhibition between cells which code different features in roughly the same visual direction.

It is interesting to note that suppression is subjectively similar to the fading of retinally stabilized images (Riggs et al., 1953; Pritchard, 1961); in both cases individual stimulus points seem to completely disappear and groups

---

<sup>4</sup>We are not interested here in uncountoured stimuli, which may be rivalrous due to a difference in color or intensity.

of points disappear together. In the case of stabilized images, it has been suggested that these groups of stimulus points which fade together actually represent single, elementary features in the image code. The group fading phenomenon is cited as evidence of feature coding in human vision. (This interpretation is suggested, for example, by Kaufman, 1974). For example, if a triangle image is stabilized on the retina, corners or edges may disappear as units. If these units actually correspond to elementary features in human vision, and if reciprocal inhibition between detectors for these features is responsible for suppression of rivalrous stimuli, then a careful systematic examination of rivalrous stimuli may allow us to discover the elementary features of human vision by psychophysical means.

So far in this section, I have discussed what may be called a feature detection theory of rivalry. Its principal postulates are, again: 1) images are coded by activity in feature-specific neurons, and 2) suppression is due to inhibition between neurons which code different features. Since the feature-specific neurons found by Hubel and Wiesel are principally binocular, we shall assume the feature detectors of this model are also binocular, so that the proposed rivalry mechanism resides in box B of Figure 2:8. Thus the feature detection theory corresponds to architecture 3, in which suppression is mediated by pathway P4.

The feature detection theory has difficulties both on theoretical and empirical grounds. For example we must account for the fact that inhibition between rivalrous features occurs when the features are presented to opposite eyes, but not when they are presented to the same eye. Suppose that a short vertical bar is presented to the left eye, and a short horizontal bar to the corresponding region of the right eye. These features activate horizontal and vertical bar detectors in cortex (box B, Figure 2:8) as shown in Figure 2:10a. The detectors are binocular, so will be activated when the appropriate features is presented to either eye. Therefore the same detectors are activated by the "plus" sign presented only to the right eye in Figure 2:10b. In the first case, rivalry occurs, so we postulate reciprocal inhibition between the detectors. Inhibition "suppresses" activity in one or the other detector. However, if this inhibition between binocular feature detectors actually exists, rivalry and suppression should also occur in case b. That is, when one views a "plus sign" with one eye, he should tend to see either a vertical line segment with the horizontal suppressed, or vice versa. In fact, entire plus signs are easy to see even with monocular viewing: both line segments seem to have equal brightness and there



is no rivalry sensation.<sup>5</sup> As this example makes clear, suppressive inhibition should occur between detectors for dissimilar features only when the detectors are activated by stimuli which are presented to different eyes. This is a dilemma for the feature detection theory of rivalry, since we have assumed that feature detectors are binocular. The simplest solution to the problem is to abandon this assumption. We might suppose instead that stimuli presented to the two eyes are coded separately, but still in terms of elementary features. Each monocular feature detector will inhibit only detectors for other features in the other eye. A model based on these interactions between monocular codes corresponds to architecture 1, as described in section 3 of this chapter. Merits of this type of model will be considered shortly.

On the other hand, it may not be necessary to abandon the idea of binocular feature detectors if we suppose interactions between detectors are properly dynamic. Note that in the above example there are more features present when the

<sup>5</sup>Campbell and Howell (1972) have observed "monocular rivalry." When two sinusoidal gratings which differ in color and orientation are superimposed and monocularly viewed, one or the other may seem to dominate at a given moment, and dominance alternates between grids over time. Monocular rivalry seems to be peculiar to sinusoidal grids. The effect is reduced if the grids are made the same color or if square wave grids are used. Monocular rivalry certainly does not occur frequently under normal viewing conditions, while binocular rivalry is common. Eye movements may help explain the effect.

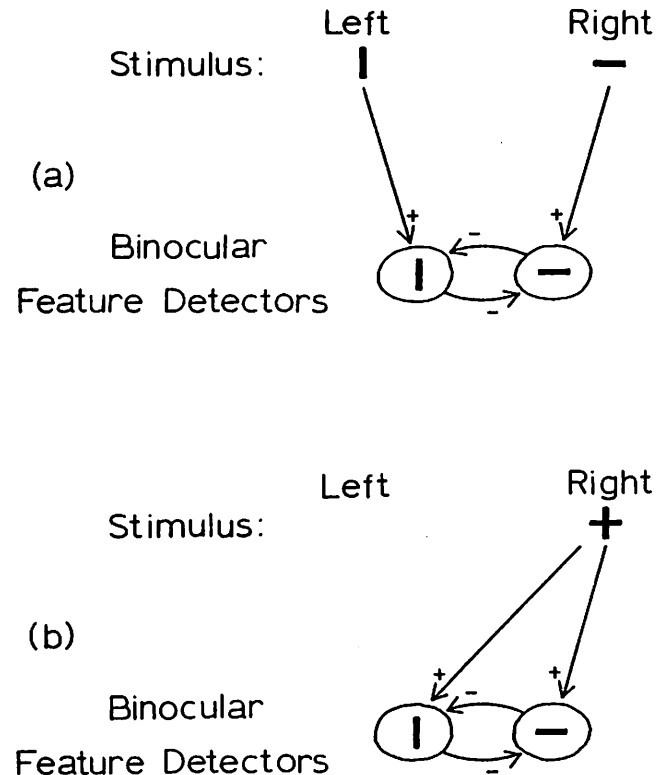


Figure 2:10

Binocular feature detectors for vertical and horizontal bars are shown stimulated by features presented to opposite eyes in 'a', and a single eye in 'b'. If inhibition between detectors causes suppression in 'a', then it should cause suppression in 'b' as well.

two bars are presented to one eye, as a "plus sign," than when they are presented separately to the two eyes. In particular, there are features which describe the spatial relation of the two line features, i.e. elementary features which record the fact that the two line segments bisect each other and form four right angles. The various features which are excited when the two lines are presented to a single eye form a consistent set of features. On the other hand, if only the horizontal and vertical line segment detectors are excited, as in binocular presentation, the set of excited features is incomplete or inconsistent. The binocular feature detection theory of rivalry may work if there exist processes which organize or "assemble" elementary features: when the features are an inconsistent set, they cannot be fully assembled and the system disregards, or "suppresses" extra, inconsistent features. Thus by a dynamic consistency checking process at a binocular level, we may explain what subjectively appears to be eye specific suppression.

There is, however, another difficulty with a binocular feature detection theory of rivalry. When suppression occurs it is not specific to the rivalrous features but extends to all features occurring within a local region of the same eye. Thus when a particular feature is suppressed, nearby features which are not rivalrous may also be suppressed. This effect was clearly demonstrated in the case of uncorrelated random dot stereograms, Figure 2:7a. While it may be possible to

account for this area suppression phenomena in terms of dynamic feature organizing processes, within the context of a binocular feature detection theory, that theory must be very complex indeed. I shall not consider the theory in any greater detail, because there are much simpler theories which follow from the alternative assumption that suppression takes place at a monocular level (box A).

The observation that suppression spreads over regions of a monocular image suggests not only that suppression occurs at a monocular level, but that inhibition responsible for suppression is directed at all ~~image~~ features, or stimulus points, within the region, rather than just at the specific features which are rivalrous with features seen by the other eye. An "area theory" for suppression follows from these observations. According to this theory, suppressive inhibition is not feature-specific, but directed at all stimulus points within a "target area" of one monocular image.

The area theory corresponds to a model architecture in which suppression occurs in box A of the flow diagram in Figure 2:8. In addition, the control pathway must be intrinsic, P1. Efferent control, via P3, cannot work for the following reasons. Information concerning the eye of origin of particular stimulus features is not retained in box B, so the proposed efferent suppression signals must be the same for both eyes. Since the inhibition is not feature-specific,

all features should be suppressed in the same target area of both visual fields.

Within box A, image information may be coded in terms of activity in neurons with concentric and antagonistic center-surround receptive fields. This type of code is known to exist in cats and monkeys in the monocular cells of the lateral geniculate body, so is reasonable on anatomical and physiological grounds. It is also attractive for more theoretical reasons.<sup>6</sup> First, the level of activity in neurons of this type is roughly equal to the second spatial derivative of the stimulus intensity function within the receptive field. Thus image contours are particularly good stimuli for center-surround neurons, so these neurons respond well to just the type of stimulus which tends to dominate in rivalry situations. Furthermore, when the interocular interactions include spread of inhibition and hence the possibility of recurrent disinhibition (as in Figure 2:6), there will be a tendency for a number of center-surround cells which are stimulated by a contour in one eye, to become dominant together, while another group of cells which are stimulated by a different contour in the other eye may be suppressed together. This cooperative behavior can account for the apparent feature-specific nature of suppression.

---

<sup>6</sup>Sperling (1970) has also assumed center-surround coding in his model for rivalry and suppression.

In this section, several arguments have been advanced to discredit feature detection theories of rivalry and suppression, and an alternative area theory has been introduced. All suppression architectures defined in Section 3 have been eliminated, except the one which places suppression at a monocular level under intrinsic control. It has been suggested that image coding in this system is in terms of activity in cells with center-surround receptive fields. This code is further developed in the next section, while the credibility of the area theory as a whole will be established in later sections in which the model is simulated.

## 2.6. Scale Factors

Here, and in all subsequent discussion, we shall assume that information is coded in terms of activity in monocular cells with antagonistic center-surround receptive fields. Suppose that the stimulus image at a given moment in time is described by the intensity function  $F(E, \theta)$ , where  $\theta$  is direction in the visual field, and  $E$  designates the eye ( $E = \text{Left}$  or  $\text{Right}$ ). The stimulus excitation  $S(E, \theta)$  of a given neuron which has a receptive field centered at  $\theta$  is given by the convolution:

$$S(E, \theta) = \int F(E, \phi) g(\phi - \theta) d\phi$$

where  $g$  is a weighting function which characterizes an antagonistic center-surround receptive field, see Figure 2:11a. It is not important to define  $g$  quantitatively, but we assume it is symmetric about zero, triphasic, and  $\int g(\phi) d\phi = 0$ . In the absence of inhibition from other cells, the activity level of a cell,  $A(E, \theta)$ , is proportional to the stimulus  $S(A, \theta)$ . However, we assume that activity in any cell which codes information for one eye ( $E$ ) has an inhibitory influence on cells coding information in the corresponding region of the other eye, ( $\hat{E}$ ). The strength of the inhibition directed at the cell with location  $\hat{\theta}$ , is proportional to the activity level  $A(E, \theta)$ , and is maximal when  $\hat{\theta} = \theta$ , but decreases monotonically as  $|\hat{\theta} - \theta|$  is increased. In particular suppose that there is a weighting function  $w(\phi)$  associated with interocular inhibition, which is symmetric about  $\phi = 0$ , and which decreases monotonically to  $w(\phi) = 0$  for large  $|\phi|$ . For example,  $w(\phi)$  may resemble a gaussian distribution, as in Figure 2:11b, but again the quantitative description of  $w$  is not important.

The inhibition at point  $\hat{\theta}$  of eye  $\hat{E}$  is due to activity in a cell centered at  $\theta$  in eye  $E$  is:

$$I(\hat{E}, \hat{\theta}) = A(E, \theta) w(\hat{\theta} - \theta)$$

This inhibition may cause "suppression" of activity  $A(\hat{E}, \hat{\theta})$  (that is, it may reduce  $A(\hat{E}, \hat{\theta})$  to zero) if it exceeds some

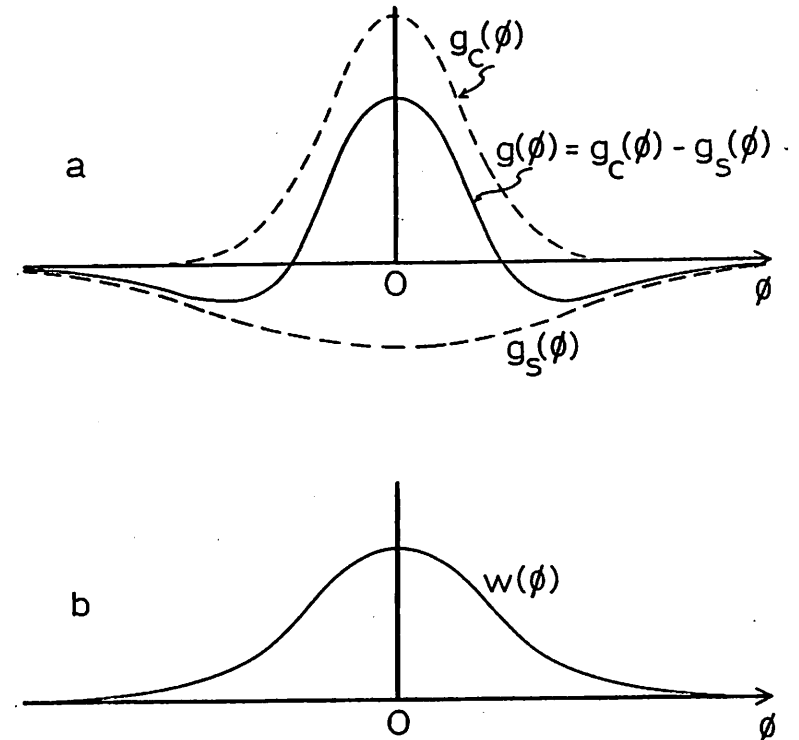


Figure 2:11

Weighting functions used to compute neuron responses and inhibitory interactions.

critical value. Note that the range over which this suppression may occur increases with  $A(E, \theta)$ , and decreases with increases in  $A(\hat{E}, \hat{\theta})$ .

If it is assumed that the receptive fields of all cells are characterized by the same weighting function,  $g$ , and that all interocular interactions are characterized by the same weighting function,  $w$ , then the model is a "single channel" model of rivalry. Alternatively, we may consider a "multi-channel" model in which all cells have a center-surround receptive field organization, but the dimensions of this receptive field may vary from cell to cell.<sup>7</sup> Suppose that receptive fields differ only in scale. Then a scale parameter,  $s$ , may be associated with the stimulus convolution:

$$S(E, \theta, s) = \frac{1}{s} \int F(E, \phi) \cdot g\left(\frac{|\theta - \phi|}{s}\right) d\phi$$

Furthermore, suppose (for reasons which will be clarified later) that inhibition between cells of different eyes is strongest for cells with the same value of  $s$ , and decreases

---

<sup>7</sup>The "channels" described here should not be confused with the principal and control channels discussed in Section 3. Other evidence based on the visibility of sinusoid gratings and supporting a multi-channel model, in the present sense, has been obtained by Graham, 1974.

monotonically as the difference in  $s$  values increases. Therefore, if we let  $G$  represent this monotonic decreasing, symmetric relationship, we have the following description of interocular inhibition:

$$I(\hat{E}, \hat{\theta}, \hat{s}) = A(E, \theta, s) w(\hat{\theta} - \theta) G(\hat{s} - s)$$

The weighting function  $G$  may have the same shape as  $w$ . Figure 2:11b. The scale parameters vary from the limit of resolution to several degrees.

The various functions used here to define the multi-channel model will be made explicit in later sections of this chapter. In the remainder of this section, I will consider experimental evidence which suggests a multi-channel rather than a single channel model.

Two elements from opposite eyes which have receptive fields centered on the same point, and which are characterized by the same receptive field scale, are said to "fuse" if reciprocal inhibition of one by the other is not sufficiently strong to suppress activity in either element. (The possibility of "fusion" of elements centered in slightly different visual field positions, i.e. disparate elements, will be considered in the next chapter). According to this definition of fusion, elements with receptive field scales within one range of values may fuse, while other elements in the same

area of the visual field but with larger or smaller receptive fields may be unfused and rivalrous. As was noted in Section 2, it is this possibility of partial fusion in the multi-channelled model which will allow us to explain co-existence of rivalry and stereopsis in stereograms of Kaufman and Julesz, as well as stereopsis with dissimilar and diplopic images.

Two predictions which can be made with the model are that the range of suppression associated with a given monocular feature should depend on contrast, which determines  $S(E, \theta, s)$ , and feature size, which determines the scale of optimally stimulated cells. These predictions are consistent with the observation of Crovitz and Lockhead (1967) that line contrast in the stereograms of Figure 2:3a affects the range of suppression, and my own observation that complete suppression of a grid by another orthogonal grid can occur even if the bar widths of the grid are greater than two degrees.<sup>8</sup>

---

<sup>8</sup> Another prediction of the model is that a grid presented to one eye will suppress another orthogonal grid of the same spatial frequency presented to the other eye about 50% of the time. It may be more or less effective in suppressing grids of other spatial frequencies. We may expect to find a balance where grids of different frequencies are equally likely to suppress one another, by varying the contrast of one of the grids. It may be possible to obtain an estimate of the weighting function  $G(\bar{s} - s)$  by systematically measuring these balance points for many combinations of grid frequencies.

Evidence in favor of a multi-channel model also comes from the stereograms in Figure 2:7. It was observed in Section 2.2 that a random dot pattern of about 50% density is very effective in suppressing another uncorrelated pattern of the same density. Areas of suppression could be identified by adding small, isolated pencil marks, or "monocular flags," to either half image - the flags were suppressed along with nearby dots in the random pattern. Now suppose we construct a stereogram in which a random dot pattern is presented to the left eye while a uniform, uncountured field is presented to the right. If a small pencil mark is made in the uniform image, this mark becomes dominant and suppresses nearby random dots in the other image! Thus when the pencil mark is isolated in the uniform field (Fig. 2:7c), it is a far stronger stimulus for suppression than are dots of the random dot pattern. But when the same pencil mark occurs in one random dot pattern (Fig. 2:7b), it is relatively ineffective in suppressing dots in another random dot pattern.

This behavior is contrary to what should be expected from the single channel model. According to that model, suppressive strength should increase monotonically with amount of contour or with number of dots per unit area. On the other hand, these results appear to be consistent with the multiple channel hypothesis. A random dot pattern is a good stimulus for units in which the receptive field center is about the same size as stimulus dots. Units with much

larger receptive fields are poorly stimulated since both the center and surround subregions will be exposed to large numbers of dots. A single isolated pencil mark will be a fairly good stimulus, even for units with large receptive fields, since when it stimulates one subfield of these units, its effect is not balanced by other dots in the antagonistic subfield. It is these channels in which elements have large receptive fields that control dominance in Figure 2:7c.

A final example of scaling is demonstrated by the stereograms of Figure 2:12. These stereograms are constructed of pairs of different sized black disks, which appear concentric when binocularly viewed. In the combined image, the smaller disk appears entirely black and there is a white region just outside its boundary. The larger disk appears black just inside its boundary, but this grades gradually to gray and then white as one follows a radial path from the outer disk boundary to the inner disk boundary. We are interested now in the apparent intensity profile of this gray region. I have drawn a rough estimate of the radial intensity profiles for each of the stereograms in Figure 2:12. These profiles reflect the horizontal spread of inhibition associated with each contour. If the sizes of both disks are doubled or quadrupled (compare Figures 2:12a,b,c), the profile does not change significantly except in scale along the radial axis. Thus it seems that the range of contour interactions increases

proportionally with size of the stimuli (disks). On the other hand, if only one disk is varied in size (Figures 2:12c, d), there are qualitative changes in the intensity profile.

## 2.7. Summary

Many aspects of the binocular rivalry and suppression phenomena have been considered and specific hypotheses about the neural circuits which underlie these phenomena have been proposed. These hypotheses will become the defining assumptions of the neural network model which is described in the next section. Computer simulations of the model will also be described and these are in substantial agreement with all empirical data discussed thus far. Specific hypotheses may be summarized as follows:

1. The apparent disappearance of regions or features of a visual image in binocular suppression is due to the elimination of the neural activity which codes this visual information by means of neural inhibition.

When a stimulus is not consciously perceived, we say that it is suppressed. The above conclusion states that suppressed information is also not available for subconscious information processing. Evidence for this conclusion is of two sorts. According to the present interpretation of data,

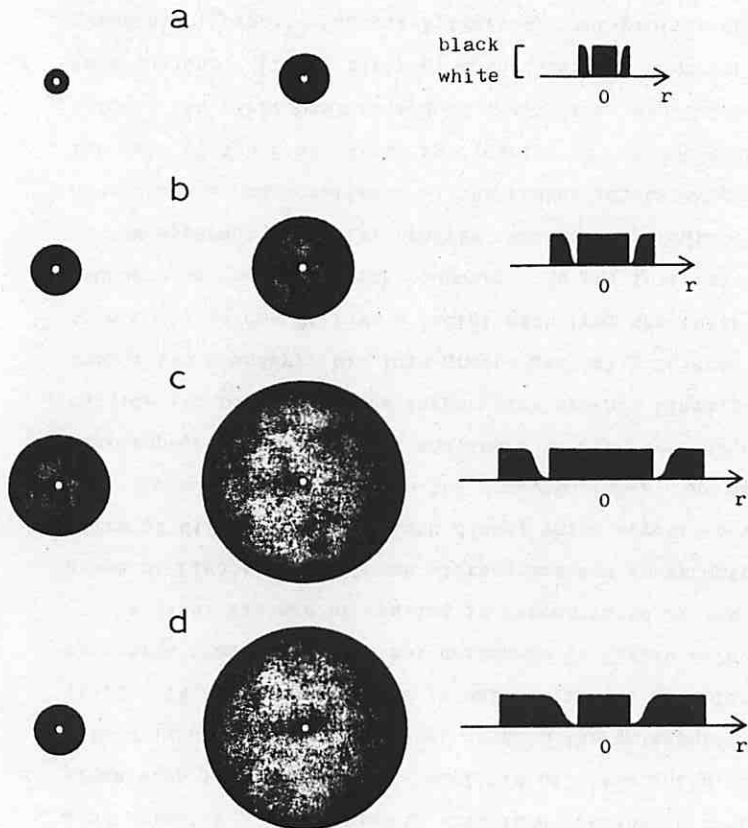


Figure 2:12

In each of these stereograms the center dots are superimposed so the smaller disk appears concentric with the larger disk. The small disk "carries with it" a "halo" of white. The extent of the halo is indicated roughly by the apparent intensity profiles to the right.

stereopsis is impossible if all image information from either eye is suppressed. Also it was concluded that one cannot control ocular dominance by appropriately directing his "focus of attention" and suppressed information is not available to visual search tasks.

2. Dominance and suppression are under afferent/intrinsic control.

Certain stimulus qualities, such as contrast and feature size, seem to determine ocular dominance. The alternative hypothesis, that dominance and suppression are under efferent control is discredited by the same arguments as were mentioned above in support of hypothesis 1. In addition, it has been shown that with complex rivalrous stereograms, the combined image is not necessarily pieced together from subregions of the two half images in a way that yields a semantically meaningful or familiar perception. This was true with the face stereogram, Figure 2:9.

3. Suppressive inhibition is directed not at individual elementary features, but at monocular afferent activity which codes all stimulus points falling within a local region of the monocular visual field.

Support for this hypothesis comes from the observation



that when a particular feature is suppressed, nearby stimulus points are also suppressed, even if they are not rivalrous with stimuli in the other eye. The uncorrelated random dot stereogram of Figure 2:7 provides a good illustration of this effect.

4. Suppressive inhibition is recurrent and spatially diffuse.

If a stimulus presented to one eye is dominant, it will inhibit stimuli presented to the corresponding point and neighboring points of the other eye. On the other hand, if the stimulus is not dominant but suppressed, it will not inhibit stimuli to the other eye. This type of recurrent, spreading inhibition is postulated to account for the observation that regions of dominance by one eye tend to spread over the visual field and may include subregions which, on the basis of strictly local stimuli, would normally be dominated by the other eye. Dominance at one point results in disinhibition of neighboring points of the same eye, thus increasing the probability that these points will be dominant as well.

5. There is a distinct state of binocular, sensory fusion.

Fusion rather than suppression may account for binocular single vision when the stimuli to the two eyes are not too different. This conclusion follows from two types of psychophysical evidence. Again we note that stereopsis fails when one or the other half images is completely suppressed due to rivalry. This implies that stereopsis mechanisms rely on unsuppressed information from both eyes. Singleness of vision with normal correlated stereograms is best attributed to sensory fusion. Suppression, if it were to occur with correlated stereograms, should show the same tendency to spread over extended regions of the visual field, as it does with uncorrelated, clearly rivalrous, random dot stereograms. This was shown not to occur.

It was necessary to modify the notion of fusion to account for a number of stereopsis and rivalry phenomena. In a "multi-channel" code, it is possible for fusion to occur in some channels but not necessarily in all channels for a given visual direction. (See hypothesis 7).

6. At the level of the visual system at which binocular suppression occurs, visual information is coded by activity in neurons with (monocular), antagonistic center-surround receptive fields.

No specific experimental support is offered for this hypothesis. However, as will be made clear in the model

simulation in the next section, this type of code is adequate to account for the various phenomena which have been discussed here. Neurons with this type of receptive field respond well to image contours, and when this fact is coupled with hypothesis (4) of recurrent spreading inhibition, we may account for the apparent feature-specific rivalry which occurs between differently oriented line elements.

The alternative hypothesis, that coding is in terms of elementary features (Hubel and Wiesel's simple and complex cells) could be incorporated in the model and result in essentially the same predicted behavior. The center-surround code is preferred here because it is simpler to model and because this will be more consistent with subsequent discussion of the lateral geniculate body as a possible locus of rivalry interactions.

#### 7. The code is "multi-channelled."

All neurons which code image information at the level of binocular rivalry interactions are assumed to have the same center-surround receptive field organization. However, the size or scale of the receptive fields varies from neuron to neuron. Neurons which have receptive field sizes falling within a narrow range may be called an "information channel." The entire image is coded by each channel, but the scale of

image details which predominate differs between channels. This multi-channel assumption is critical to the explanation given here of a number of phenomena such as the "paradoxical" co-existence of fusion and rivalry within the same area of the visual field, and the observation that the range of suppression increases with stimulus size.

This concludes the list of hypotheses. It is interesting to note that the neural structure for binocular combination described by these hypotheses is about the simplest of any considered here. While one will generally favor a simple model to a more complex model when both account for empirical data, it was not this consideration which guided formulation of the present structure. Rather, the simple structure proved more consistent with the psychophysical data than did alternative structures.

The fundamental conclusion of this study is that binocular combination seems to occur at an early stage of visual processing and seems to be mediated by rather unsophisticated neural processes. In suppressing image features, the system effectively throws out visual information. One may presume that there is a reason for disposing of information; for example, the visual system has limited computing resources so it may not be able to simultaneously process two different images presented to different eyes. But it is most interesting that decisions about what information to suppress can

be made so early in processing without reference to semantic or other high level information.

This observation has important implications for computer vision: it should be possible to implement binocular combination in computer vision with low level, non-semantic, and relatively unsophisticated image processing.

#### 2.8. The Model, Part I: Image Code

To simulate binocular combination of rivalrous stereograms, we must 1) code the half images, 2) combine the two half image codes into a single binocular code, and 3) decode the binocular code in order to determine if the combined image generated by the model matches empirical data. In this section I shall be concerned with image coding and decoding, and in the next section I shall describe the procedures for combining codes and describe several simulations.

The coding (and code combining) procedures are based on the seven hypotheses formulated in previous sections of this chapter. These procedures are implemented in a way which facilitates computer simulation. Thus each input image for the model is digitized and expressed as a two-dimensional array of numbers such that each number represents the image intensity at a corresponding point in the real image. The image code is expressed as several arrays of numbers, one array for each code "channel." We may imagine that every

code element corresponds to a neuron in the visual system with a concentric center-surround receptive field. The elements within one channel all have receptive fields of the same size, while elements of different channels have different size receptive fields.

The numerical value of each code element is obtained by convolving an appropriate weighting function with the input image. These convolutions are performed very rapidly by a recursive algorithm. This rapid coding is possible because the receptive field scales which characterize various channels differ by powers of 2, and the elements of each code array are positioned so that the inhibitory surround of an element in one array may be obtained from the excitatory center of an element in another array. The code may be represented as a hierarchy of arrays, as shown (for one dimension) in Figure 2:13, and I shall refer to particular arrays by the level they occupy in the hierarchy. Thus a level 1 array has elements with the smallest receptive fields, and the size of the receptive field of an element in level  $n$  is twice that of an element in level  $n - 1$ . Note that the level  $n$  array has half the number of code elements as level  $n - 1$ . (Actually, in a 2 dimensional code, the number of elements differs by a factor of 4). However the reduced number of elements still "covers" the image because their receptive fields are larger.

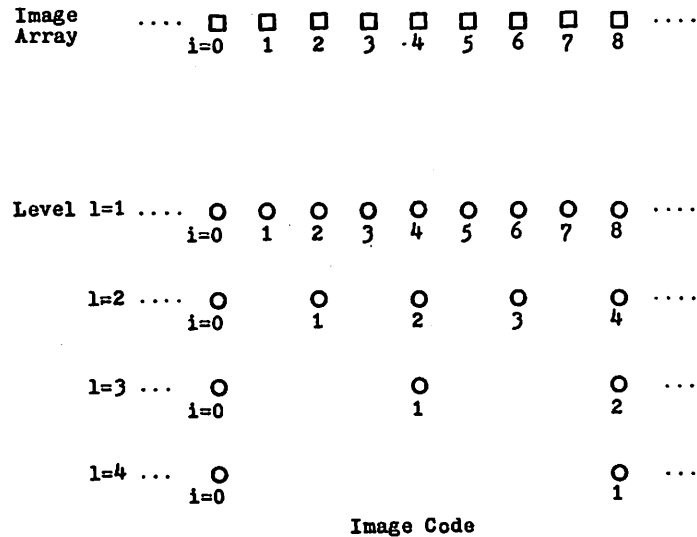


Figure 2:13

The "image array" is the function  $F(i,j)$  which specifies image intensity at a regular array of points. Circles represent code elements. These are arranged in arrays and each array is referenced by its level in a hierarchy of arrays. The horizontal position of an element within the array corresponds to the position of its receptive field center with respect to the image array. Only one dimension of the two dimensional array is shown here.

The weighting function which defines the receptive field of each element may be divided into two parts,  $g_c$  and  $g_s$ , which determine the contributions of the center and surround respectively. Both  $g_c$  and  $g_s$  are bell shaped curves which are normalized so that the area under the curves is equal to 1 (in appropriate units). However, the width of  $g_s$  is approximately twice that of  $g_c$ , as shown in Figure 2:11a. If we let  $S(i,j,l)$  be the code value of the element in position  $i,j$  of level  $l$ , then this also may be expressed as the difference between the center and surround contributions:

$$S(i,j,l) = S_c(i,j,l) - S_s(i,j,l)$$

The  $S_c$  and  $S_s$  contributions are determined by the following recursive procedure.<sup>9</sup>

Let  $F(i,j)$  be the input image array, and  $L$  be the number of code levels.

For  $l = 1$ ,

<sup>9</sup>The value obtained for  $S$  by this procedure may be positive or negative. One interpretation of a code element is that it corresponds to a neuron with a center-surround receptive field, and the level of activity, or spike frequency, of the neuron is  $S$ . To accommodate both positive and negative  $S$  values, we may suppose that each code element is actually a pair of neurons, both of which have the same receptive field but in one, the center is excitatory (this codes positive  $S$ ) and in the other, the surround is excitatory (this codes negative  $S$ ). Alternatively we may suppose that  $S$  is a departure from a non-zero resting frequency.

$$S_c(i,j,1) = F(i,j)$$

For  $l > 1$  and  $l < L$ ,

$$S_c(i,j,l) = \sum_{n=-2}^2 \sum_{m=-2}^2 w(n,m) S_c(n+2i,m+2j,l-1)$$

For  $l = L$

$$S_s(i,j,l) = 0$$

For  $l > 1$ ,  $l < L$ ,

$$S_s(i,j,l) = 4 \underbrace{\sum_{n=-2}^2 \sum_{m=-2}^2 w(n,m)}_{\substack{i+n \text{ even,} \\ j+m \text{ even}}} S_c\left(\frac{i+n}{2}, \frac{j+m}{2}, l+1\right)$$

Thus the central contribution,  $S_c$ , to  $S$  at level  $l$  is simply the intensity value  $F$  at the corresponding image point.<sup>10</sup>

<sup>10</sup>The code as described here represents a linear sum of image intensity over the receptive field. However, if  $F(i,j)$  is the log of the image intensity plus one (so the  $F(i,j)$  is always positive), then the same coding algorithm may be used, but with a slightly different interpretation. A code element represents the log of a kind of integrated product of image intensities in the central receptive field region, divided by a similar integral over the surrounding receptive field region. The code in this case represents image contrast independent of background illumination.

The central contributions to  $S$  at other levels is obtained recursively by a weighted sum over a 5 by 5 window of the values obtained at the next lower level. The surround contribution,  $S_s$ , at the highest level,  $L$ , is assumed to be zero, while the surround contributions at other levels is determined recursively from values obtained for  $S_c$  at the next higher level. The same weighting function applies for both computations, but in computing  $S_s$ , only about one out of every four terms indexed in the 5 by 5 window is actually included in the sum. This reflects the fact that each code element has one quarter as many parent elements as daughter elements in the hierarchy.<sup>11</sup> Note that in computing  $S_s$ , only those values of  $n$  and  $m$  are used which when added to  $i$  and  $j$  respectively result in an even number. (The reader interested in details of indexing should refer to Figure 2:13). The weighting function is designed in such a way that the total of the weights of terms used in this double sum is always  $\frac{1}{4}$ , regardless of the values of  $i$  and  $j$ , (see below). This is why the sum is multiplied by 4.

The choice of weights  $w(n,m)$  is subject to these constraints. First, the total weight within the 5 x 5 window

<sup>11</sup>The exact number of parents an element has depends on its position in the hierarchy of elements. The average is about 6. However, weights are such that each parent contributes effectively four times as much to  $S_s$  as each daughter contributes to  $S_c$ . The reader interested in details of this relationship is advised to study Figure 2:13 and work through several examples.

should equal 1:

$$\sum_{n=-2}^2 \sum_{m=-2}^2 w(n,m) = 1$$

Second, the weighting pattern is symmetric about  $n = 0$ ,  $m = 0$ , so for each  $n, m$

$$w(n,m) = w(n,-m) = w(-n,m) = w(-n,-m).$$

And third, the weights are distributed so that all stimulus points  $F(i,j)$ , contribute equally to the code at each level of the code hierarchy. This condition is satisfied when for  $i = 1,2$  and  $j = 1,2$

$$\underbrace{\sum_{n=-2}^2 \sum_{m=-2}^2 w(n,m)}_{\substack{i+n \text{ odd,} \\ j+m \text{ odd}}} = \frac{1}{4}$$

For example, the following weighting pattern satisfies these constraints: (the  $i, j$  entry in this array is equal to  $w(i, j)$  times 48).

j=	-2	-1	0	1	2
i=-2	0	1	2	1	0
-1	1	4	6	4	1
0	2	6	8	6	2
1	1	4	6	4	1
2	0	1	2	1	0

The image code is now fully defined. Still there are several interesting features of the code which should be pointed out. First we may note that the actual receptive fields of elements at various levels of the code hierarchy have the desired triphasic symmetric pattern. Example receptive fields are shown in one dimension in Figure 2:14 for the first 4 levels of the hierarchy.<sup>12</sup> In each case, a dashed line shows the receptive field of the neighboring elements at the same level. It is apparent that the density of elements is such that the central portions of neighboring elements partially overlap while non-neighbors do not. This provides full coverage of the image without undue redundancy.

It is worth observing that each code element approximates a local second derivative. If the local second derivative of an image is sampled at frequent enough intervals, and if in addition, the absolute intensity and gradient of the image

<sup>12</sup>The weight used here and in other one dimensional examples are 1/12, 3/12, 4/12, 3/12, 1/12.

### 2.9. Model, Part II: Binocular Combination

In this section, I shall describe several procedures for deriving a single binocular code from two codes for the monocular images. I begin with the basic combining procedure, then show how this procedure can be refined.

Suppose the left and right half images of a stereogram have been coded in the way prescribed in the previous section. Each code element has an address, given by its level,  $l$ , and its  $i, j$  position within that level, and a value,  $S(l, i, j)$ . "Corresponding elements" of the two half image codes are pairs of elements, one from each code, which have the same address. In the basic procedure for generating a binocular code, the value of each binocular element,  $S_B(l, i, j)$ , is obtained from the values of the two corresponding elements of the monocular codes,  $S_L(l, i, j)$  and  $S_R(l, i, j)$  by the following rule:

$$\text{Let } \Delta = |S_L(l, i, j) - S_R(l, i, j)|$$

Then

$$S_B(l, i, j) = \begin{cases} \frac{S_L(l, i, j) + S_R(l, i, j)}{2} & \text{if } \Delta \leq \alpha \\ S_L(l, i, j) \text{ or } S_R(l, i, j), \text{ whichever} \\ \text{has the larger absolute value if } \Delta > \alpha. \end{cases}$$

Thus we distinguish two states. Fusion occurs when the values of the left and right code elements do not differ by more than  $\alpha$ . In this case, the value of the binocular element is simply the average of the values of the monocular elements. On the other hand, suppression occurs when the difference in these values exceeds  $\alpha$ . In ~~suppression~~ the value of the binocular element is set equal to the monocular element with the largest absolute value. No provision is made here for partial suppression or spread of suppression.

A network of formal neurons which could approximately mediate this simple type of binocular combination is shown in Figure 2:17. The output of each neuron is equal to the sum of its excitatory inputs minus the sum of its inhibitory inputs times the coefficients  $C$  and  $1/C$  indicated inside each neuron symbol. If inhibition exceeds excitation, the output is zero. A couple of examples will illustrate how this network functions. Suppose  $C$  is just slightly less than 1, and suppose  $S_L = 0$ . Then the output to the right monocular cell will be  $CS_R$ , and the output of the binocular cell will be  $S_B = S_R$ . Suppose, next, that  $S_L = S_R = S$ . Then the output of each monocular neuron is  $CS/(1+C) \cong S/2$ , and the output of the binocular neuron is  $2S/(1+C) \cong S$ . Note that  $C$  must be less than 1 for this solution to be stable. Also  $C$  less than 1 implies  $S_B$  will be slightly larger when both  $S_L$  and  $S_R$  equal  $S$  than when one monocular input is  $S$  and the other is

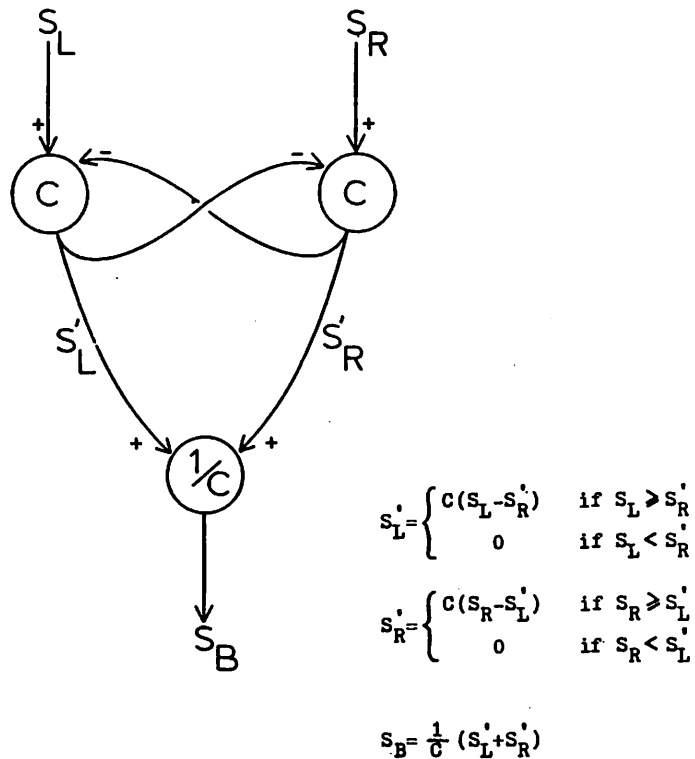


Figure 2:17

This three neuron net combines two monocular inputs,  $S_L$  and  $S_R$ , into one binocular output,  $S_B$ . If  $S_L$  and  $S_R$  are about the same they both contribute to  $S_B$  (fusion), but if one is greater than the other by a factor of  $1/C$  or more, only the larger input contributes to  $S_B$  (suppression). Note that this net is like that shown in figure 1:18 except that here there is no fatigue factor and  $C$  is only slightly greater than 1.

zero. This implies a slight increase in brightness when one views an image with both eyes open over viewing with one eye closed. Such an effect has in fact been observed in psychophysics experiments (De Silva and Bartley, 1930).

Now suppose  $S_L$  and  $S_R$  are both non-zero, and that they are unequal. If either input exceeds the other by a factor of  $1/C$  or more, that input will dominate while the other is completely suppressed.

A simulation of binocular combination of orthogonal bars (see stereogram c in Figure 2:1) is shown in Figure 2:18. Figures a and b show the monocular images. These were coded and combined, then the combined code was decoded to yield the binocular image shown in Figure c. This result is in substantial agreement with the psychophysical result: every contour of the monocular images appears in the combined image and there is a gray level gradient around the inner squares.

There is however one point of disagreement between model and psychophysical results. While the intensity change across the contours around the central square has the right magnitude and direction, the absolute value of the intensity itself is not right: the image intensity should appear white just outside these contours in Figure 2:18c. Instead, this intensity is a gray. This problem is peculiar to cases where regions of dominance by one eye are immediately surrounded by regions of dominance by the other eye. It is due to the fact



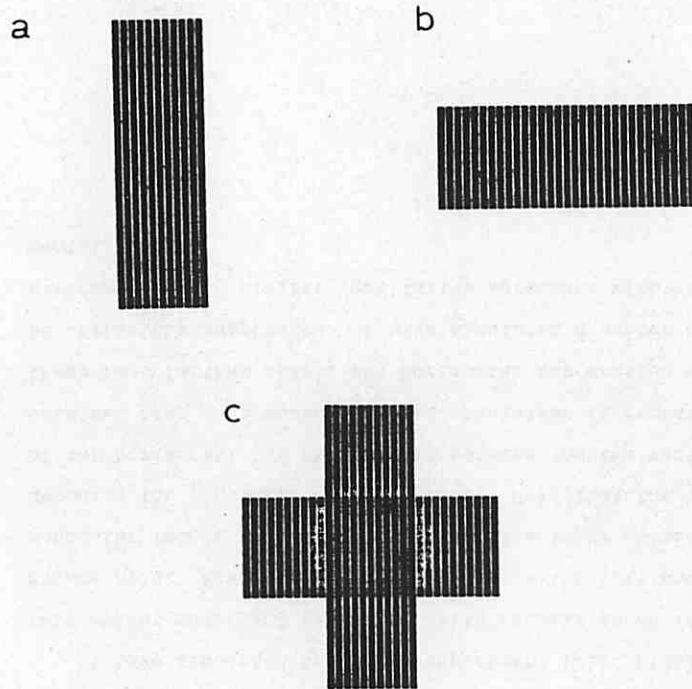


Figure 2:18

These computer generated CRT displays show the intensity patterns of the left and right input images, a and b, and the resultant combined image, c. This simulation shows substantial agreement with psychophysical results with the stereogram in figure 2:1c.

that center-surround coding approximates a second derivative of the image intensity function, so that changes in intensity are directly represented, while absolute intensity levels are not. The problem may not be real: we have not specified the later stages of visual processing which will interpret the binocular code. Also, there are a number of modifications to the code and decoding procedure which could correct the problem, but which I have not yet implemented. For example, absolute image intensity levels could be carried by a separate population of projection neurons. This information may be incorporated with the binocular center-surround code during decoding to re-establish absolute background intensity whenever dominance changes eyes.

The above procedure for binocular combination may be refined to include spread of inhibition. Ideally, this would be achieved by directing inhibition from one element in one code at the corresponding element in the other code and at immediate neighbors of that element. However, a slightly different strategy was used in the present computer model in order to cut down on time-consuming iterative processing. A weight  $W(l,i,j)$  was associated with each code element of the monocular codes. Initially this weight was set equal to the code value  $S(l,i,j)$ , but in the process of combining the monocular codes, the values of individual weights might be modified. In particular, it is the weight rather than the

S value which determines the dominant element of each pair of corresponding elements, and whenever an element in one code is suppressed, the weights of neighboring elements in that code are reduced. High level elements are combined first, and within each level, the elements with the largest weights are combined first.

I have simulated Kaufman's experiment (Fig. 2:3a) using this weight modifying procedure, with results shown in Figure 2:19. Again, figures a and b show the left and right monocular images while figure c shows the image obtained by decoding the binocular combined code. Note that the section of the horizontal bar which falls between the two vertical bars has been suppressed. If the simulation is repeated with these bars farther apart, the horizontal bar section will not be completely suppressed. I have simulated a number of other stereograms with similar qualitative agreement with experimental results.

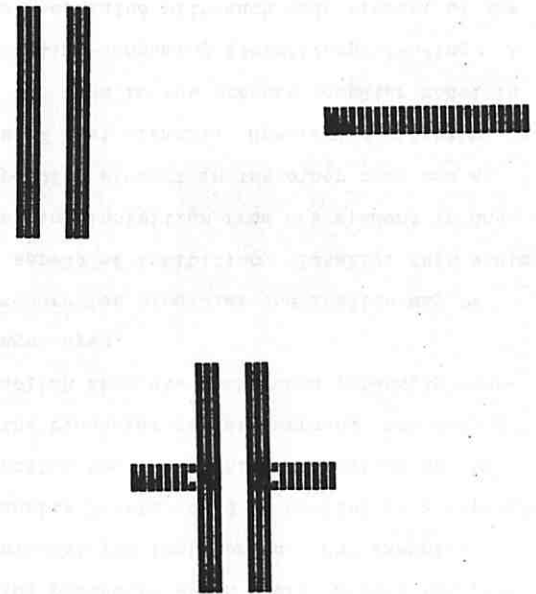


Figure 2:19

This is a computer simulation of binocular combination of the stereogram shown in figure 2:3a. The input images are shown in a and b, and the combined image in c.

CHAPTER III  
A MODEL FOR STEREOPSIS

Introduction

In this chapter I shall consider two questions relating to stereoscopic depth perception: how does the brain combine bits of information from the two eyes in sensory fusion, and how does it resolve the local stimulus matching ambiguities described in Chapter 1? That these processes may result in perception of depth is of secondary importance in this study. Sensory fusion occurs when a feature presented to one eye combines perceptually with a similar feature presented to the other eye, so that only one copy of the feature appears in the perceived visual field. The curious property of sensory fusion is that a feature presented to a given locus in one eye may fuse with a feature presented at any point within a region of the other eye, at different moments in time, while it cannot fuse with two features at different positions within that region simultaneously. It is this "plasticity" of retinal correspondence which needs to be explained.

The stimulus matching ambiguity arises when a feature in one eye could fuse with any one of several features in the other eye. When the ambiguity is properly resolved, the

way in which points are matched locally fits into an overall pattern of fusion and contributes to global perception.

Both the fusion and ambiguity problems have been solved, in principle at least, in projection field models of stereopsis. Several of these models have been proposed in recent years, and these will be critically reviewed in the first section of this chapter. On the other hand, there are several details of stereopsis phenomena which seem to be inconsistent with the projection field structure. A "double projection field" model will be proposed in the following sections to account for these phenomena.

There are other theories of stereopsis which do not postulate projection fields, and do not assume that sensory fusion occurs in binocular vision. A number of these theories have been reviewed by Kaufman (1974). (Kaufman also points out a number of difficulties with projection field models, and these will be considered later in this chapter). As was discussed in the last chapter, suppression is the principal alternative to fusion as an explanation of binocular single vision. There several arguments were given in favor of the fusion mechanism, so pure suppression theories of stereopsis will not be reviewed here. However, suppression does play a role in stereopsis, so both fusion and suppression will be incorporated in the model proposed here.

### 3.1. Projection Field Models for Stereopsis

Stereopsis models have recently been proposed by Julesz (1971), Dodwell (1970), Sperling (1970), Dev (1975), Nelson (1975) and Marr (1974). My objective in this section is to briefly outline these models and to identify specific strong and weak points of each. For the most part, these models may be discussed as a group, for they represent variations on a basic projection field model proposed earlier by Boring (1933) and others. The dipole model of Julesz was the first to address the local stimulus matching ambiguity problem, but since it fits least well into the projection field format, it will be discussed last.

The basic neural architecture of the projection field model is shown in Figure 3:1. Image information from each eye is projected into a layered, retinotopically organized, neural network, the "projection field," in such a way that at each layer, the coded half images from either eye come together with a slightly different disparity. The image information is coded in terms of elementary features, which are represented by neural activity in "labeled lines." These labeled lines are the axons of projection cells with feature specific receptive fields. The axons pass through the projection field at an angle, making contact with many "match cells" along the way. Match cells have an input from each eye and are excited when the two inputs code the same image

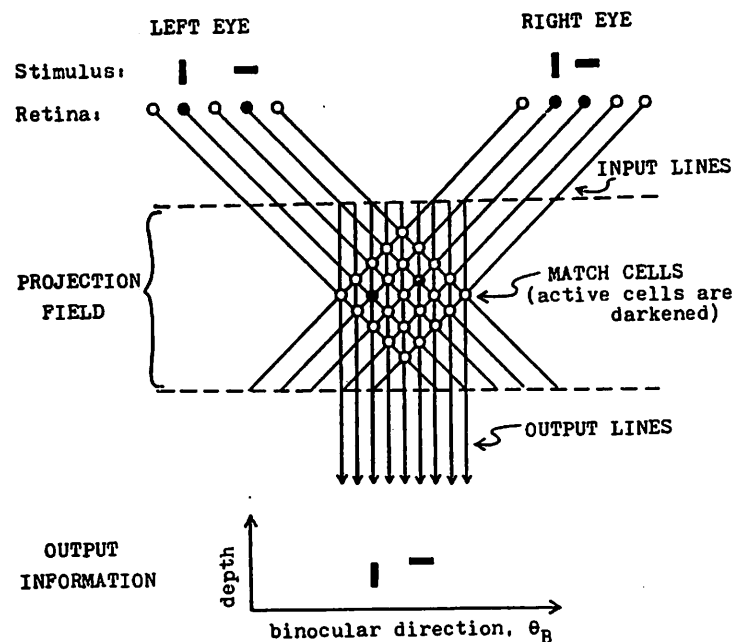


Figure 3:1

Basic projection field architecture.

feature, or "match." When a match cell is stimulated binocularly by a single object in space, the horizontal position of the match cell in the projection field codes the direction of the object, while its vertical position codes depth.

There is also an array of output lines from the projection field. Once active, a match cell will "enable" information flow from an input line to an output line. In this way the projection field is essentially a switching network. Sensory fusion occurs when an active match cell causes input signals on its two input lines to pass onto a single output line. Plasticity of retinal correspondence is simply accounted for by changes in match cell activity in the projection field. It is frequently assumed that the output lines are the axons of the match cells, so that these are also labeled lines coding feature, binocular direction and depth. There are, however, important variations on the match cell concept in the individual models.

One difficulty with the projection field concept is the possibility of multiple fusion. This is illustrated in Figure 3:2, where pairs of similar features presented close together in the two eyes activate four match cells. There are two extra matches, or "ghost images," in this case, and the number increases rapidly as more features of the same kind are added to each stimulus image. The solution which is generally proposed to the ghost problem is that there is

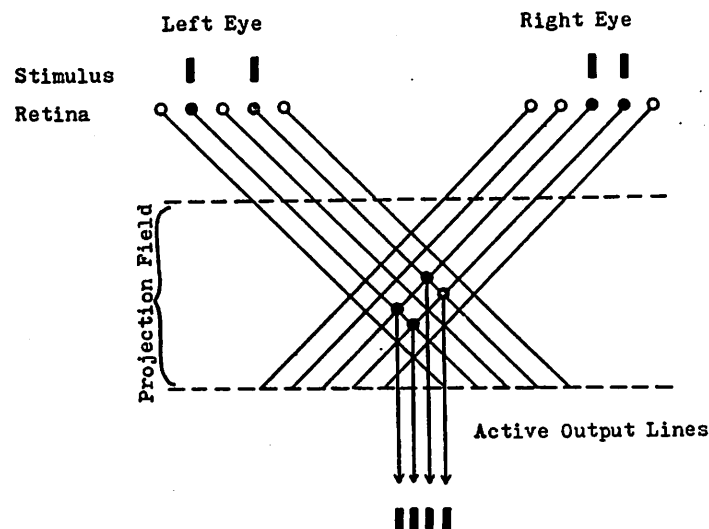


Figure 3:2

When several copies of a given feature are presented to each eye, multiple fusion of each can occur, so that extra "ghost" images should be perceived.

mutual inhibition between match cells which code a given feature in roughly the same direction but at different depths. The inhibition is reciprocal and once one cell becomes active, the cells which it inhibits cannot become active. This depth-domain inhibition, though frequently postulated in the models, is usually not well defined.

The local stimulus matching ambiguity may also be resolved by interactions between match cells. This problem is particularly apparent when the stimulus is a random dot stereogram. Of the many possible matches which occur in the projection field when viewing such a stereogram, only those which correspond to dense surfaces in depth should actually be activated, while others are suppressed by depth domain inhibition. The local matching constraint which leads to a dense surface is that matches of neighboring features in the input images should be made at the same disparity. It is therefore generally proposed that nearby match cells in the same depth plane facilitate one another. When one image feature is fused, the activated match cell stimulates its neighbors, which code the same depth, but slightly different directions, thus increasing the probability that they also will become active.

To summarize, in the basic projection field structure, there are three populations of cells. Two of these code the monocular stimulus images and provide the input to the pro-

jection field. The third population, the match cells within the projection field, code the binocular, or "cyclopean" image. It is proposed that depth domain inhibition between match cells which code the same binocular direction eliminates ghost images, while space domain facilitation between match cells which code the same depth resolves the matching ambiguity.

The recent stereopsis models are variations on this theme. Each is strong in accounting for some aspects of stereopsis, but weak in others. Dodwell seems not to have considered either the ghost image or matching ambiguity problems, and his mechanism for sensory fusion is unclear. For these reasons, the model will not be considered here. Still it should be pointed out that the model has several strong features: the model can respond to monocular as well as binocular stimulation, depth and image feature information are coded in separate cell populations (we shall see that this has advantages), and it is possible for a given binocular input to have variable apparent direction when fused. There is empirical evidence for this variability, and it cannot be accounted for in the other projection field models.

Sperling also does not address the matching ambiguity problem in his model, but his is the only model to include both stereopsis and binocular rivalry. However, the space-domain inhibition in his model which accounts for the spread

of ocular dominance with rivalrous stimuli, has the effect of discouraging the formation of extended regions of match cells activity in one depth plane when the stimulus is not rivalrous. Marr is more precise than the others in specifying appropriate inhibitory interactions in the projection field, but he does not account for sensory fusion. Dev and Nelson emphasize the matching ambiguity problem, and their models can account for stereopsis with a random dot stereogram. However, their approach is to detect a signal in noise by means of averaging in the space domain and differentiation in the depth domain, and it is not sufficiently precise in its treatment of individual features.

These four models will now be outlined along with the dipole model of Julesz. The various authors use rather different terms to identify cell classes within their models, and to identify the structure as a whole. To avoid confusion, all models will be described in the projection field terminology introduced above.

Nelson. Of the models to be considered here, Nelson's most nearly conforms to the basic projection field scheme. Only a few points need to be added to the above description. The input lines to the projection field code simple line shaped features, or, for convenience in describing model response to random dot stereograms, small white and black squares. The depth domain inhibition is directed vertically

through the projection field along lines of constant binocular direction, and the strength of inhibition between two cells decreases with the vertical separation of the cells. Space domain facilitation decreases with horizontal separation in a similar way. Each cell facilitates and inhibits itself, but these recurrent inputs balance and cancel one another. Also the numbers of cells coding different disparities is maximum for zero disparity, which is represented by the middle layer of the projection field, and decreases monotonically with increased disparities. This means that monocular stimuli will usually activate match cells in the zero disparity plane because the greater cell density means there is more lateral facilitation.

Dev. This model is almost identical to Nelson's, but is of greater interest because it has been defined quantitatively and has been simulated on the computer. The principal difference between the models is that Dev introduces a fourth class of cells. These are inhibitory and intrinsic to the projection field, and mediate the depth domain inhibition. Active match cells facilitate their neighbors in the horizontal direction, but do not directly inhibit neighbors in the vertical direction. Instead, each active match cell which codes a particular binocular direction stimulates a single inhibitory cell, which then recurrently inhibits all match cells coding that direction. This means that different

cell populations are responsible for the excitatory and inhibitory interactions, consistent with "Dale's Law." But it also means that the strength of inhibition between match cells does not decrease with their vertical separation, as in Nelson's model. It is not clear that this is a disadvantage for Dev's model, but decreasing inhibition is a feature Nelson uses to explain several psychophysical phenomena.

Since Dev's model is defined quantitatively, it can be analyzed mathematically. For example, it can be shown that once the projection field is in a stable equilibrium state, only one match cell will be active per visual direction. Also system response to changes in the input stimuli will show hysteresis effects like those associated with ambiguous stereograms. This analysis is given in Appendix B, where it is also shown that the failure of Dev's simulations to demonstrate these properties was due to an incorrect choice of network parameters.

Sperling. Sperling's model differs from the other models in several major respects. His model incorporates two projection fields. The first, or primary projection field serves the depth and fusion functions of projection fields in the other models, while the secondary projection field only computes depth, and is responsible for stereopsis with large disparities. Here I shall describe only the primary projection field.

There are three classes of cells within the projection field, and examples of these cells are shown in Figure 3:3. Match cells in Sperling's model are intrinsic to the projection field. Their function is to (somehow) "enable" passage of signals from input cell axons to output cell dendrites. As in other models, match cells coding the same binocular direction reciprocally inhibit one another. However cells coding the same depth do not facilitate one another. There are two types of output cells: one class codes image feature and direction information, while in the other, depth is coded by the magnitude of cell activity. This separate representation of depth and feature information means that the number of output cells is roughly equal to the number of input cells, since relatively few cells are required to code depth. By contrast, the models of Dev and Nelson have many times as many output cells as input cells.

Match cells mediate rivalry as well as fusion. Each match cell is associated with an output cell and is excited by an input from only ~~one eye~~, which codes the same feature as the output cell. In addition, the match cell receives inhibitory inputs from the other eye which codes complementary, or rivalrous features. An active match cell enables information flow only from its excitatory input onto the output cell. Thus the output can be stimulated binocularly only when features in the two eyes match. Each match cell which



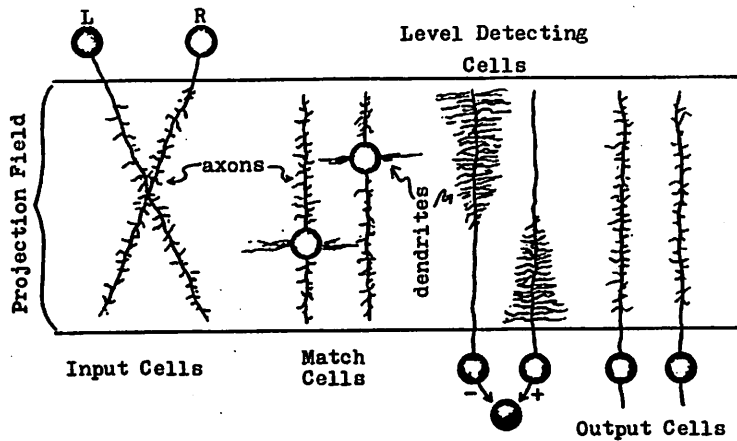


Figure 3:3

Classes of projection field cells proposed by Sperling (1970).

is excited by one eye inhibits nearby match cells in the same depth layer which are excited by the other eye. This inhibition is weak, but the resulting disinhibition of cells excited by the same eye may account for spread of ocular dominance when the binocular stimulus is rivalrous.

Marr. Marr's model has three cell populations, as in the basic projection field model: two populations to code the monocular image and a population of match cells. Unlike the other models, match cells only code depth. Depth is regarded as an "attribute" which is computed in the projection field, and "bound" to the image features. These features are assumed to be more complex than the simple oriented line features used in other models, although the nature of the features is not specified. Coding in terms of many specialized features means less computation is required to resolve matching ambiguities than with a few simple features. Marr does not specify how feature information is coded in the output of the projection field, but it seems that the input lines are meant to pass through the projection field and serve as output lines as well, with the depth attribute added along the way. If this is the case, then output lines are monocular and the model does not explain sensory fusion.

On the other hand, depth computation in Marr's model incorporates several important refinements. As in other models, cells coding different depths reciprocally inhibit

one another. However, in this model, the mutually inhibiting cells fall along the diagonal path followed by input lines. This arrangement insures that only one depth value is associated with each feature, and that each feature of one monocular image is matched with, at most, one feature of the other image. One flaw of other projection field models is their failure to make this "use once" condition explicit. In this model the space domain facilitation is strong for only a short period of time after a match cell becomes active. It functions as a short term bias for neighboring features to match at the same depth, but the bias quickly dies out, so does not force matches when they are inappropriate.

Finally it should be mentioned that a given match cell can be stimulated by any one of a number of features. These features are all very different, so that it is very unlikely that any two features will occur nearby in an image. The match cell is therefore not feature specific, but when it is binocularly stimulated, it is most likely that the stimulation is by the same feature in both images. This system allows a degree of economy: while a match cell is needed for every combination of depth and direction in the projection field, it is not necessary to have one for every feature as well.

Julesz. The familiar dipole model proposed by Julesz (1971) has several structural features which are similar to projection field models. The visual images are represented by binary codes in two parallel arrays of magnetic dipoles.

Sensory fusion occurs when a magnet of one array attracts, and "locks onto", a nearby magnet of the other array which has a complementary polarity. The orientation of dipoles when locked together represents depth. Springs between neighboring dipoles of each array are intended to play the role of space domain facilitation in the projection field models and constrain neighboring dipoles so that they match at the same orientation, or depth. (The actual constraint imposed by the springs is somewhat different from this, as we shall see). Since each image feature is represented by a single dipole in this model, there is no possibility for multiple fusion. Thus there is no problem with ghost images.

One criticism of the dipole model has been that the search for matches must be serial, since each dipole can be in only one orientation at a time. In response to this criticism, Julesz has recently proposed a more elaborate version of the model in which there are many pairs of dipole arrays, and in each, the image from one eye is represented in a different initial disparity relative to the image from the other eye (Julesz, 1976). These array pairs correspond to the depth planes of the projection field models, and their introduction brings with it the possibility of multiple fusion. Following a binocular stimulus presentation, those arrays which happen to have nearly the correct relative disparity will quickly lock together. This initiates a second

"readout" phase of binocular processing, in which all dipoles coding the same feature in different arrays are forced into a common alinement, which is the orientation of the one dipole of the group that is most strongly held by a dipole from the other eye. The alinement mechanism eliminates multiple fusion, so corresponds to depth domain inhibition in the projection field models.

The principal virtue of the dipole model is that it clearly illustrates organizing phenomena in stereopsis. However there are difficulties with the model which should be mentioned. For example, neural units which behave like rotating magnets are difficult to define with precision, and the output from the dipole arrays has not been specified. Also there seems to be a critical flaw even in the physical model. The problem is that the force exerted by a spring increases in proportion to the stretch of the spring. Thus the springs in the model may have very little influence on the relative orientations of neighboring dipoles when they are moderately out of alinement. In particular, there will be a binding force which must be overcome to pull two locked dipoles apart. There will also be critical spring stretch at which its force equals the binding force. The spring mechanism will be powerless to modify dipole alinements when springs are stretched less than this critical amount, while it will prevent any binding which requires that springs be stretched

more than the critical amount. The system behavior will depend on the force constants of the springs. If these constants are large, then all dipoles will be pulled into common alinement over extended areas of the array, as is intended, but around edges of these areas, the transition from one alinement to the other will be gradual, and local fusion will be impossible. On the other hand, if the spring constant is small, neighboring dipoles may not be brought into common alinement, even when they code image features which could be fused at a common depth.

### 3.2. Problems of Projection Field Models

Existing projection field models have a number of problems which should be identified before a new model is proposed. Several of these problems can be resolved by careful definition of cell interactions within the projection field, while others necessitate a restructuring of the projection field itself.

Problem 1: Inefficient image coding. A problem with most projection field models is that too many cells are postulated to serve the stereopsis function. This follows from the assumption that a different cell must exist to code every feature, direction, and depth combination. There must be several hundred complete copies of an image code within

the projection field<sup>1</sup>. As Kaufman points out, it seems unlikely that the number of cells devoted to image coding in the visual system should be multiplied by this factor simply to compute depth information, in view of the fact that stereopsis is an evolutionarily new visual ability, and an ability of seemingly little importance. (About 15% of humans do not have stereopsis or are stereopsis-deficient, according to Julesz, 1971. Those that have good stereopsis still notice very little change in their perception of depth when they close one eye). We might expect such large numbers of cells to be involved in stereopsis if stereoscopic depth were computed in a neural structure which was developed for some other perceptual function, such as computation of depth from motion parallax.

The models of Sperling and Marr illustrate ways the number of cells in the projection field might be reduced. Of these, the most important is the idea that depth and

---

<sup>1</sup>The number of functionally different layers in the projection field equals the number of resolvably different positions a feature presented to one eye may occupy and still be fused with a feature presented to the other eye. The dimensions of Panum's fusional area in the vicinity of the fovea is about 14' arc horizontally, and 7' vertically (Ogle, 1950). Binocular resolution is about 1/3 minute of arc, which leads to an estimate of perhaps 700 resolvably different points at which the features may be fused. The actual number may not be quite this large if resolution is less for large disparities. On the other hand, the number may be many times greater if fusion is possible with a similar resolution over the extended fusional area discovered by Fender and Julesz (1967).

image feature information should be coded in different populations of cells. In Sperling's model, the image features are coded only once within the projection field, and yet sensory fusion is possible since a great many input lines make contact with each output line. Unfortunately his model requires that there be a match cell for each of these input to output line junctions. Marr's suggestion that a single match cell may serve a number of features also reduces the total number of match cells required in the projection field. This economy is counterbalanced however by the fact that the image is coded in terms of complex features, since many distinct populations of specialized feature coding cells are required to code an image.

Three conclusions can be drawn from consideration of efficient coding: features should be simple, so that the number of different types required to code an image is minimal; match cells should not be feature specific, so that a single match cell can mediate fusion of a number of features; and depth and feature information should be coded in different cell populations. Other support for this last conclusion has been given in Chapter 1, Section 4, where depth information was part of the segment representation. Also we should note that perceived depth depends on the degree of eye convergence, as well as image disparities. Neural combination of these types of information is easier if depth associated with disparity is coded separately from feature

information.

Problem 2: Space domain facilitation. If match cells code image features, then lateral facilitation between neighboring cells in the same depth plane should never be so strong that it can activate match cells which do not also receive an appropriate retinal stimulus. If this were to happen, one should "see" a host of ghost images around each real stimulus in the visual field. This lateral activation of match cells frequently occurs in Dev's computer simulations. Strong lateral facilitation also causes match cell activity to be large in the interior of a region which is seen at a constant depth, but small around the edges of such a region. Since level of activity corresponds to vividness of the depth perception, this result is contrary to psychophysical experience. When one views a random dot stereogram, the sensation of depth seems strongest around the edges of central target region.

Arbib (1974) describes an experiment by Jenkins in which a matrix of bright dots was embedded in noise field and presented stereoscopically. When the same image was presented to both eyes, the bright dots were difficult to see, but when the dots were presented with increasing disparity relative to the noise, they became easier to see. Presumably, enhanced visibility does not occur until the background noise is fused. Fusion organizes image features, so that

attention is drawn to just those points which cannot be fused with the noise. In contrast to this, simulations of Dev's model show that it has a strong tendency to suppress isolated points which do not fuse at the same depth as points in a larger surround. This effect would be greatly reduced if space domain facilitation were minimized.

One way to avoid the problems inherent in lateral facilitation is to postulate that the facilitation should be short-lived, as Marr has done. Another possibility is that lateral facilitation should take the form of disinhibition, as in Sperling's model. In any event, these difficulties are largely avoided when depth and image features are coded in separate cell populations. Then lateral facilitation can take place between depth coding cells without activating feature cells. As was suggested in the first chapter, this lateral activation of depth coding cells could account for the appearance of anomalous surfaces in random dot stereograms.

Problem 3: Depth domain inhibition. In most projection field models, inhibition is directed vertically along lines of constant binocular direction. The inhibition is reciprocal and sufficiently strong to prevent simultaneous activity in more than one match cell. Only Dev is explicit about this inhibitory mechanism, and while mutually inhibitory cells were simultaneously active in her computer simulations,

this is not because of any fault in the assumed structure, or because cell behavior was "too linear," but simply because weights associated with the net interactions were not correctly chosen (see Appendix B). However inhibition should not be directed vertically through the projection field but diagonally along the input projection lines, as in Marr's model. The role of inhibition is not to prevent fusion of random dot stereograms in two depth planes at once, as in the Dev and Nelson models, but to prevent multiple fusion of individual image features. In Dev's model, some features may fuse more than once, while others cannot fuse, because of vertically directed inhibition. The inhibition should not decrease with distance, as in Nelson's model.

Problem 4: Stereopsis without fusion. A fundamental assumption of projection field models is that stereoscopic depth perception is the result of sensory fusion. However depth may be perceived when the images appear double (Richards, 1971) and when different features are presented to the two eyes. For example, Frisby and Julesz (1975) obtained stereopsis when differently oriented lines are presented to the two eyes.

Nelson accounts for depth with diplopia in the following way. When a stimulus is presented to one eye, it will activate a match cell in the middle, zero disparity, layer of the projection field, because of natural bias which favors

activation of these cells. If a matching stimulus is then presented to the other eye but at a large disparity with respect to the first, it too will activate a cell in the middle layer of the projection field. If the disparity is not more than about two degrees, the two stimuli may also activate a third match cell which codes their disparity. Since Nelson assumes that depth domain inhibition decreases with distance, and the binocularly activated cell is at an extreme disparity level, the three match cells will remain active. Therefore depth is perceived because of the binocularly activated match cell, while the image is double because of the two active match cells at zero disparity. This explanation cannot be correct, since it implies that one should see three rather than two images.

Sperling's model offers a more satisfactory explanation of depth with diplopia, since he postulates the existence of a secondary projection field which responds to large, widely separated image features, but passes only depth information on to later stages of visual processing. Feature information, which may not be fused, is passed by the primary projection field.

The fact that stereopsis can be obtained with diplopia and dissimilar features can be explained without multiple fusion or multiple projection fields. Depth with diplopia is a special case of depth with dissimilar, unfused features. When a monocular stimulus is presented, we assume

that a match cell in the zero disparity layer is activated because of a natural bias which favors match cells which code small disparities. When stimuli are presented to both eyes, which may be the same or different, lateral facilitation favors activation of match cells which are on the same depth plane and close together. As is illustrated in Figure 3:4, this bias will favor activation of match cells which do not lie in the zero disparity plane. Match cells will actually be activated in the depth layer at which the two bias effects balance.

### 3.3. Scale Factors

An image coding system was proposed in the last chapter which had two basic characteristics: coding was in terms of elementary features which correspond to cells with concentric center-surround receptive field organizations, and there were a number of populations of these coding cells which differ from one another only in the size, or scale, of their receptive fields. A number of rivalry related phenomena could be explained in terms of such a coding system. It will now be argued that scaled coding is appropriate for projection field models of stereopsis as well. I will begin by describing an original experiment I performed during the summer of 1975 in Dr. Julesz's laboratory at Bell Telephone Laboratories. The results of this experiment may be explained

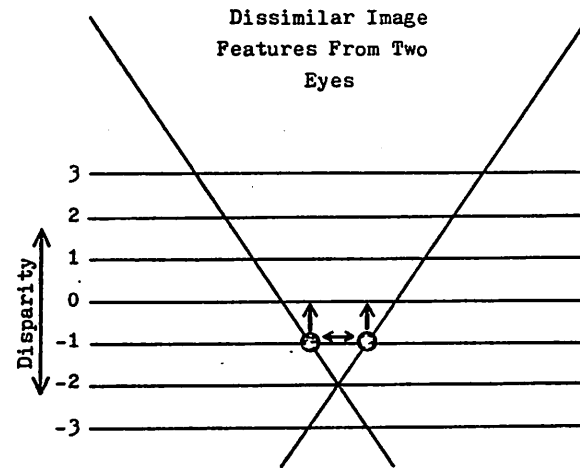


Figure 3:4

This figure shows the interaction of dissimilar features in the projection field. The horizontal arrows indicate the effect of lateral facilitation, which is balanced here by the bias favoring cells which code small disparities.

either in terms of scaled image coding or in terms of a serial stereopsis mechanism. The experiment is best motivated in terms of the serial processing interpretation.

The projection field model of stereopsis accounts for plasticity of retinal correspondence by supposing that all pairs of points which could possibly be corresponding at some moment in time are "prewired" to match cells that uniquely code each binocular combination. However, a particular retinal correspondence is in force only when the appropriate match cell is activated. As has been mentioned, this way of accounting for the plasticity of retinal correspondence is terribly inefficient in that it requires that there exist separate match cells to code each possible retinal combination. In addition, it has been necessary to postulate a complex system of inhibition to insure that only the right number of match cells are active at a time. Let us suppose that instead of coding disparity information in terms of "labeled" cells, this information is coded dynamically in terms of a pattern of neural activity in a general purpose network. For example, it might be coded somehow in terms of an imposed phase relation between activity coding image information from the two eyes. Such a system could be more economical than a projection field in its requirements for specialized cells, and while no model has been proposed of this type, we should consider whether there is empirical evidence in favor of such a model.

As was pointed out in the first chapter, processing in spatially organized networks which code information in terms of organized patterns of neural activity may be characterized as locally serial, since small regions of the net can be organized in only one way at a time, but globally parallel, since many separate regions of the net may be differently organized before a global organization is attained. In contrast to this, processing in the projection field model is inherently parallel before a pattern of match cell dominance is achieved.

Julesz (1964) argues that processing responsible for stereopsis must be serial on the basis of an experiment in which he presented observers with ambiguous random dot stereograms. These stereograms were "biased" to favor fusion of one surface by making 10% of the dots in the target area unambiguous so they could only be fused in that surface. When the stereograms were presented briefly, observers almost always saw the biased surface. Julesz argues that if depth planes were searched sequentially, the search would stop when either the biased or unbiased surface was encountered, since even the unbiased surface gives a good overall match. He concludes that both surfaces must be evaluated at the same time by a parallel mechanism to account for selection of the one which offers a slightly better overall match.

If Julesz is correct, then his experiment offers strong support of a projection field type of model. I believe,



however, that these results can be explained just as well with a serial model. As has been noted, the initial phase of processing in a net which is locally serial may be effectively parallel. Furthermore, once a nearly global organization is obtained, occasional mismatched points will serve as "seeds" for rapid net reorganization.

If stereopsis is actually processed in a projection field structure, then there should be a brief moment after the presentation of an ambiguous stereogram and before patterns of inhibition have been fully established when multiple match cells will be active for each image feature. If this multiple activation could be detected, then we would have strong support for a parallel processing model of stereopsis. The experiment I will now describe was an attempt to detect such activity. The results were negative, but this is not necessarily strong evidence against a projection field model.

Subjects were presented with a sequence of four random dot stereograms in rapid succession on a CRT screen.<sup>2</sup> The

---

<sup>2</sup>Stereograms were 48 x 48 arrays of black and white dots viewed through a prism stereoscope at a distance of 50 cm. A lens was placed in front of each eye which doubled the apparent distance of the screen, while reducing by half the visual angle subtended. With this arrangement, the center to center distance between neighboring dots was 3 min. arc. The stereograms were generated on a PDP-11/20 computer and displayed with special hardware designed by Walter Kropfl on a Hewlett-Packard 1300A video monitor. For more details of this display system see Julesz and Chang (1976).

first and last stereograms simply depicted a uniform surface at the depth of the CRT screen, while the second and third stereograms were ambiguous so that a central target area could be seen either in front of, or behind, the screen. (See Figure 3:5). The first, or pretest stereogram, was presented for .64 seconds, and its function was to alert the observer and provide a reference plane against which he could judge the depth at which he saw subsequent stereograms. The fourth, or posttest stereogram was also presented for .64 seconds and its function was to stop processing of the third stereogram. The second stereogram was presented for 100 msec. or less and its purpose was to initialize match cell activity corresponding to two depth planes, A and B. The disparity of these planes would be selected randomly each trial from a set of twelve combinations (see Figure 3:6). There was a bias (usually 6%) which favored surface A, so the observer almost always fused this stereogram at surface A when the presentation time,  $t_2$ , exceeded about 50 msec. However, with such short presentation times, it is reasonable to expect that any match cells initially activated by surface B could not have been completely suppressed. The purpose of the second test stereogram, stereogram 3, was to detect any residual activity associated with surface B. Thus surface D of this stereogram had the same disparity as B, while surface C had disparity of the same magnitude

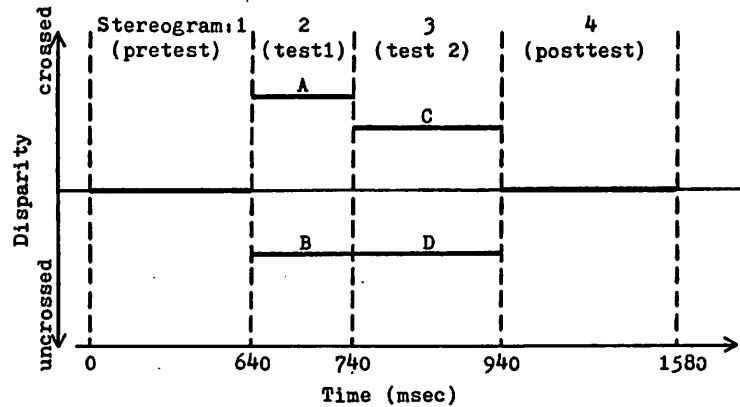


Figure 3:5

This diagram shows the relative disparities and presentation times of the four stereograms of a test sequence. Stereograms 1 and 4 are fused with zero disparity, while stereograms 2 and 3 are ambiguous, so can be fused at either the crossed or uncrossed disparities indicated.

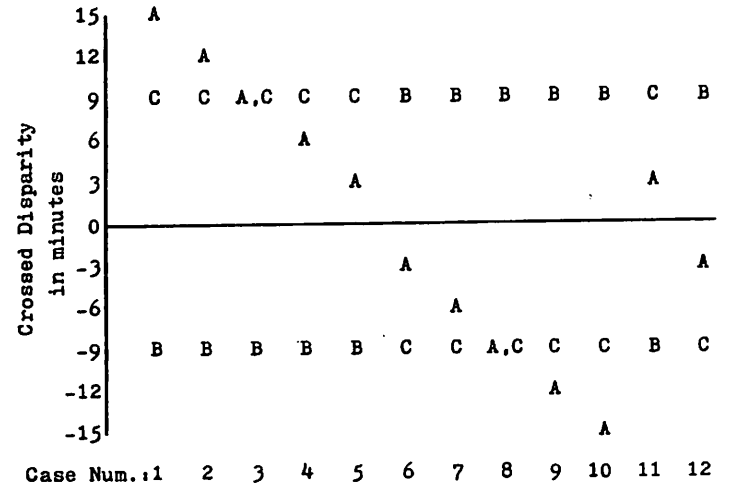


Figure 3:6

This figure indicates the 12 sets of disparity values assigned to A, B, and C. (d=b).

but opposite in sign. Due to constraints imposed by the technique used to generate ambiguous stereograms, surface A was always closer in depth to C than to D. The second test stereogram was presented 200 msec. Following each trial, the observer reported whether he had seen the second test stereogram in front of or behind the screen by pressing an appropriate button. He could press a third button to indicate that he was unsure.

It was anticipated that the observers would fuse surface D more often than C if stereopsis was processed in a projection field, and if there was residual activity associated with the unseen surface B. On the other hand, a preference for surface C was anticipated if processing was serial, since the dominant surface A was closer to C than to D. Only one observer has been extensively studied, and the results of a typical experimental run for this observer are shown in Figure 3:7. This run consisted of 8 trials each of the 12 combinations of value assignments for A, B and C indicated in Figure 3:6. The 96 trials were presented in random order<sup>3</sup>.

It is evident in this figure that perception of surface C is preferred to D. The results are not completely symmetric.

<sup>3</sup>Combinations numbered 11 and 12 were included as controls. In these cases, there was a bias of 3% favoring surface D. Since this should cause D to be fused, if the subject's responses were not consistently D in these cases this would indicate that his judgment of surface depth was poor.

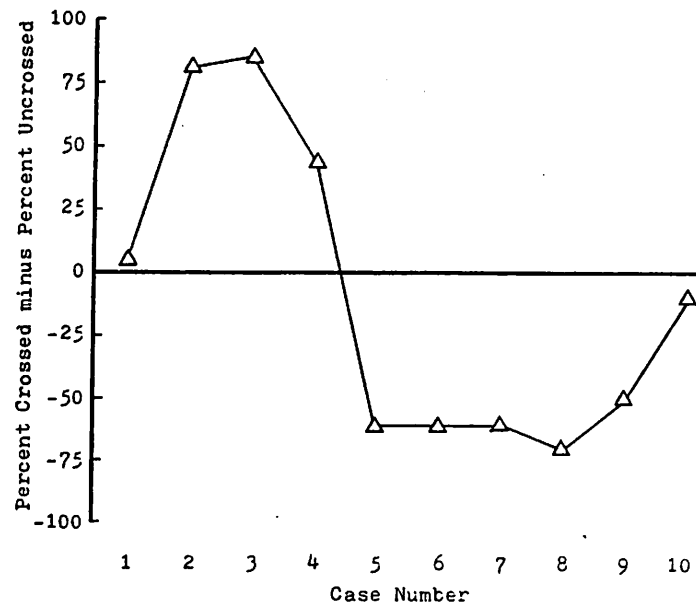


Figure 3:7

This graph shows the combined results of five runs. In each run the 12 cases were presented 5 times in random order. Cases 11 and 12 were controls and are not shown. Each data point shows the number of times a particular case was reported up (crossed) minus the number of times it was reported down (uncrossed) expressed as a percentage of the total number of times the case was presented.

For example, in both cases 5 and 6 the surface with uncrossed disparity is seen even though disparity values are exactly reversed. This effect is common and reflects a natural bias of individual observers to see ambiguous stereograms either always with crossed or always with uncrossed disparity. To eliminate the contribution of this bias, the results of symmetric cases (1 and 10, 2 and 9, etc.) may be combined. Figures 3:8 and 3:9 show results of other experimental runs in this format. (Note that these graphs show the "percent preference" of surface C to surface D as the difference between the percentage of trials in which C is reported and the percentage in which D is reported).

The effect of the presentation time of stereogram 2 on the perception of stereogram 3 is shown in Figure 3:8. When stereogram 2 is not presented at all,  $t_2 = 0$ , we see that surfaces C and D are seen with almost equal probability. But with all non-zero presentation times, C is preferred. Thus no matter how briefly stereogram 2 is presented, no residual activity associated with B is detected.

The strength of the bias in favor of surface C due to the presentation of stereogram 2 can be quantified by determining the number of points which must be made to mismatch in C but not D, so that these two types of bias balance and C and D are seen with equal probability. Figure 3:9 shows the effects of making various percentages of points in C mismatch.

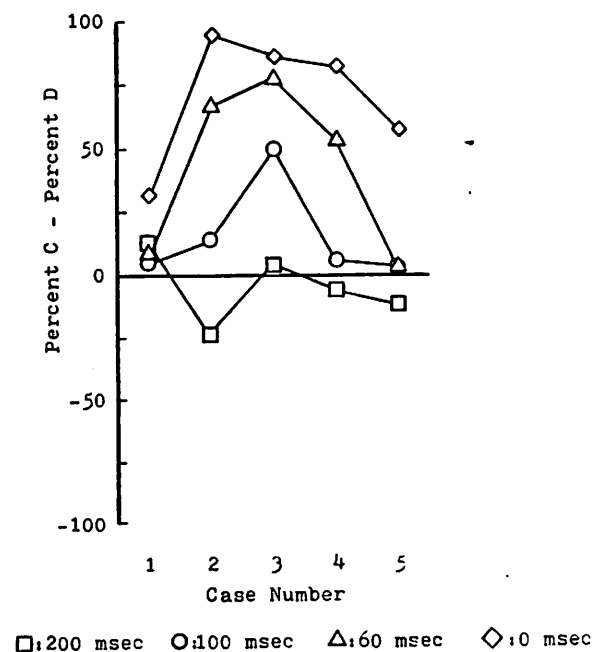


Figure 3:8

The time of presentation of the first test stereogram was varied in the four runs shown here. The second test stereogram was presented for 200 msec in each trial. Again in each run, the 12 cases were presented 8 times each. Here the results of symmetric cases are combined and expressed in terms of the percentage of times C was reported minus the number of times D was reported.

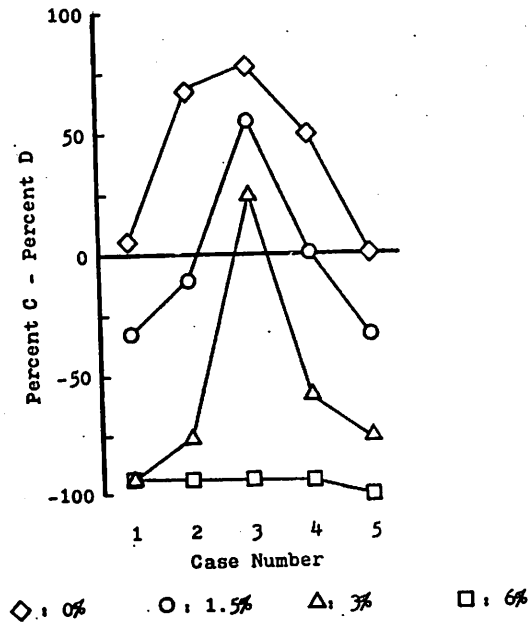


Figure 3:9

In three of these four runs a bias favoring D has been added to the second test stereogram by making a small percentage of points unfuseable at disparity C. A bias of only 1.5% balances the effects of the first test stereogram, except when the disparity of A is the same as the disparity of C (case 3).

The overall conclusion of this study is that the residual cell activity which would indicate the existence of a projection field structure could not be detected. This result is not consistent with Dev's model, and presumably would also be inconsistent with the other models if they were quantitatively defined. However, it is a negative result, so does not rule out the possibility that a modified version of the projection field model could be proposed with which it is consistent. One such modification is the adoption of a scaled code.

Suppose each half image is coded in the manner adopted in the last chapter for the rivalry model. In particular, suppose that the coding elements are cells with concentric center-surround receptive fields and that these elements may be divided into a number of populations, or channels, such that within each channel all cells have about the same size receptive fields, but cells in different channels have different size receptive fields. For each of these size channels in the input code, there will be a population of match cells in the projection field which they contact. Thus match cells of the projection field may also be divided into groups according to the size of the feature they process. Cells in each of these groups are arranged as a projection field, with depth domain inhibition and space domain facilitation, while cells within different groups are, at most, loosely coupled. It is therefore convenient to think of the projection field

as being composed of a number of subprojection fields, each of which is characterized by the size or scale of the image features which it processes.

It may be assumed that the number of cells per depth plane and number of functionally distinct depth planes are about the same in all projection fields. This will mean that the different subprojection fields will cover different proportions of the visual field and different ranges of disparity: the projection field which processes the smallest features will cover a small portion of the visual field near the fovea and disparities up to a few minutes of arc, while the projection field which processes the largest features will cover most of the visual field and disparities up to about two degrees.

Following the extended concept of fusion suggested in the last chapter, we assume fusion may occur in one projection field while rivalry and suppression occur in others. In this way, we can account for the coexistence of rivalry and fusion in the Kaufman's stereogram shown in Figure 2:4b, and in the stereograms devised by Julesz and Miller (1975) in which the two half images of a stereogram are correlated within one spatial frequency band but uncorrelated within another. When these bands are separated by an octave or more, depth may be perceived despite rivalry.

The use of this scaled code in the projection field model need not be defined in greater detail here, since it is

a straight forward extension of the code described for rivalry.

When a stereogram is presented to this system, fusion of large features will occur first since these features are low resolution and can tolerate the most ocular misalignment. Pairs of projection fields are coupled if they process features which differ only slightly in scale. This means that when match cells are activated in one projection field, they will bias match cells which code about the same depth in another. In this way, initial fusion of low resolution features will guide fusion of progressively higher resolution features, and the time required to aline the eyes and resolve the stimulus matching ambiguity will be minimized.

Now suppose that after a stereo image is completely fused in this system, the disparities of some regions within the image are gradually changed. Changes in match cell activity will be required in order to retain fusion. These changes will occur first and most frequently in the high resolution projection field, while activity in low resolution projection fields may need to change little or not at all. In this way, low resolution fusion will remain intact and will help guide rapid refusion in the higher resolution projection fields.

Essentially the same changes in match cell activity will occur when the system is presented with the sequence of stereograms used in the experiment described above. Surface C

in higher resolution projection fields, even though in these projection fields, the bias surface and the ambiguous surface compete.

#### 3.4. Modified Projection Field Model

In view of the fact that several models of stereopsis have recently been proposed, it is appropriate to ask whether yet another model is needed at this time. The model which will be outlined here can be justified on several accounts. First, care has been taken to identify and correct weak points of existing models, as described above. In addition, this model introduces two fairly major modifications to the projection field concept. These include the incorporation of the scaled coding scheme described in the last section, and the division of the binocular projection field into two coupled monocular projection fields which will be explained below. Finally there is practical value to be gained by continuing to refine the theoretical model. This will help guide physiological studies of binocular vision on the one hand and may provide powerful techniques for binocular computer vision on the other.

Before we specify the modified projection field model, there is one final type of empirical data which must be considered. This data has to do with apparent visual direction in binocular vision and is completely inconsistent with the

existing projection field structures.

When disparate image features are fused, the apparent direction of the binocular image is generally not the same as either image viewed monocularly. This displacement phenomenon, or allelotropia, can be approximately accounted for in the projection field model: as an examination of Figure 3:1 will show, the binocular direction associated with each match cell is exactly halfway between the directions of its monocular stimuli. The important point to observe is that the relation between monocular and binocular directions is uniquely determined in the projection field. Unfortunately this strict relationship is not found in binocular vision, as has been demonstrated by Pitblado (1966). Also Charnwood (1951) has found that if one varies the brightness of one image of a stereo pair, the perceived direction of the binocularly combined image shifts in the direction of the brighter half image.

A related point may be made with the help of Figure 3:11. The two eyes are shown fixated on a point 'a' at the edge of a surface A which is in front of and partially occludes another surface B. The points of surface B which lie between b and d are seen by the left eye but are occluded for the right eye. When these points are represented in the projection field, point 'a' will be represented at the binocular direction  $\theta_B$  indicated in the figure. Point d should also

be represented in its binocular direction. If points between b and d are represented at the same depth as d, which might happen due to lateral facilitation, then they too should appear in the binocular direction, even though they are seen monocularly. This leads to the prediction that points between c and b of surface B should appear directly behind points between a and e of surface A, and in particular, that point f should appear to the right of point a. This prediction is contrary to what actually happens when one views overlapping surfaces as depicted in the figure: the point f and other background points which are hidden from one eye are either suppressed, so do not appear in the binocular image, or appear displaced from their binocular direction, so that they appear to the left of point a. These two outcomes may be demonstrated with a simple Julesz random dot stereogram if one adds monocular flag points on either side of the boundary between target area and ground, as in Figure 3:12. When the stereogram is fused, at least one of these flags will be suppressed, which indicates that the dots in the neighborhood of the suppressed flag and in the same half image are also suppressed. Still depth is clearly seen in the region of monocular suppression. In addition, there is no apparent change in the angular separation of the flags in the half image where suppression did not take place, although they are seen at different depths.

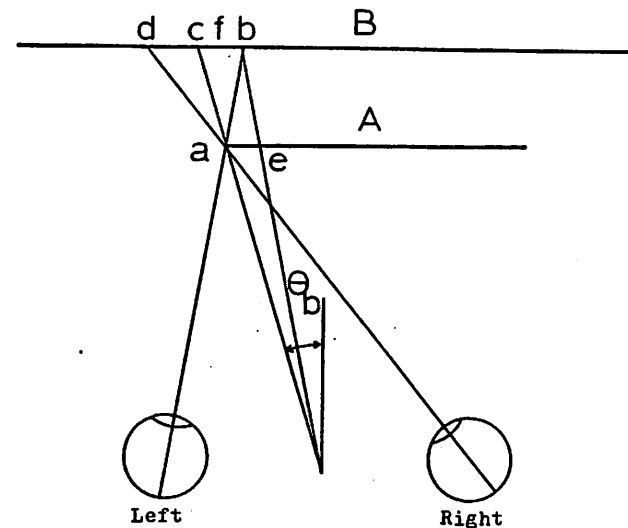


Figure 3:11

If points on the surfaces A and B which are visible to both eyes are fused in a projection field, then these points should appear in their binocular directions. The binocular direction of point a is shown as  $\theta_a$ . The monocularly viewed point should also appear in their binocular directions if they are seen at the same depth as d. This leads to the prediction that f should appear to the right of a.



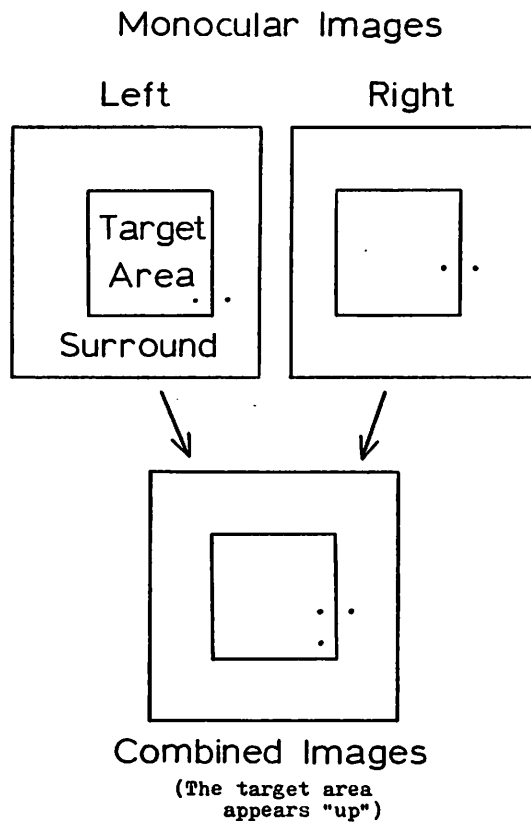


Figure 3:12

This figure shows the placement of monocular flag points on either side of the target-surround boundary in a random dot stereogram. When the stereogram is fused one of the points disappears, indicating a region of suppression.

Two conclusions are drawn from these demonstrations. First, monocular half images must be shifted independently prior to binocular combination. This will account for variability of binocular direction. Second, nearby points of the same half image cannot be shifted by very different amounts. The variability of apparent binocular direction is inconsistent with the assumed geometry of the projection field and constitutes perhaps the strongest evidence against such models.

To account for these observations, we suppose that there are two monocular, or half projection fields, rather than one binocular projection field. The new arrangement is shown in Figure 3:13. Each half projection field is structured in the same way as the projection fields considered up to this point, except that there is an input from only one eye, and the output is monocular. Match cells code a displaced monocular direction and relative depth, as will become clear in a moment. Match cell activity is constrained to one layer of the half projection field by depth domain inhibition. The two half projection fields are coupled so that patterns of match cell activity are coordinated in a way that results in alignments of similar features in the two half images. Sensory fusion occurs outside the projection fields in a layer of binocular cells, as shown in the figure. When images presented to the two eyes are rivalrous, the interactions between half projection fields cause suppression.

Each monocular image is coded in terms of scaled center-surround features in the manner described in the last section, while a number of simple features may converge on the binocular cells to make them responsive to more complex features. To illustrate the operation of this system, assume that Figure 3:13 illustrates only those cells with receptive fields within a small range of sizes and that only the  $l_2$  input line is stimulated. This will cause the activation of a match cell in the middle layer of the left projection field since this layer,  $L_0$ , represents a zero monocular displacement and it is assumed that there is a natural bias favoring activation of cells which code small displacements. The output of the projection field will then stimulate binocular cell  $B_2$ . The active match cell inhibits other match cells in both projection fields. The areas covered by this inhibition are shown in Figure 3:14. Note that the pattern of inhibition in the left projection field insures that the input line activates only one match cell and that it constrains the activation of other nearby match cells by other input stimuli. This is consistent with the observation made in the discussion of Figure 3:12 that the angular distance between nearby monocular flags does not change noticeably even when they are seen at different depths. How they can be seen at different depths will be indicated shortly.

As shown in Figure 3:14, there is a diffuse inhibition

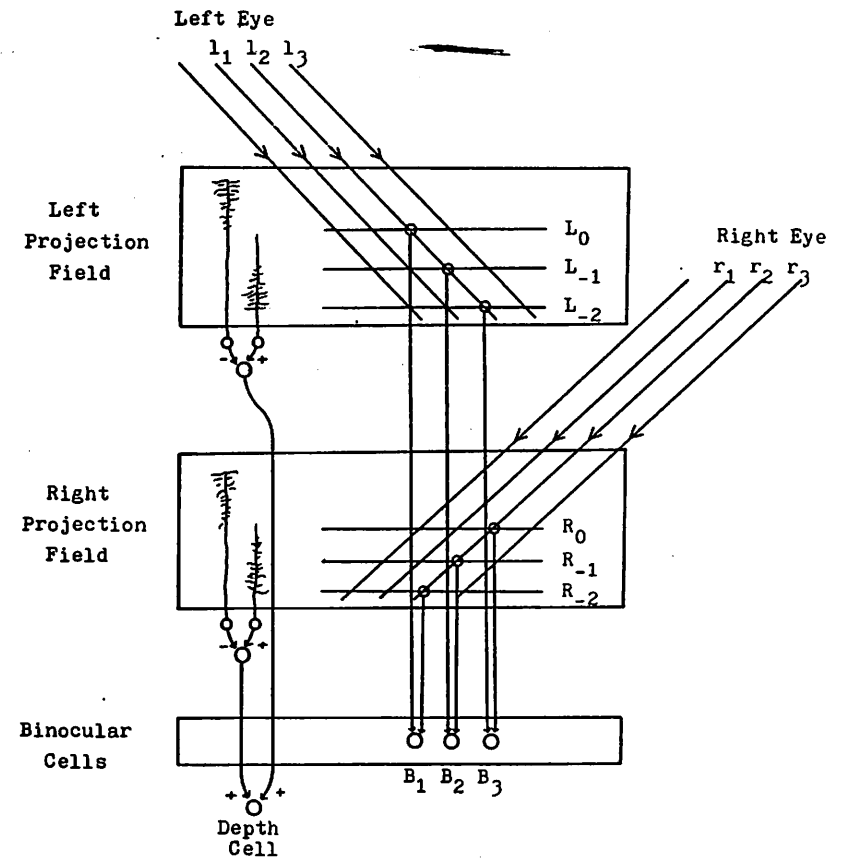


Figure 3:13

Coupled monocular projection fields.

due to an active match cell in one half projection field which is directed at match cells of the other projection field. This inter-field inhibition is much weaker than the intra-field inhibition. Inhibition between projection fields causes suppression when stimuli from the two eyes differ significantly in magnitude, as in the rivalry model proposed in the last chapter. It also serves to aline matching monocular inputs so that they project to the same binocular cell. Alinement results from an assumed decrease in the magnitude of inhibition for cells coding roughly the same direction as the active match cell in the other projection field. With this in mind, suppose input lines  $l_2$  and  $r_2$  are both stimulated. If  $l_2$  activates the match cell in layer  $L_0$  of the left half projection field, then  $r_2$  will be constrained to activate the match cell in layer  $R_2$ , and both outputs of the projection fields will stimulate the single binocular cell  $B_2$ . However several other pairs of match cells could have been excited instead. For example, if  $r_2$  had activated the match cell in layer  $R_{-1}$ , then  $l_2$  would have been constrained to activating the match cell in  $L_{-1}$ , and binocular cell  $B_3$  would be stimulated. This accounts for the variability of binocular direction.

As was stated above, there is a bias within each projection field favoring activation of match cells near the median layers. The pair of match cells which is normally activated

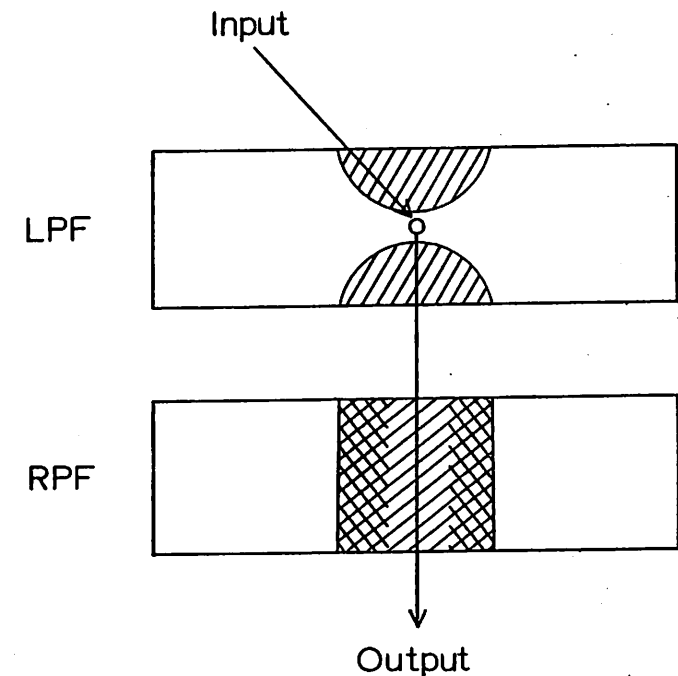


Figure 3:14

Domains of inhibition associated with a match cell.

by stimulation of  $l_2$  and  $r_2$  will be the one for which this bias is balanced in the two half fields. Thus if both stimuli are of the same strength, then the match cells in  $L_{-1}$  and  $R_{-1}$  will be activated, but if the stimulus to  $l_2$  is stronger than the stimulus to  $r_2$ , then the match cells in  $L_0$  and  $R_{-2}$  will be activated. This effect accounts for the observation made by Charnwood (1951) that changes in the brightness of one half of a stereo pair results in a shift of the binocular image in the direction of the brighter half image.

Depth in this model is represented by a separate output from the projection fields. For each binocular direction in each projection field there is a cell which codes the vertical coordinate of match cell activity. This could be achieved in the manner postulated by Sperling in his model, as shown on the left side of Figure 3:13. The outputs for depth cells of the two half fields are summed in a binocular depth cell. With this arrangement, any pair of match cells which can be activated by two matching monocular inputs will result in the same binocular depth. When one input is suppressed, binocular depth is determined entirely by the output of the activated half projection field.

Suppose the stimulus to this system is a random dot stereogram in which a central target area is seen at one depth, while the surround is seen at another. The pattern of match cell activity in this case might look like that shown

in Figure 3:15. There is binocular fusion in the central target area and in parts of the surround. However there is suppression of surround points near the edge of the target area. Again this type of suppression may be observed in actual stereogram with the technique shown in Figure 3:12. In the present model, fusion cannot be maintained over an abrupt boundary due to the broad distribution of inhibition postulated within a half projection field (see Fig. 3:14). An abrupt change in apparent depth is still obtained at the target boundary, because on the target side of the boundary both half projection fields contribute to the depth sensation, whereas on the ground side, one of these contributions drops out abruptly. Similarly if there is an isolated point in the stereogram which cannot be fused at the same depth as its neighbors, local rivalry and suppression will result, so that one projection field does not contribute to the depth signal of that point and the point stands out in depth.

While this model is only outlined at the present time, it is my belief that it has the basic components necessary to account for all the stereopsis phenomena considered in this chapter. The interaction described here in general terms can be made specific with the help of further analysis and computer simulations. It is appropriate also to consider whether any anatomical structure in the visual system could correspond to this paired projection field system. Of course

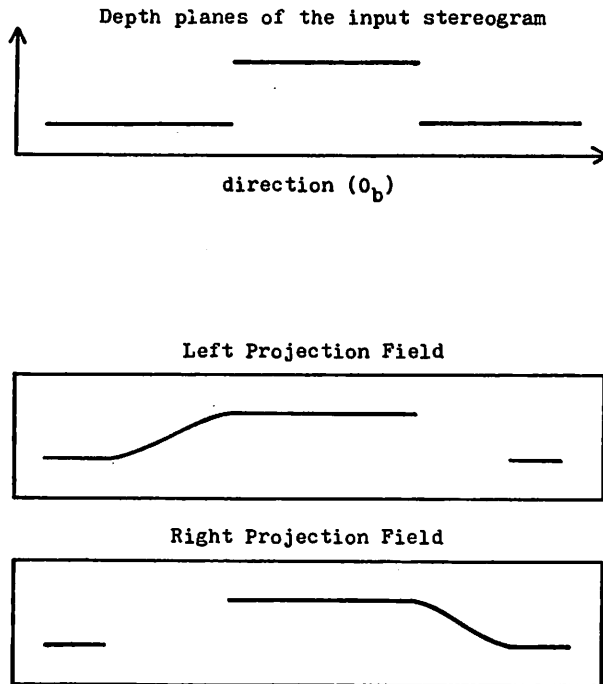


Figure 3:15

The patterns of neural activity within the projection fields are shown for a random dot stereogram stimulus depicting surfaces in depth. Notice that activity is suppressed in one or the other projection field at the boundary between target and surround regions of the stereogram.

it is not necessary that the cells be arranged neatly into layers or that the two half projection fields be anatomically separated in the brain. However it is interesting to note that the structure of this model resembles that of the lateral geniculate body in many respects. Cells within the LGB have center-surround receptive fields and are segregated into lamina which are monocularly activated. These could correspond to half projection fields of the model. In addition there is inhibition between lamina which could correspond to the inhibition proposed here to mediate suppression due to binocular rivalry, and to coordinate match cell activity for sensory fusion. The possible role of the LGB in binocular vision is further explored in Appendix A.

CHAPTER IV  
A PROCESS MODEL OF APPARENT MOTION

Introduction

If a visual pattern is presented very briefly to one position in a subject's visual field and again after a slight delay to another position, then he may actually perceive a single pattern which is continually present over the time of the two discrete stimulations and which moves smoothly from the position of the first to the position of the second. This motion is known variously as apparent motion, stroboscopic motion, optimal motion or beta motion. Perceptually (at the conscious level) it may seem very much like real motion. The same display, with a slightly different temporal or spatial displacement between the two patterns, may result in one of several qualitatively quite different perceptual experiences. For example, one may see the two stimuli as distinct, separate and stationary patterns in the visual field, and still have a sense of motion between them. This type of apparent motion is called objectless motion or phi motion. Alternative perceptions do not include the experience of motion.

Apparent motion is a visual illusion, and yet it may reveal processes which underlie normal visual perception. For

example, apparent motion experiments allow us to examine the stimulus integration processes of perception. The critical difference between phi and beta motion is that two patterns are perceived in the first case, but a single moving pattern in the second. Thus in beta motion, the stimuli are somehow "integrated" into the perception of a single object. Also apparent motion phenomena give us information about the mechanisms in the visual system which are responsible for motion perception. Since, in the case of phi motion, the motion experience is not accompanied by the perceived displacement of any objects, we may conclude that motion detection and stimulus integration processes are separate in the visual system. However, the two processes must interact in important ways, as is shown by various variations on the apparent motion phenomena which will be described here. The trick in developing a model for apparent motion is to account for these process interactions.<sup>1</sup>

The theory I propose accounts for continuity of object perception and stimulus integration in the following way. I suppose that perception of an object is accompanied by the

---

<sup>1</sup>Phi motion is generally considered "paradoxical" because no visible object seems to change position. However another interpretation is possible if one supposes that the motion is perceptually associated with an invisible or unseen object. Such an object might be inferred to be moving across the display screen leaving discrete visible stimulus points like footprints along its path.

development of a corresponding pattern of neural activity within the visual system. The object is perceived continuously only so long as this pattern of activity is maintained. For integration of isolated stimuli, as in beta motion, the pattern of activity must move within the supporting neural structure from the location stimulated by the first pattern presentation to that stimulated by the second. I will describe a model to show how neural structures might support this pattern of moving activity and consider the constraints that must be placed on the stimulus pattern in order to produce motion.

I associate the sensation of motion which accompanies certain stimulus presentations with the activation of motion detecting elements. These fall into two classes. The first class includes neurons in the visual cortex which respond to the motion of optical images over the retina. The second class is made up of neurons in the integration structure postulated above, and which respond to the motion of patterns of activity within that structure. Thus the perception of motion depends both on image motion and the motion of activity patterns. Conversely, these detectors will influence activity pattern motion in the integration structure.

In the first section of this chapter I will describe seven phenomena involving apparent motion. These will indicate the considerable range of motion behaviors which must be

explained by a model. The examples also help motivate the model which is developed in the remainder of the chapter.

#### 4.1. Apparent Motion Phenomena The Basic Paradigm and Examples

I begin by adopting some notational conventions:

$P_1, P_2 \dots$	stimulus patterns
ISI	interstimulus interval
ICI	intercycle interval
S	spatial separation of two patterns
D	duration of a pattern presentation
I	total intensity = integration of light intensity over the area of the pattern (which is assumed small) and the time of display
d	distance between periodic elements of a single pattern
$ISI_\beta$	the ISI which yields beta motion that is most like real motion
$M_1, M_2 \dots$	alternative paths over which motion may be perceived in an ambiguous display

The basic experimental paradigm for producing apparent motion is this: One stimulus pattern,  $P_1$ , is presented as a brief flash of duration  $D$ , which is followed a short time later by the presentation of a second pattern,  $P_2$ , also for a time  $D$ . The delay between the onset times of patterns  $P_1$  and  $P_2$  is called the interstimulus interval, ISI. In this simplest display, the same sequence,  $P_1$  then  $P_2$  is repeated

a number of times at regular intervals. The time between the onset of one presentation of  $P_1$  and its onset in the next presentation is the intercycle interval ICI (Fig. 4:1). One is usually interested only in the sequence  $P_1, P_2$ , so ICI is made much greater than ISI, and each repetition of the sequence is considered a separate event. We should note, however, that events defined in this way are not independent. The condition  $ICI \gg ISI$  means apparent motion will be seen repeatedly from  $P_1$  to  $P_2$  but not from  $P_2$  to  $P_1$ . Yet the subject will learn the repetition cycle and anticipate the appearance of  $P_1$  followed by  $P_2$ . Kolers (1972, p. 162) believes this anticipation is necessary for apparent motion and he notes that after several cycles, when apparent motion has been established, it is possible to delete  $P_2$  from the cycle and still get apparent motion from  $P_1$  to the accustomed location of  $P_2$ .

In general, the patterns  $P_1, P_2$  may be arbitrary and different. When they are different, there is a complex array of apparent motion as  $P_1$  is transformed into  $P_2$ . The motion is constrained in these cases by figural parameters which are not included in the present model. I will consider only cases when  $P_1$  and  $P_2$  are identical except that  $P_2$  is displaced a distance  $S$  relative to  $P_1$ . The patterns are single disks, lines or rows of dots. A basic display in which the figures are single disks is shown in Figure 4:2. It is to be understood that the  $P_1$  and  $P_2$  subfigures are present at different times. In the basic paradigm, the two figures are presented

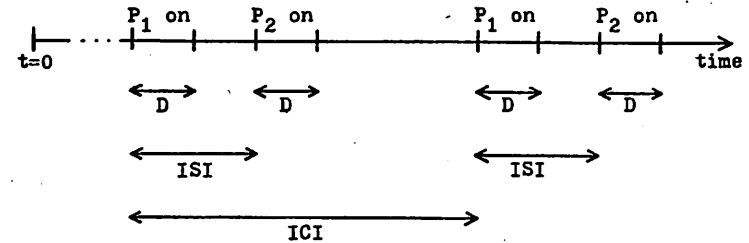


Figure 4:1 The time periods defined for an apparent motion display.

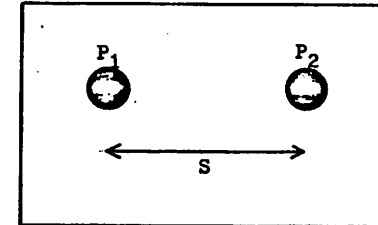


Figure 4:2 The simplest apparent motion display.



monocularly. An important variation is one in which  $P_1$  is presented only to one eye and  $P_2$  to the other.

Apparent motion may also be stimulated with a sequence of patterns  $P_1, P_2, \dots, P_k$  presented one at a time in order. Each figure is related to its predecessor in a fixed way, i.e.  $P_n$  is the same pattern as  $P_{n-1}$  but displaced a distance  $S$ . Here ISI refers to the time between the onset of one pattern,  $P_n$  and the next  $P_{n+1}$ . We will examine a sequential display of this sort in which the basic figure element is a dot as in Figure 4.3.

An important variant of this sequential display is one in which each pattern is a row of identical figures. Figure 4.4 shows patterns  $P_1, P_2, P_3$  and  $P_4$  of such a sequence. Each pattern is a horizontal row of dots and pattern  $P_i$  is displaced a distance  $S_H$  to the right and  $S_V$  down relative to pattern  $P_{i-1}$ . New dots are added to the left end of the row as dots pass off the screen to the right. This display is ambiguous since the observer may perceive apparent motion over one of several paths  $M_1, M_2, \dots$  as shown in the figure for a single dot. All dots appear to move in parallel except under certain conditions which will be noted later. Motion over path  $M_1$  occurs when a dot in the row presented at time  $T_N$  appears to move to the position of the nearest dot in the row presented at time  $T_{N+1}$ .

Consider now a two disk display of the sort shown in

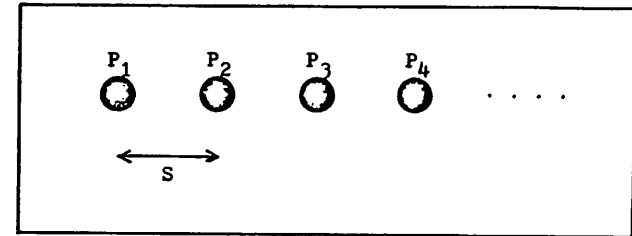


Figure 4.3 A sequential apparent motion display.

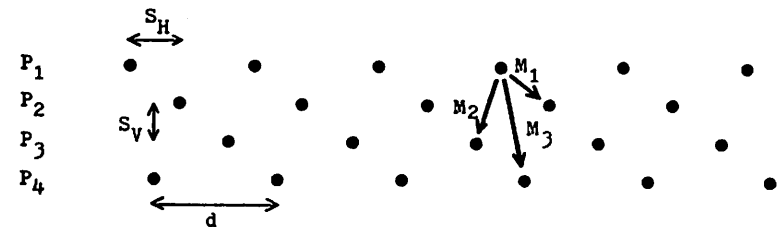


Figure 4.4 The first four patterns from an ambiguous periodic display.

Figure 4.2. Whether or not apparent motion is observed with this display will depend on the values given the parameters ISI, S and I. D, it turns out, is not a critical parameter in the case of small, briefly presented disks, since the visibility of the disk is related to the integral of the light intensity over the area of the disk and over the presentation time D. This is the parameter I.

Given fixed but appropriate values of S and I, four qualitatively different perceptions will be obtained with different ISI values, as mentioned in the introduction. We may distinguish three "critical" ISI values,  $ISI_1$ ,  $ISI_2$  and  $ISI_3$ , such that  $ISI_1 < ISI_2 < ISI_3$  and

- if  $ISI < ISI_1$ , one perceives both patterns  $P_1$  and  $P_2$  as if they were simultaneously presented with no sensation of motion.
- if  $ISI_1 < ISI < ISI_2$ , one perceives the two patterns as distinct and simultaneous, and yet there appears to be motion between them. This is phi motion.
- if  $ISI_2 < ISI < ISI_3$ , one perceives beta motion, i.e., a single pattern is seen to move from the location of  $P_1$  to the location of  $P_2$ .
- if  $ISI_3 < ISI$  one perceives distinct, stationary patterns appearing in succession.

This division of the perceptual experiences into four categories is actually a convenience more than an absolute rule. There is a gradual rather than an abrupt change in the qualitative nature of perception as ISI is changed slowly from 0

to a value greater than  $ISI_3$ . In addition, for a given ISI, I, and S, the type of perception is always the same. For example, for a given sequence of presentations of  $P_1$  and  $P_2$ , one may see phi motion for several cycles, then beta motion, then phi motion again, and so on. The a truer description of perception in terms of ISI would be a set of probabilities. The effect of varying parameters I and S is given by Korte's laws which will be described below.

We now turn to descriptions of several apparent motion phenomena which will later be explained in terms of the model.

Phenomenon 1: Korte's Laws. A theory and model of apparent motion will have to account for the four qualitatively distinct perceptions described above: simultaneity, phi-motion, beta-motion, and succession. It will also have to explain how the probability of obtaining any one of these perceptions depends on the parameters I, S and I. There is controversy in the literature concerning just what this dependency is. For the case of beta motion, the most widely accepted description is given by Korte's "Laws" which may be summarized in the following equation:

$$\frac{ISI \cdot I}{S} = K \quad \text{where } K \text{ is constant.}$$

This equation is based on data obtained by Korte as follows (see Kolers, 1972, p. 21): First, initial values of S and I

were chosen and ISI was varied until subjects reported seeing optimal beta motion. Then S and I were changed systematically by small amounts, and new values of ISI determined which yielded optimal motion.

Notice that the above equation implies that for a given stimulus intensity I, there is a fixed velocity  $V_{op}$  which will yield optimal beta motion.  $V_{op}$  increases with I:

$$V_{op}(I) = \frac{S}{ISI} = \frac{I}{K}$$

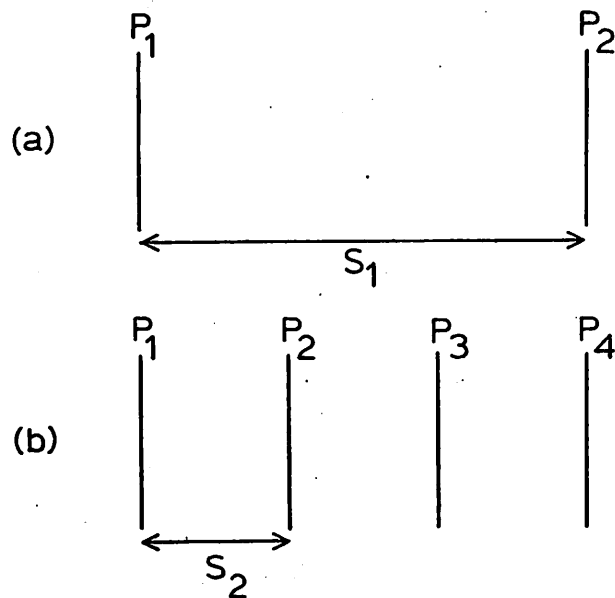
This interpretation assumes motion occurs only over the period ISI and is of constant velocity. These assumptions have not actually been substantiated for the two pattern display. One should bear in mind also that Korte's Laws are intended only to give parameters for optimal motion, that is, motion which is most like real motion. Beta motion can be observed over a range of other velocities.

As mentioned above, Korte's Laws are not universally accepted by psychologists. Kolers (p. 22) interprets data of Neuhaus to arrive at an alternative formulation. This data implies that velocity  $V_{op}$  depends on I in the way described by Korte, but that it is also proportional to S. Thus from the Neuhaus data, we may conclude that it is not the velocity but the total time, ISI, which is constant (for fixed I) in all cases of optimal beta motion.

Phenomenon 2: Sparse stimulus sequences and loss of beta-motion. An experiment by Kolers (p. 36) suggests that there is a fundamental difference between beta motion with two patterns  $P_1$  and  $P_2$  and motion with a sequence of patterns  $P_1, P_2 \dots P_k$ . The patterns Kolers used in this experiment were vertical lines. A display would consist of the sequential presentation of  $2^k$  such patterns where k is an integer between 1 and 10. Examples with  $k = 1$  and 2 are shown in Figure 4:5. The displacement of patterns for each value of k, i.e.  $S_k = S_1 / (2k - 1)$  was chosen so that the distance between the first and the last pattern was always the same.

It was discovered that smooth beta motion could be obtained with just two lines,  $k = 1$ , or with a relatively dense sequence of lines,  $k \geq 5$ , but not for sparse sequences  $k = 2, 3$  or 4. To put this result another way, suppose  $ISI_1$  is the interstimulus interval which results in optimal beta motion with the two lines a distance  $S_1$  apart, shown in Figure a. If stimulus patterns are added between the two as in Figure b by letting  $k = 2$ ,  $ISI_2 = ISI_1 / 3$ ,  $S_2 = 1/3 S_1$ , the result is not an enhancement of the quality of apparent motion, but a considerable degeneration of the motion.

Conversely, if  $ISI_2$  is the interstimulus interval which results in optimal beta motion between just the first two lines of Figure 4:5b, this choice of ISI would not result in beta motion when patterns  $P_3$  and  $P_4$  are added to the sequence.



$$S_2 = S_1/3$$

Figure 4.5  
Sequences of 2 and 4 patterns.

This result indicates that apparent motion is not simply the result of stimulating motion sensitive elements, since one would expect multiple stimulation within the receptive field of such elements to be better than just two stimulations.

Phenomenon 3: Dense sequences and "retrospective memory readouts." In the case of dense sequences of patterns convincing beta motion can be obtained, but there are new phenomena which deserve attention. I describe here an experiment by Ross (1972) in which a sequence of dots is displayed on a CRT, see Figure 4.6. Each dot is an individual pattern in our notational convention, and each pattern is displaced a small distance  $S$  to the right of its predecessor.

Suppose that an interstimulus interval is chosen which results in smooth beta motion. Ross observes that while a single dot may appear to move smoothly over the central region of the display, at either end that perception breaks down and one perceives instead a number of isolated, stationary dots. It would seem that several dots have to be presented in this display before motion is suggested and a moving dot perception attained. This accounts for stationary dots on the left. But why doesn't the integrated perception endure to the end of the line? The number of isolated dots on the right end is roughly the number presented in the final 100 msec. of the display. At first glance, there would seem to be a paradox here - how does the moving dot perception know enough to break down 100 msec. before the end of the sequence?

Ross suggests the following explanation: perception at a given moment is based not just on the stimulus at that moment but on all stimuli received within a time window, i.e. within the previous 100 msec. Thus there is a 100 msec. short term memory of the stimulus input. In this display, the stimulus within any window is a sequence of dots which may be integrated into the perception of a single moving dot, or perceived individually as a row of dots. The movement perception is maintained until the end of the row when the next anticipated dot fails to appear. At that time, there is a shift in perception from moving dot to the isolated dots still in memory, in effect, a "retrospective memory readout."

Phenomenon 4: Hysteresis and the Land Square. A fundamental question about apparent motion is whether it is the result of direct sensation, that is, the result of stimulation of motion selective elements in the retina, or whether it is the result of processes which organize the visual input in the cortex. One way to experimentally distinguish between these types of processing is to look for hysteresis or fatigue phenomena.

Hysteresis should be expected if organizing processes are involved as a consequence of the stability of particular organizations once formed. To illustrate this point, suppose there is a stimulus pattern which can be organized in either of two ways, depending on the value given to some parameter.

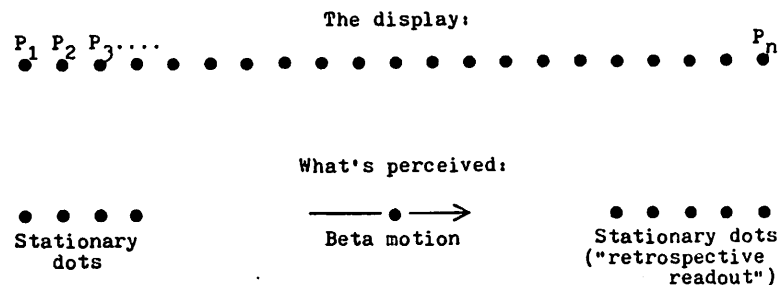


Figure 4.6 Beta motion and "retrospective memory readout" with a dence sequence.

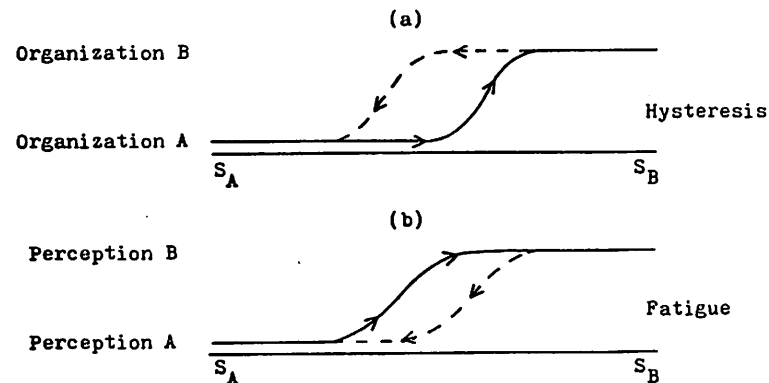


Figure 4.7

Then if  $S_A$  is a value of that parameter which favors organization A, and  $S_B$  favors organization B, the actual organization, while  $S$  is changed from  $S_A$  to  $S_B$ , might be as shown by the solid line in Figure 4:7a, while the dashed line shows the organization as the parameter is changed in the other direction, from  $S_B$  to  $S_A$ . Here, organizations correspond to perceptions. On the other hand, if direct sensation is responsible for a perception, then the reverse of hysteresis may occur, due to fatigue of the elements responding to a particular stimulus, as in Figure 4:7b. An interesting feature of the Land Square display which will be described here can be interpreted as hysteresis phenomena, and thus may be taken as evidence that apparent motion involves organizing processes.

A Land Square (Ross, 1972) is generated by plotting a dense sequence of dots on a CRT in a square pattern, as shown in Figure 4:8. In effect, each edge of the square is a row of dots, like that of example 3. The phenomenon of interest now is obtained when an ISI is chosen which results in beta motion of four single dots, one along each edge of the square. As before, there is a breakdown of the moving dot perception at the beginning of the row and those dots which are plotted in the final 100 msec. are seen at the end of the row. Now, however, we focus our attention on a new feature of the display, that is, the corners of the square do not appear to

meet at right angles, but are "pinched," as shown in Figure 4:8b.<sup>2</sup> We may interpret the pinched corners as an hysteresis effect, if, as shown, the moving dot perception travels beyond the end of the row of supporting stimuli before it reorients 90° for motion along the next edge. The model will have to account both for overshoot and the breakdown of the moving dot perception at the corners.

Phenomenon 5: Apparent motion in spatially periodic ambiguous displays. Here we consider the perception which results when one observes a spatially periodic pattern of the sort shown in Figure 4:4. As noted previously, the display is ambiguous, since for a given  $S_H$ ,  $S_V$  and  $d$ , several different perceived velocities are presumably possible. It is useful here to define "motion parameters" associated with these paths. Let  $s_i$  be the distance between successive dots along path  $M_i$ , let  $t_i$  be the time between their presentations:  $t_i = i \cdot \text{ISI}$ , and let  $v_i$  be the velocity of apparent motion along the path:  $v_i = s_i / t_i$ .

While motion is possible over any one of several paths, we may anticipate that for certain values of the "display parameters,"  $S_H$ ,  $S_V$ ,  $d$  and  $\text{ISI}$ , motion over certain paths will strongly be preferred. Experiments confirm this expect-

---

<sup>2</sup>Ross does not actually describe the shape of the "pinched" corners, so Figure 9b is somewhat speculative.

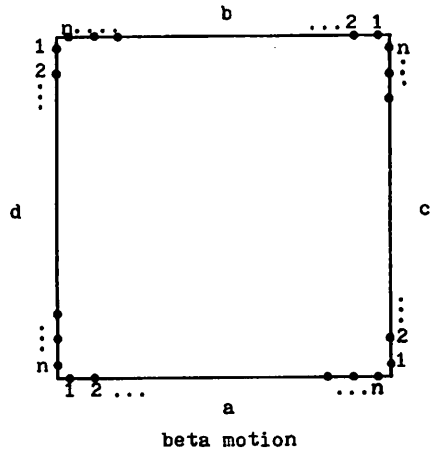


Figure 4:8a Order of dot presentation:  
 $a_1 b_1 c_1 d_1 a_2 b_2 c_2 d_2 \dots a_n b_n c_n d_n a_1 b_1 c_1 d_1 \dots$

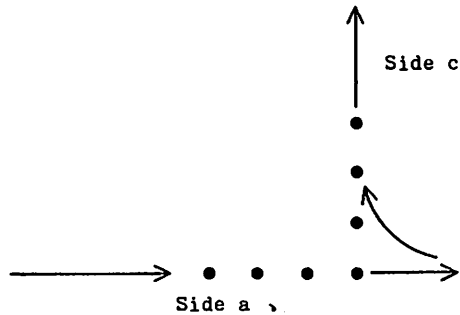


Figure 4:8b A "pinched" corner of the Land Square, interpreted here as overshoot of a moving dot perception.

tation. (Experiments described in this section are my own).

For a given assignment of values to the display parameters, an observer would usually see motion over a single path. A small change in any one of the parameters could then result in a slight change in the direction and/or speed of motion, still over the same path, or in a more or less abrupt transition to the perception of motion over another path. However, for certain choices of the display parameter values, motion over two paths might appear either to alternate or to coexist. In the latter case, motion was not paradoxical; no dot seemed to move in two directions at once. Rather the brightness of individual dots would appear to vary, and while the dots moved in one direction, waves associated with this brightness modulation would appear to move in another. Displays which yield motion in two directions were called "transition points," since a small change in any display parameter would strongly diminish the perception of one or the other of the motions.

The above phenomena may be accounted for in the following way. The ambiguous apparent motion display is a stimulus simultaneously for motion in many directions. However, for particular display parameters, it is a stronger stimulus for motion in one of these directions than for the others, and it is that motion which is actually perceived. We may represent the effective strength of the stimulus for a particular direction of apparent motion by E, and hypothesize that E

is a function of the motion parameters  $s$ ,  $t$  and  $v$ . Thus when motion is perceived over path  $M_i$ , the following inequality should hold:

$$(Eq. 1) \quad E(s_i, t_i, v_i) > E(s_j, t_j, v_j) \text{ for all } j, j \neq i.$$

On the other hand, in "transition" cases where motion is perceived simultaneously or alternately over two paths,  $M_i$  and  $M_j$ , we have:

$$(Eq. 2) \quad E(s_i, t_i, v_i) = E(s_j, t_j, v_j).$$

In order to determine the functional form of  $E$ , many transition points were experimentally determined for two subjects. These points were obtained for a wide range of  $s$ ,  $t$  and  $v$  values, and were assumed to represent conditions under which equation 2 was satisfied. The data reported here are for one of these subjects, but results for the second subject, for whom less data were obtained, are consistent with data reported.<sup>3</sup>

The procedure for locating a transition point was this. Initial values were assigned to the display parameters  $S_V$ ,  $S_H$ ,  $d$  and  $ISI$ , then the stimulus was presented continuously

---

<sup>3</sup>Results reported here are of a preliminary study. A more extensive followup study is planned.

for a period of several minutes. The subject could control  $ISI$  without interrupting the stimulus presentation, and, by systematically varying this parameter, could determine that value of  $ISI$  which yielded what he judged to be equal strength apparent motion in two directions. The values of  $S_V$ ,  $S_H$ ,  $d$  and  $ISI$  which corresponded to this transition point were recorded, and the above procedure was repeated with new initial values assigned to these parameters.

It should be noted here that the subject made his observations while fixating a point displayed in the center of the screen. Motion sensitivity for foveal and peripheral vision proved to be significantly different, and stimulus points within approximately two degrees of the fixation point would frequently appear to move slower and over a different path than points in the periphery. The data reported here are for peripherally perceived motion. It should also be noted that the subject could only select  $ISI$  values which were integral multiples of 2.5 msec. This allowed only rough determination of  $ISI$  values at transition points, which ranged from 7.5 to 30 msec.

The data were analyzed in terms of the motion parameters  $s$ ,  $v$  and  $t$ , which were computed from the stimulus parameter values at the observed transition points. The first graph in Figure 4:9 shows 27 data points obtained for transitions between motion over paths  $M_1$  and  $M_2$  in terms of their



respective velocities. An additional 7 data points were obtained for transitions between  $M_1$  and  $M_3$ , and  $M_2$  and  $M_3$  and are shown in a similar way in the second graph of Figure 4:9. Two relationships emerge when the data are represented in this way. First, for transitions between  $M_i$  and  $M_j$ ,  $i < j$ , it appears that  $v_j = v_i - c_{ij}$ , (with  $c_{ij}$  positive), as is suggested by the solid lines of slope 1 in the figure. Scatter of points about these lines can be reasonably attributed to probable error in ISI determination, as noted above. Second, the intercepts  $c_{ij}$  appear to be logarithmically related to the ratio  $t_j/t_i$ . ( $c_{12} = 5.9$ ,  $c_{13} = 9.4$ ,  $c_{23} = 3.4$ ). Thus a general expression for transition between paths  $P_i$  and  $P_j$  is this:

$$v_j = v_i - A/B \ln(t_j/t_i)$$

which may be rewritten in the symmetric form:

$$(Eq. 3) \quad A \ln(1/t_i) - Bv_i = A \ln(1/t_j) - Bv_j$$

$A$  and  $B$  are constants and must have the same sign to account for the sign of  $c_{ij}$ . The three solid lines in Figure 4:9 were obtained from this expression by finding the single value of the ratio  $A/B$ , 8.52 cm/sec., which resulted in a least squared error fit of the lines to the three groups of data.

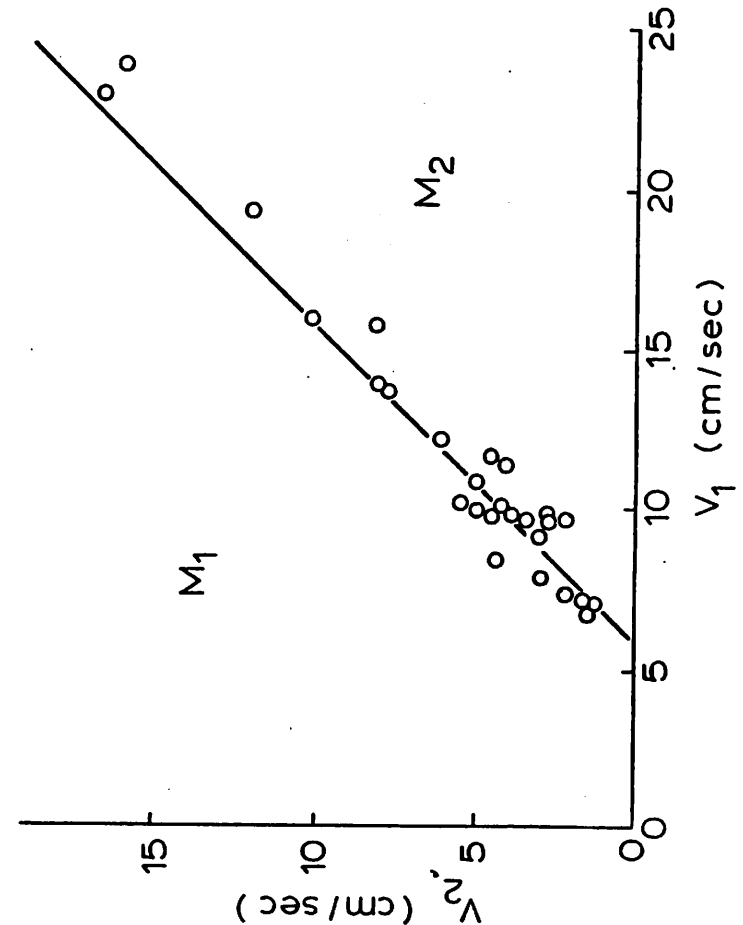


Figure 4:9 See next page for caption.

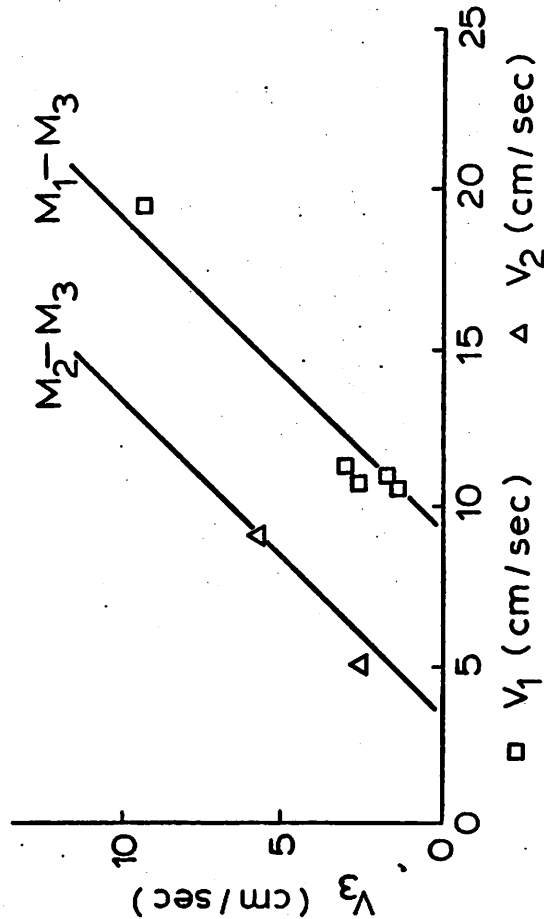


Figure 4.9 The data shown here indicates the values of  $v_i$  and  $v_j$  for various display conditions that were judged by one subject to be transition points for which motion over paths  $M_i$  and  $M_j$  appeared equally compelling. Transitions between  $M_1$  and  $M_2$  were observed most frequently, and to avoid confusion, these points are plotted separately on the preceding page. A velocity of 1 cm/sec on the CRT screen corresponded to an angular velocity of approximately 1.2 deg/sec for the observer. The solid lines are obtained from equation 3 with  $A/B=8.52$  cm/sec.

It now follows that E may be expressed as a function of a single argument:

$$(Eq. 4) \quad E(s, t, v) = \hat{E}(A \ln(1/t) - Bv + C).$$

It was also observed that motion  $M_i$  dominated when  $v_j > v_i - C_{ij}$  for  $i < j$ . Therefore, to be consistent with inequality (1),  $\hat{E}$  must be a monotonically increasing function of its argument with B positive. The constant A must have the same sign as B, as noted above. While it is not possible to determine the explicit relationship of  $\hat{E}$  to this argument from the present data, two possible relationships are of interest. The first and simplest is that  $\hat{E}$  is just the identity function, so that

$$(Eq. 5a) \quad \hat{E}(s, t, v) = A \ln(1/t) - Bv + C.$$

The second is that  $\hat{E}$  is proportional to the exponential

$$(Eq. 5b) \quad \hat{E}(s, t, v) = \frac{C'}{t^A} e^{-Bv}$$

Here,  $C' = \ln C$ . Notice that in equation 5a, the contributions to stimulus strength related to time and velocity respectively are additive, while in equation 5b, these contributions are multiplicative. Future experiments might focus on determining whether either the additive or multiplicative property is

correct. At the same time, brain theory studies may help show which form of  $\hat{E}$  has the most realistic neural implementation. We may note, however, that according to equation 5b, the strength of the stimulus is a power function of the frequency of stimulation  $f$ ,  $f = 1/t$ . This is intuitively reasonable. Moreover, equation 5a has the difficulty that the stimulus strength,  $\hat{E}$ , becomes negative for sufficiently large  $v$ .

Finally it is worth observing that the general form of  $E$  derived here, equation 4 (and hence the particular forms in equations 5a and 5b) is consistent with Korte's laws. Recall that the apparent motion display for which these laws were developed consists of only two stimulus points which are presented sequentially with temporal separation  $ISI$ , spatial separation  $S$  and intensity  $I$ . According to Korte's laws, the strongest, or "optimal," sense of apparent motion is evoked when  $ISI \cdot I = K \cdot S$ , where  $K$  is a constant. Thus, for a fixed  $I$ ,  $ISI$  is proportional to  $S$ . This indicates that there is an optimal velocity,  $v_{op} = S/ISI = I/K$ , for apparent motion which is independent of dot separation  $S$ . Noting that  $v = s/t$  in equation 4, we may determine that value of  $t$  which yields maximum  $E$  for a given value of  $d$  by setting the derivative of  $E$  with respect to  $t$  equal to zero. This yields  $t = Bs/A$ , again indicating constant velocity and corroborating Korte's laws.

The role of intensity in determining effective stimulus strength was not investigated in the experiment reported here. However we may speculate, on the basis of the correspondence between equation 4 and Korte's law, that the ratio  $A/B$  is, in fact, proportional to display intensity.

Sperling (1975) reports results of another type of apparent motion experiment which lends support to the above analysis. The display used for this experiment was a sequence of dots plotted one at a time in a row, as in Figure 4:3. This display is characterized by two parameters,  $ISI$  and  $S$ . For various combinations of values assigned to these parameters, the subject judged the "quality of perceived motion," which he rated on a scale of 0 (no motion) to 10 (smooth, continuous motion). Results of the experiment for one observer are shown in Figure 4:10.

While the experiment Sperling describes and my own are very different in concept, it seems likely that "quality of perceived motion" reported in his experiment should be related to the "stimulus strength,"  $E$  derived above. In fact, a qualitatively good match can be obtained between Sperling's results and a slightly modified version of  $E$ :

$$(Eq. 6) \quad \hat{E} = \frac{C}{(t+\delta)^A} e^{-B \frac{x+\epsilon}{t+\delta}}$$

Here very small quantities  $\delta$  and  $\epsilon$  are added to  $x$  and  $t$

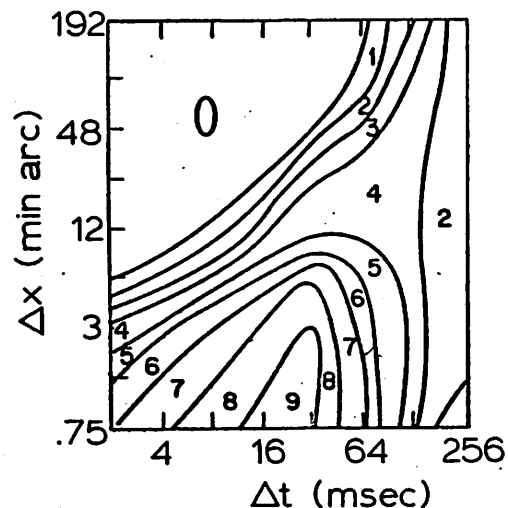


Figure 4:10

Estimates of the quality of perceived movement on a scale from 0 (no motion) to 10 (smooth continuous motion). This graph was reported by Sperling (1975) and is based on experiments by Miriam Kaplan.

respectively. This modification means that  $\bar{E}$  does not become arbitrarily large when  $x$  and  $t$  are made very small. It is reasonable that such a limit should exist in a biological system. If  $x$  and  $t$  are made equal to  $S$  and  $ISI$ , then the values of  $E$  which correspond to Sperling's results are as shown in Figure 4:11. Here the ratio  $A/B = 8.52 \text{ M/sec.} = 10.2 \text{ deg/sec.}$  as obtained in my experiment. Values of other parameters could not be obtained from my data and were chosen to give a reasonably good match to Sperling's results. In particular,  $\gamma = 4 \text{ min. arc}$ ,  $\delta = 1 \text{ msec}$  and  $A = \frac{1}{2}$ . It should be noted that these values of  $\gamma$  and  $\delta$  are below the limits of spatial and temporal resolution for peripheral vision, so represent reasonable lower limits on  $x$  and  $t$  values in equation 6. A similar degree of match between Sperling's results and equation 6 can be obtained for a range of values of  $A$ , so the choice  $A = \frac{1}{2}$  made here is not critical. The value of  $C' = 50$  was chosen so that the maximum  $E$  was 9 as in Sperling's results.

The differences that exist between Figure 4:10 and 4:11 may be due to any one of several display related factors. The size and brightness of individual dots may differ in the two experimental displays. If, as has been suggested,  $B$  is inversely related to dot intensity, then increasing intensity should displace curves shown in Figures 4:11 to the right. Also my results were obtained for peripheral vision. If they had been obtained for foveal vision then again the value of

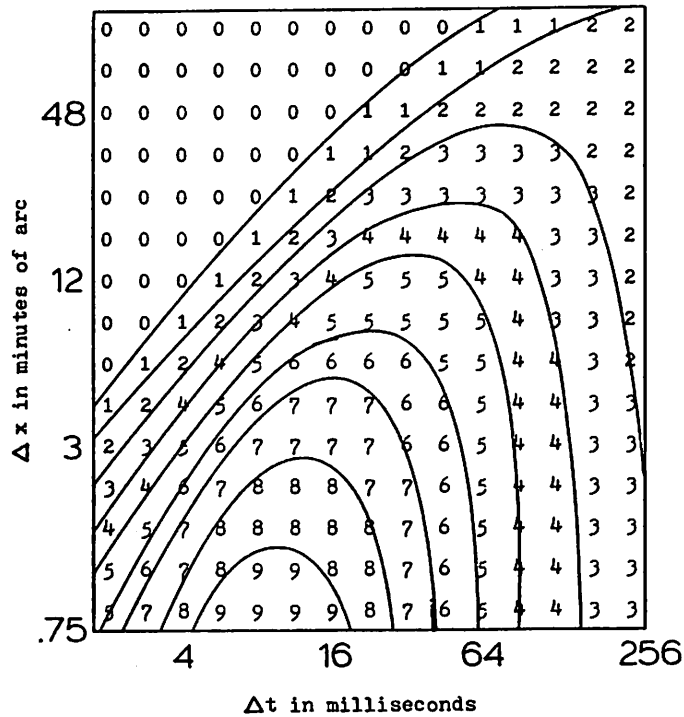


Figure 4.11

Values of  $\bar{E}^2$  (equation 6) which correspond to Sperling's results of figure 4.10.

B would be decreased and the curves would be displaced to the right. Either of these two factors could account for the difference in position of regions of best motion with respect to the time axis in the two figures. The other respect in which the two figures differ most clearly is in the shape of the region in which quality of motion was judged to equal 4. However, this difference should not be considered too significant on the basis of the present data due to the subjective nature of measurements in Sperling's experiment.

Phenomenon 6: Depth and motion in random process stereograms. Ross has recently devised a generalization of the Julesz stereogram in which the random pattern of dots is continually changed, while the target-surround relationship remains unchanged (Ross, 1974, 1976). To present one of these "random process stereograms," a computer is used to continually generate a stream of pseudo-random points. The points are displayed on two cathode ray tubes which are arranged so that only one is visible to each eye. As in Julesz's stereogram, the patterns presented to the two eyes are identical, except that points falling within a pre-defined target region of the display are displaced slightly in one eye relative to their position in the other. It is this displacement which causes disparity and hence the illusion of depth when the two images are stereoscopically fused.

In the random process stereogram, each point is displayed only once to each eye and very briefly (for about 10 sec), but

since all points displayed over a 100 msec. period appear to be simultaneous, one perceives a dense and ever changing random pattern.

This method of dynamic stereogram generation makes it possible to introduce delays between the time at which a dot is presented to one eye, and the time its displaced partner is presented to the other. Ross has found that fusion and depth may be perceived, even with the relatively large delays of several tenths of a second (or more). However, the introduction of delays in excess of .70 msec. brings about a perceptual change which is of particular interest to us here, that is, subjects observe apparent motion between the left and right members of each point pair, so that the field of view seems to be streaming either to the right or to the left. Furthermore, Ross has found that without disparities between left and right image points, depth can still be seen - on the basis of differences in the time delays applied to target and surround points.

Since different time delays in the zero disparity display result in different rates of apparent motion, as well as depth, it would seem possible that a motion parallax depth mechanism may be involved. But if this were the case, the sense of depth should not depend on the fact that the display was binocular. To test the motion parallax hypothesis, Ross (personal communication) modified the display to be monocular.

To use the present notation, each random point was presented twice to the same eye, the second time shifted a fixed distance  $S$  to the right of the first presentation, and following the first by an interstimulus interval  $ISI$ , which would result in apparent motion between the pair. The displacement  $S$  was the same for all points, but the  $ISI$  was different for target and surround points. This display produced apparent motion or streaming, as in the binocular presentation, but it did not produce depth. Thus, motion parallax cannot be the mechanism of depth perception here. Ross therefore suggests that depth is computed on the basis of time delays by a previously unknown mechanism. He notes also that interocular time delays which are correlated with depth occur whenever one is rotating his eyes, so a time delay depth mechanism which makes use of this information could exist.

The model described here will offer an alternative explanation not involving binocular time delay elements.

Related Phenomena. The above six phenomena are the principal phenomena which motivate the model and which will be discussed in relation to the model in the next section. There are, however, several related phenomena which should be mentioned briefly.

Shepard and Judd (1975) have measured the quality of apparent motion between two patterns which are line drawings of fairly complex three-dimensional objects seen in different orientations. For appropriately chosen  $ISI$ , observers

perceived a ridged object rotating in depth as line elements of the first pattern moved in unison to the corresponding elements of the second pattern. However for other ISI, a qualitatively different type of apparent motion was perceived in which line elements of first pattern moved independently to the nearest line elements of the second pattern. It seems likely that very different apparent motion mechanisms are responsible for these two types of perception. The ridged motion requires sophisticated global pattern matching and occurs for ISI in excess of 300 msec. The nonridged motion occurs for ISI less than 200 msec, and seems to involve only local pattern matching. This nonridged motion probably involves the same mechanism as the other apparent motion phenomena considered here.

Another phenomenon which should be mentioned involves a variation of the ambiguous display in which patterns are presented alternately to the left and right eyes. To avoid stereoptic fusion in this case, the patterns have to be vertical rows, see Figure 4:12. Also a fixation point is provided to each eye to facilitate pattern alinement. The displacement  $S$  is chosen to be  $2/3d$ , as shown in the figure. This means that the displacement between the successive patterns presented to each eye individually is  $2S$  or  $4/3d$  and this is equivalent to  $1/3d$ . Thus, when the observer closes either eye, the row of dots appears to move upwards, but when

he opens both eyes, the apparent motion reverses abruptly in direction. We conclude from this that monocular motion sensitive elements cannot be critical in determining apparent motion, at least with this display. Fairly large values of  $d$  (about one degree or larger) were required here to avoid binocular rivalry.

A final observation may be made with respect to the display shown in Figure 4:13a. In this display  $S_H = \frac{1}{2}D$  so that three types of motion are possible as shown. Motions right and left are of equal velocity and symmetrically oriented about the vertical. Suppose that ISI,  $D$  and  $S_V$  are chosen so that motion may be seen to the right or left but not vertically. When ISI is not too large, say 35 msec or less, one will see not one row but several moving together as in Figure 4:13b. Multiple rows may be seen here just as multiple dots can be seen in Ross's display of phenomenon 3. The rows form a pattern which moves down and to the right. As might be expected, the top rows of the pattern do not seem as bright as the bottom rows. However, I have noticed another curious property of the pattern which might not be anticipated. For the purpose of illustration, suppose that motion is seen to the right. The row spacing does not seem to be uniform; while the vertical distance between the lower rows appears to be less than  $S_V$ , the distance between the upper rows seems to be greater than  $S_V$ . In addition, each row seems to be shifted slightly to the right relative to the row above it.

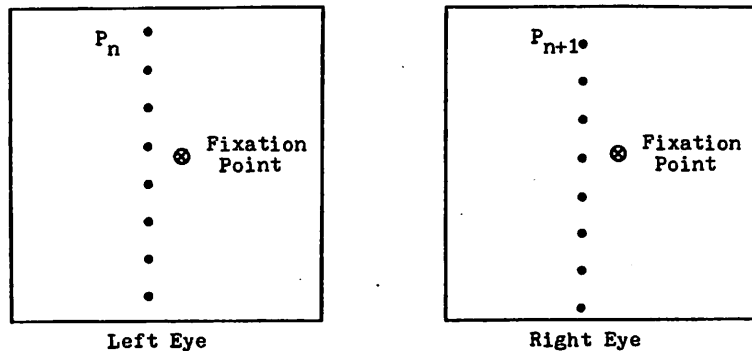


Figure 4:12

A binocularly segregated display which shows that monocular motion sensitive elements cannot play a critical role in apparent motion with spatially periodic displays.

These spatial distortions are illustrated in an exaggerated way in Figure 2:13b. My feeling is that these effects are related to the pinched corners of the Land Square, phenomenon 4, and reflect a tendency for dots when first presented to seem displaced in the direction of perceived motion.

Summary of apparent motion phenomena The following is a list of the important points in the above examples, which we would like to explain with any apparent motion model.

1. (Phenomenon 1) How do we account for the four qualitatively distinct perceptions obtained with the basic display (Figure 4:2): simultaneity, phi motion, beta motion and succession?
2. (Phenomenon 1) How do we account for the parameter constraints for beta motion expressed in Korte's Laws?
3. (Phenomenon 1) As an alternative to #2, can we account for the Neuhaus result, that for a given stimulus intensity, the velocity of beta motion is proportional to distance, that is ISI is constant?
4. (Phenomenon 2) Why should beta motion occur between two patterns but not over a sequence (sparse) of patterns?
5. (Phenomenon 2) Is perceived motion with a dense sequence due to a different visual mechanism than beta motion between two stimuli?
6. (Phenomenon 3) What visual system processes are involved in "retrospective memory readout"?
7. (Phenomenon 4) How do we account for motion hysteresis



of the sort revealed by the Land Square?

8. (Phenomenon 5) How do we account for the selection of a single direction of motion in ambiguous spatially periodic displays?

9. (Phenomenon 6) Can depth perception based on time delays be accounted for in an apparent motion model?

#### 4.2. The Apparent Motion Model

The fundamental assumptions of the model which will be described here were given in the introduction to this chapter: it is postulated that 1) perception of an object is accompanied by the development of a corresponding pattern of neural activity within the visual system, 2) the object is perceived continuously only so long as this pattern of activity is maintained and 3) for integration of isolated stimuli, as in beta motion, the pattern of activity must move within the supporting neural structure from the location stimulated by one pattern presentation to that stimulated by the next. These assumptions characterize the idea of the stimulus organization model discussed in the first chapter where moving patterns of neural activity were moving segment representations. Apparent motion will be explained in terms of processes which initiate activity and cause it to move within the neural structure. As in stereopsis, where activation of disparity detectors is not adequate to account for perception,

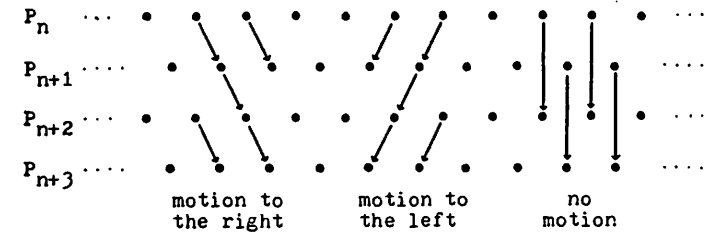


Figure 4.13a Successive patterns when  $S=d/2$ , and three ways in which dots in successive patterns might correspond.

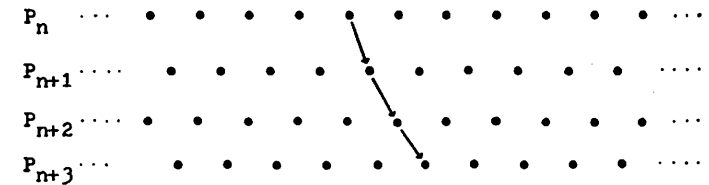


Figure 4.13b When ISI is adjusted so that several rows of dots appear to be simultaneously on, and moving together down and to the right, then the spacing of the rows may seem distorted as shown here.

here activation of motion detectors, while important, cannot account for perception.

The apparent motion model incorporates features of several other models. Following Burt (1975), the system for representing segments and segment motion consists of three neural layers, as in Figure 4.14. It is supposed that there are two types of stimulus input: image features and image motion. These are shown as arising from detectors in the retina, but it is assumed that motion detection occurs at a later stage of processing, probably in visual cortex, and in any event, previous to stimulus integration of the sort considered here. Feature information leads to formation of segment representations in the OL (object location) layer. Activity in the OV (object velocity) layer then causes the pattern of activity in OL to move. It is imagined that OV activity corresponds to perceived object velocity. The original model incorporates eye rotations as well. Once these activities have become established, segments will move in anticipation of image motion. Any differences in image and segment motion are detected in the DV (difference in velocity) layer and the difference is added to OV activity. Large differences may lead to reorganization.

It is supposed that mechanisms of stimulus organization responsible for combining the many motions stimuli for a single segment (see discussion of Figure 1:16 in Chapter 1)

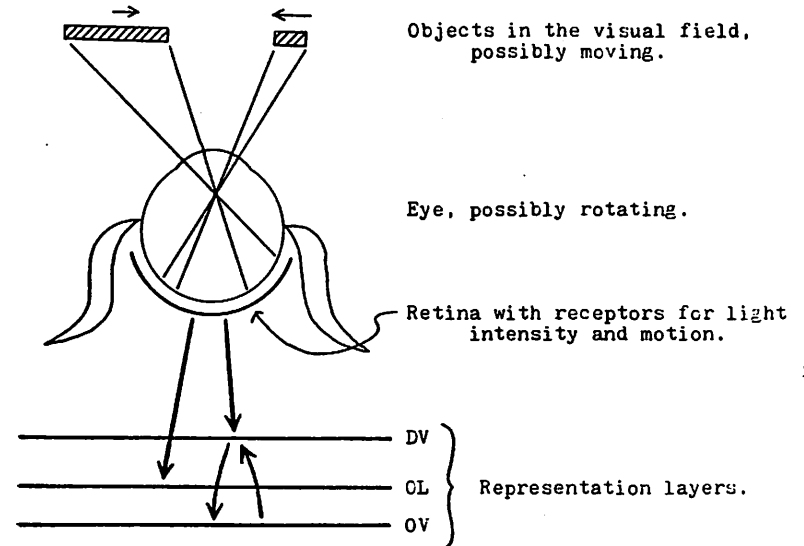


Figure 4:14

A model for the representation of motion in the visual system. Each arrow indicates a topographic projection.

are incorporated in layers OV and DV. But that type of integration will not be considered here. The interactions of activities in these layers which lead to motion are defined in Burt (1975).

The OL layer is assumed here to have the same response characteristics as the cortex model of Wilson and Cowan, (Wilson, Cowan, 1973). Thus a brief stimulus to a small region of the net and of sufficient energy will initiate a characteristic response cycle in which net activity grows rapidly to a high level, then dies more slowly to its initial level. This response would be confined to the local region of the stimulus, as in the Wilson-Cowan model, if it were not for possible influence of OV layer activity which makes the activity move.

There are many details of the Burt and Wilson-Cowan models which are assumed here but which will not be repeated. For example, rather than define details of neural interconnections as in these three models, I simply assume such interconnections, and look only at the resultant net properties. In particular, I will examine how the elements of the models, i.e. motion and transient net responses, can be fit together into a model for apparent motion. Properties of the model will now be developed in a qualitative way by considering in turn its response to five very elementary stimulus configurations. In the first case, I consider the net response to

a single brief stimulus. This is essentially a review of Wilson-Cowan's description of the transient response.

Case #1: Transient response to a single brief stimulus.

The OL layer is made up of two populations of cells - one excitatory and one inhibitory. The level of Ex, excitatory, activity following the presentation of three different stimuli,  $R_1$ ,  $R_2$  and  $R_3$  is shown by solid lines in Figure 4.15. The dashed line shows the level of inhibition, In, activity over time for  $R_2$ . This figure, which is based on Wilson and Cowan's results, shows two qualitatively quite different ways in which activity in the net may evolve following a brief stimulation. The energy of the stimulus determines which of these transient response patterns will result. Thus, stimulus  $R_1$ , which was presented from  $t_0$  to  $t_1$  did not reach the critical energy, so Ex activity did not continue to increase after  $t_1$ . However, the stimulus  $R_2$ , which had the same intensity as  $R_1$  but which was presented for a slightly longer time,  $t_0$  to  $t_2$ , did pass the critical energy level so that the active net response was triggered. In this case, Ex activity continued to increase after  $t_2$  reached a high maximum value, then slowly decayed. The decay of Ex activity is caused by the increased inhibitory activity, In, which develops more slowly than the Ex activity.  $R_3$  is a more intense stimulus than  $R_1$  or  $R_2$  and net response to  $R_3$  shows the effect of this increase in stimulus energy. The maximum level of Ex activity reached is not much changed, but the time taken to reach the

maximum is considerably shortened and decay of Ex activity begins sooner due to more rapid development of In activity. Thus a brief stimulus which exceeds the energy threshold will evoke a net response which follows a characteristic pattern of development and decay. Excess of stimulus energy over the threshold will increase the rate at which this response pattern evolves.

The spatial distribution of net activity at several moments over the time of the transient response is shown in Figure 4:16. Again, Ex activity is shown by a solid line and In by a dashed line. Following the stimulus, Figure a, Ex develops progressively to a maximum at  $t_3$ . The development of In activity follows Ex activity, and eventually suppresses it,  $t_4$ . In activity extends over a wider area, so has the effect of narrowing the peak of Ex activity as it develops. Thus, excitation activity develops an inhibitory surround.

These temporal and spatial response characteristics assumed for the OL layer will play a critical role in the model for apparent motion. As mentioned above, it will be assumed that OV layer activity can cause the response patterns described here to move as they develop. Also, a threshold,  $T$ , (as in Fig. 4:15) is assumed to determine the perceptual status of activity in the OL layer. That is, higher level regions of the visual system ( $I_2$  of Figure 1:10) will "see" activity in an OL layer only if the Ex component of that activity exceeds  $T$ . Only then will OL activity

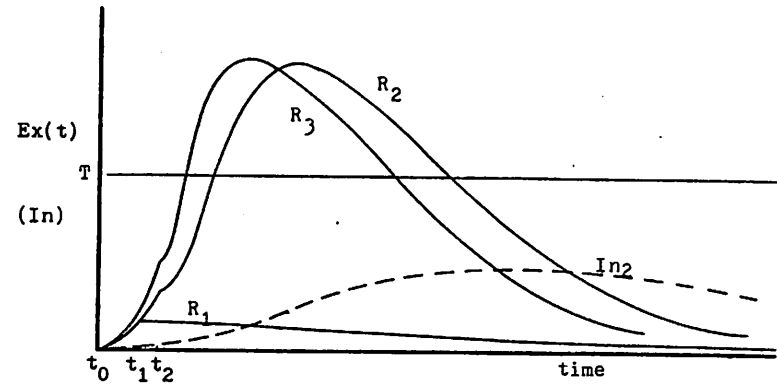


Figure 4:15 Ex activity is shown to follow one of two characteristic patterns of development following a brief stimulus.

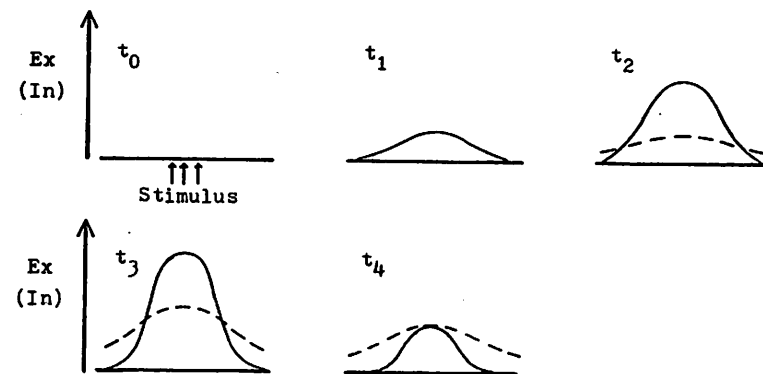


Figure 4:16 The spatial distribution of Ex (solid line) and In (dashed line) activity at five moments in time following a brief stimulus.

correspond to perceived objects.

In the next four cases, I examine the net response to pairs of stimuli.

Case #2: Temporal resolution. Suppose an observer is presented with two brief flashes to the same point in his visual field, but one following the other by a short time  $\tau$ . Psychophysical experiments show that there is a critical time  $\tau_c$  such that if  $\tau < \tau_c$ , the two stimuli will lead to the perception of only one dot (stimulus integration) while if  $\tau > \tau_c$  the distinct stimuli are perceived, one following the other (segmentation).

Now consider the response of the OL layer in the model to the two stimuli. The first stimulus initiates a characteristic response cycle. The effect of the second stimulus on Ex activity will depend on when this stimulus occurs within the response cycle of the first. In particular, there are five qualitatively different response to the second stimulus. These are associated with the five time zones indicated in Figure 4:17a, which shows the response pattern for the first stimulus. If the second stimulus falls into one of these zones, the response is as follows:

**Zone I:** A second stimulus in Zone I has an effect essentially equivalent to increasing the energy of a single brief stimulus. As shown in Figure b, the result is accelerated evolution of the response pattern. Only one dot is perceived, since activity is continually above T.

**Zone II:** The second stimulus occurs after Ex activity reaches a high level and before it drops back below the threshold T. In this case, the time during which Ex activity is above T is extended, as in Figure c. The perception is again of a single dot.

**Zone III:** The second stimulus occurs after Ex has fallen below T, but while Ex is still sufficiently high so that the second stimulus can raise activity above T for a second time, Fig. d. In this case, two dots are perceived in succession. Note that the second stimulus does not raise Ex to the same maximum level as the first, since the second occurs when inhibitory activity is well developed.

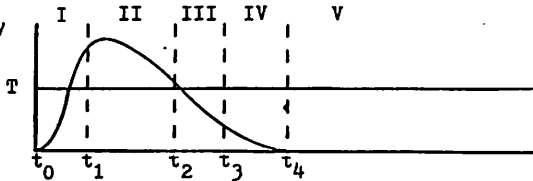
**Zone IV:** If the second stimulus falls in this region, it will fail to drive Ex above T at all, Fig. e. In this case, the second stimulus is not perceived.

**Zone V:** The second stimulus falls in the fifth time zone, that is, after net activity associated with the first has nearly died out. The second stimulus initiates its own response pattern, as in Fig. f. In this case, two distinct dots are perceived as Ex activity exceeds T on two occasions.

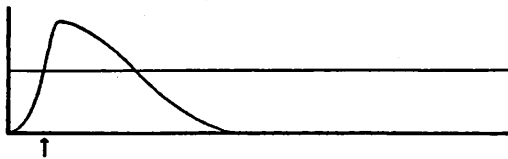
If we equate the experimentally derived  $\tau_c$  with  $t_2$  in the model, the time separating Zones II and III of Figure 4:17a, then the model behavior corresponds to experimental observation.

We might note that these time zones depend on the stimulus intensity. If more intense stimuli are used, the time

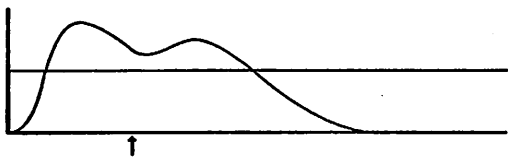
a) The Ex activity following a single stimulus, and five time zones.



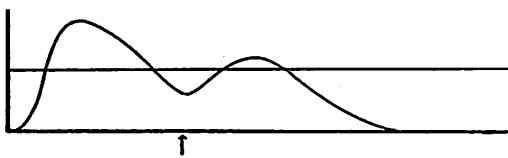
b) Ex activity when a second stimulus occurs in zone I.



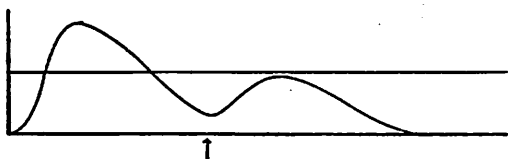
c) A second stimulus in zone II extends the time Ex is above threshold T.



d) Second stimulus in zone III. Ex exceeds T two times.



e) Second stimulus in zone IV. Ex activity following the stimulus fails to exceed T.



f) Second stimulus in zone V. Response to second is like response to the first.

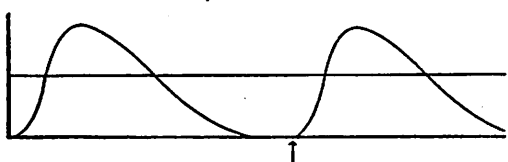


Figure 4.17 Ex activity is shown due to two stimuli, and for five different delays between stimuli.

course of the response to the first stimulus is quicker and the second stimulus can push Ex activity above the threshold from lower initial Ex values. That is, time zone III is extended, and for sufficiently intense stimuli, zone IV will not exist.

Case #3: Flicker fusion frequency. Suppose that instead of presenting two stimuli to the same position, as in case #2, a series of identical stimuli are presented to that position, one every  $\tau$  seconds. Again, there will be a critical value of  $\tau$ , call it  $\tau_c$ , such that if  $\tau < \tau_c$  an observer will see what appears to be a single constant stimulus, but if  $\tau > \tau_c$ , "flicker" is perceived and discrete stimuli are observed. This is just the reciprocal of the critical flicker-fusion frequency. The interesting psychophysical point here is that  $\tau_c > \tau_f$ . (In fact,  $\tau_c = \frac{1}{2} \tau_f$  (Boynton, 1970)). Thus one's ability to discriminate stimuli in the time dimension improves with repeated stimulus presentation.

The model predicts this phenomenon in the following way. Notice in Figure 4.17c that for  $\tau$  just less than  $\tau_2$ , the level of Ex activity in response to the second stimulus does not reach nearly so high a value as it did in response to the first stimulus. The reason for this is, of course, that the inhibitory activity, In, is well developed at the time of the second stimulus presentation. As a consequence, the Ex activity falls back to the threshold T more rapidly after the second stimulus than after the first. Thus, if a third

stimulus is to be integrated with the first two, it must follow the second by a time interval considerably less than  $t_2$ , see Fig. 4:18.

We see here an example of how a response activity in the visual system modifies subsequent processing by the system in a way that increases its power to discriminate stimulus patterns. In this example, there is an increase in time resolution. In the example below, the increase is in spatial resolution.

Case #4: Spatial resolution. There is a parallel between the perceptual situation with two or more stimuli presented at the same time but spatially displaced from one another and the above cases. Again, there will be a critical displacement  $d_c$  such that two dot stimuli which are displaced less than  $d_c$  will be integrated and perceived as one dot. Thus  $d_c$  is the limit of spatial resolution of two dots. Corresponding response patterns of the model are illustrated in Figure 4:19. But perceptual segregation of the dots results if their separation is greater than  $d_c$ .

Since the inhibitory activity in the OL layer covers a wider area than excitatory activity, two other predictions may be made on the basis of the model: 1. Spatial resolving power will improve with repeated stimulation, and 2. the spatial resolution for a row of dots is greater than for two dots. Both results follow from the increased background inhibition which effectively increases threshold of perception.

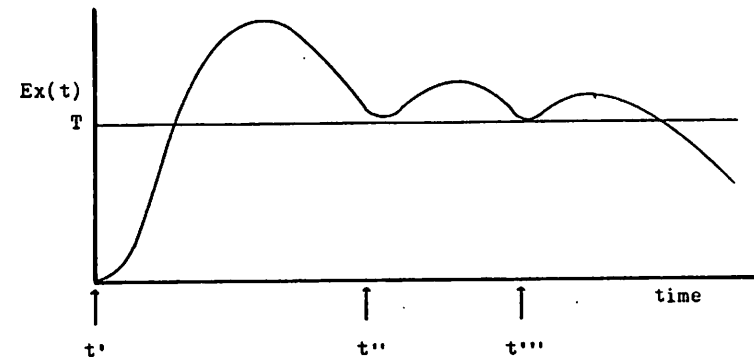


Figure 4:18

Ex is shown for three stimuli presented at times  $t'$ ,  $t''$ , and  $t'''$  respectively. With  $t'' - t' < \tau_c$  the first two stimuli will be perceptually integrated since Ex will exceed T at  $t''$ . However for the third stimulus to be integrated as well the interval  $t''' - t''$  must be much shorter than  $\tau_c$ . Since inhibitory activity is well developed at  $t''$  Ex will tend to fall below T more rapidly after the second stimulus.

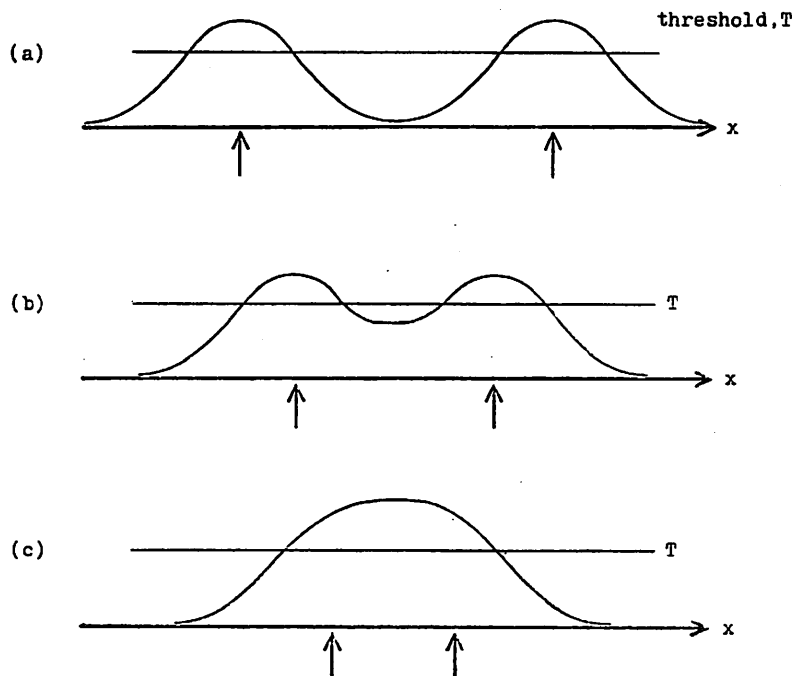


Figure 4:19

Spatial distribution of  $E_x$  activity at a time when it is maximum following simultaneous stimulation in two positions as shown. Cases a and b result in segregation while c results in integration.

Case #5: Apparent motion and Korte's Laws. Now suppose that two stimuli are presented with both a relative time delay and a spatial disparity. We want to know if the conditions under which the response pattern for the two stimuli integrate in the model corresponds to those observed for apparent motion.

As indicated previously, the underlying assumption about apparent motion in this model is that it corresponds to motion of the response pattern initiated by the first stimulus, so that that pattern's position at the time of the second stimulus is, roughly, the position of the second stimulus. Figure 4:20 illustrates the scenario. Activity is shown at various stages of development between the presentations of the two stimuli. We assume for the moment that a proper level of activity already exists in the V layer to cause this motion. Figure 4:20 shows, not only the continual motion of the activity pattern, but its growth and decline in size. The second stimulus is perceptually integrated with the first just in case: 1) it coincides with the position of the moving activity pattern at time  $t_4$ , and 2) the activity at  $t_4$  has not dropped below the threshold  $T$ . There are two issues which must be clarified. First, what are the temporal and spatial constraints on integration, i.e. what is the analogue of Korte's Laws in the model? Second, how is it that the velocity (represented as activity in the V layer) comes to be already present when the first stimulus is presented?



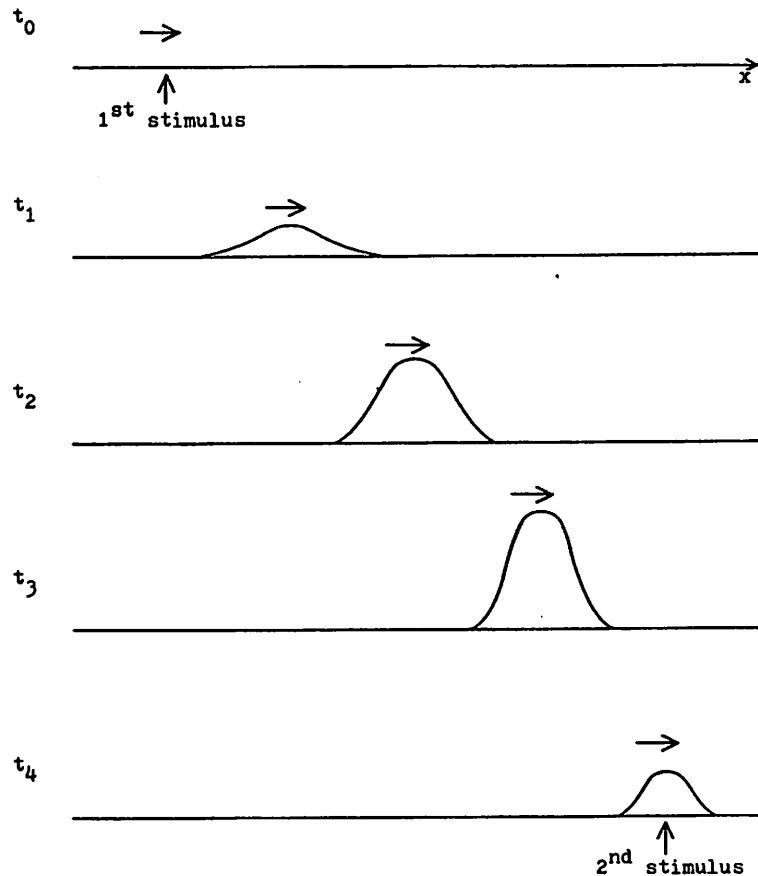


Figure 4.20 Spatial distribution of Ex activity is shown at several moments following one stimulation and up to the time of a second. The response pattern moves as it evolves, and this is taken as the mechanism of beta motion. Note that this figure is the same as 4.16 with motion added.

We begin with the question of constraints on integration. Assume now that the response pattern of the first stimulus moves at an appropriate velocity, so that success or failure of stimulus integration depends only on the stage of development of this response activity at the time the second stimulus occurs. In particular, integration occurs when the response development is in zone I or II, as defined in Figure 4.17. If we assume that the response develops at the same rate whether it is moving or stationary, then integration will depend only on the total time between the two stimuli, and this time should not exceed  $\tau_c$  defined in case #2. We note further that  $\tau_c$  is a decreasing function of the energy of the first stimulus. This energy corresponds to I defined in the experimental section of this paper. Therefore, the assumption that the rates of pattern evolution are unaffected by motion leads to a conclusion much like that drawn from the Neuhaus data (see Phenomenon 1 above), i.e., that the interstimulus interval required for beta motion depends inversely on I and does not depend on S, the separation.

On the other hand, it will not necessarily be the case that the processes causing the evolution of a response pattern and processes causing its motion will be independent. Without defining these processes specifically, it is not possible to state precisely how they interact. My experience in simulating a related system (Burt, 1975) is that a moving activity pattern decays more rapidly than a stationary pattern, and the

magnitude of this effect increases with velocity. A possible result of a velocity dependent decay rate is that optimal conditions for apparent motion between two stimuli will be in accord with Korte's Laws. While the actual behavior of a model which combines motion with a Wilson-Cowan structure has not been determined, it can be shown how this or a similar model might exhibit the properties described by Korte's Laws. Suppose the strength of a motion stimulus is determined by  $Ex(S, ISI)$ , the level of OL layer activity which is initiated by the first pattern of a stimulus pair at the time the activity has moved a distance  $S$  to the position of the second pattern,  $ISI$  sec. later. Then  $Ex$  may be the neural equivalent of  $\hat{E}$  in equation 6. It is worth observing that the transient response of the Wilson-Cowan network (see Fig. 4:15) may be approximated quite well by  $\hat{E}$  when  $x=0$ . Furthermore it is not too unreasonable to speculate that the effect of adding velocity will be to increase the rate of decay by an amount which is exponentially related to velocity. If this were the case, then  $Ex$  would indeed be equivalent to  $\hat{E}$ . As has been noted, this functional form of  $\hat{E}$  is consistent with Korte's Laws (for  $ISI$  and  $S$  large compared to  $\delta$  and  $b$ ).

I now turn to the second question cited above: how is it that the velocity appropriate to apparent motion comes to be represented by activity in the OV layer. There are two mechanisms which contribute to motion representation. First, motion detectors which respond to motion stimuli in the input

pattern have been postulated (see Figure 4:14). This stimulus motion is compared to represented object velocity in OV and the difference is used to update OV activity. (Note again that when stimulus and OV motion match, they are mutually consistent, but not necessarily alike: OV represents motion information integrated over a segment, while stimulus motion is local). OV activity, in turn, causes motion of activity patterns in OL. When an apparent motion stimulus is first presented, OV activity will be zero. Suppose there is a slight delay, 50 to 100 msec., between the time an input stimulus becomes available to the motion detectors and the time it initiates an active response in the OL layer. This delay allows some time for motion detectors to respond to two input stimuli and to initiate OV layer activity. If OV activity is not adequately established in this way, motion will be sensed due to partial OV activation, but input stimuli will not be integrated. This condition we assume corresponds to phi motion. The second source of appropriate OV activity is repetition. Kolers suggests that beta motion is only perceived after several repetitions of the stimulus sequence so that the motion is anticipated. In the model stimulus repetitions allow OV activity to become established.

#### 4.3. Apparent Motion Phenomena Explained

The major features of the model have now been described

and illustrated. In this section I will consider each of the phenomena listed in Section 1 and show how they may be explained in the context of the model. The relation of model behavior to phenomena 1 and 5, i.e. Korte's Laws, has already been discussed.

Phenomenon 2: Here we want to explain why beta motion can occur between two widely separated stimuli, but only between more than two stimuli if the spacing is very narrow. My explanation will be based on the notion that long distance apparent motion requires that the velocity be near an optimal value, as in Korte's Law. Recall that this velocity depends on the stimulus energy. Now if we suppose a response pattern travels between the first two stimuli of a sparse sequence at its optimal velocity, based on the energy of the first stimulus, then it arrives at the second stimulus with the response activity evolved to zone 1 or 2, (see Figure 4:18). As the second stimulus is integrated with the first it in effect adds energy, with the result that subsequent evolution of the response pattern is accelerated. The velocity appropriate for beta motion between the first two patterns is too slow between the second and third. Hence, a failure of further integrations, and of beta motion with sparse stimuli sequences.

On the other hand, with dense sequences, the velocity of motion is not such a critical factor. It is not necessary to slow down the response evolution so that it lasts from one stimulus presentation to the next by judicious control of the

velocity. In this case, velocities will exist which do permit beta motion between the first two and subsequent points.

To summarize, the explanation here makes use of the idea of a critical velocity for motion between widely disparate points, and the notion that this velocity must be larger for more energetic stimuli.

Phenomenon 3: The principal aspect of Ross's display which needs explanation is "retrospective readout." Two explanations may be proposed in the context of the present model. First it has been suggested that there is a delay of 50 to 100 msec. between the time a stimulus activates motion detectors in the DV layer (Fig. 4:14) and the time it triggers an active response in OL. This means that motion information is processed slightly before feature information. As one views the dense stimulus sequence of Ross's display, the termination of the motion stimulus is sensed 50 to 100 msec. prior to the time the last dot is processed in the OL layer. Even a small change in motion representation could then cause failure of integrations of final dots of the row. In this way "retroactive readout" may not be retroactive at all but a consequence of a slight difference in times at which information of different types is processed.

The second explanation does involve a retroactive memory readout of the sort Ross suggests. This requires a further refinement to the model. I have postulated that visual stimulus is mapped to the OL layer by way of neurons

in the layer  $D_1$  of Figure 1:10. We will now assume that at least some of these cells, once active, remain active for some time after the stimulus is removed. In particular, it is supposed that a significant number of these cells will fire for roughly 100 msec. following stimulation. Evidence of maintained activity even at the level of the retina has been obtained by Sakitt (1975).

Notice that a moving point stimulus on the retina (real motion) will result in smearing of the stimulus to the OL layer, since at a given moment there will be stimuli to all points in the OL layer which correspond to loci on the retina which have been optically stimulated over the past 100 msec. In order that moving activity patterns in the R layer should not be significantly altered or destroyed by this stimulus smearing, we must postulate the existence of a region of inhibition following any moving activity pattern which counteracts the residual stimulus. See Fig. 4:21. Thus we suppose that there exists a relatively insensitive region in the trail of any moving object representation, and that the width of this inhibited area is  $d = V/10$ . There must also be a facilitated region in front of the moving representation so that a representation pattern is not suppressed by another if the two are within a distance  $d$  of each other and moving in the same direction. The inhibition and facilitation must originate in the processes causing motion of the OL layer activity.

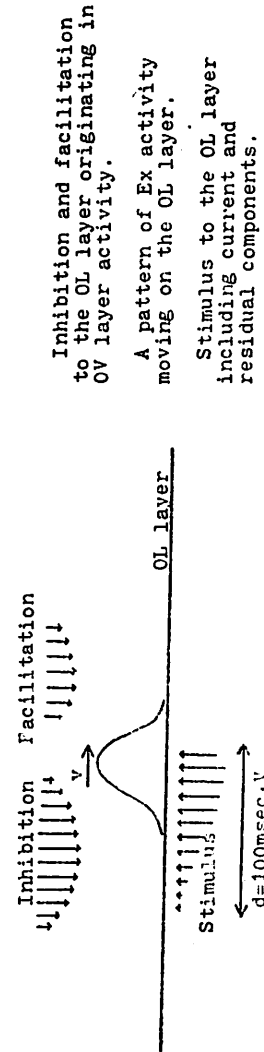


Figure 4:21

Now consider the apparent motion display used by Ross. With the proper display rate, the dense sequence of dots will yield beta motion, and one will perceive a single moving dot. A region of inhibition follows the moving dot representation, which suppresses activity in the OL layer over the area still receiving residual stimulation. When the end of the row is reached and the moving dot representation stops, we imagine that several things will happen. Activity in the OV layer will go to zero. Since this activity is responsible for the inhibition to the OL layer, this inhibition will also go to zero, which in turn will allow residual activity in  $D_1$  neurons, associated with the stimuli over the past 100 msec. to reactivate response patterns in the OL layer. This time there is a stationary response pattern for each stimulus point. Perceptually this amounts to retrospective readout.

To summarize, we argue that associated with moving representations there must be an inhibitory following zone (and a facilitating preceding zone) to compensate for stimulus smearing. The retrospective readout of Ross corresponds to a collapse of this inhibition so the OL layer responds anew to the residual stimulus. The explanation we give here for the phenomenon is essentially that given by Ross, except that here retroactive readout is specifically associated with motion perception.

Phenomenon 4: The Land Square is similar to the case above - each side of the square is a dense row of dots. The

appearance of beta motion over the central portion of the sides but of isolated dots at the ends can be explained in the same way as above. What we need to explain now is the "pinched" appearance of the corners.

As I suggested in the first section, pinching may be explained in the model as an hysteresis effect associated with the moving dot representations. We imagine that a dot representation actually moves slightly beyond the end of an edge before the absence of continued stimulation leads to the complete collapse of the moving representation. Before this collapse occurs, the initial stimulus points of the next edge are being plotted, Fig. 4:22b. At the moment when the moving representation collapses, Fig. 4:22c, there may be apparent motion, presumably of the phi type, between that dying representation and the current stimulus point on the next side, as shown. At the same time, there is a retroactive readout phenomenon making the individual stimulus points perceptible.

Phenomenon 6: The similarity of motion perception to stereopsis which was discussed in Chapter 1, Section 5 leads one to suspect that the processes responsible for these perceptions might be the same or tightly coupled. An interesting interpretation of the combined motion and depth due to time delay observed by Ross is that it is evidence of this coupling. Both types of perception are based on small stimulus disparities and are obtained even though there is no actual disparity in the display used by Ross. However it could be that

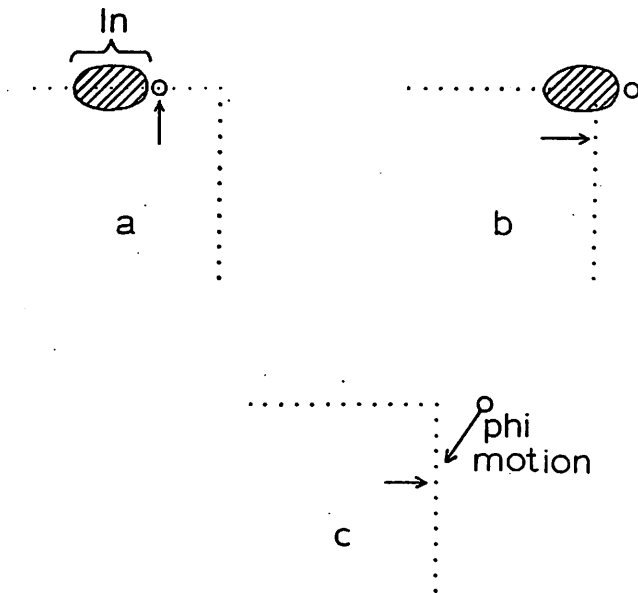


Figure 4:22

Stimulus and perception part ways at the corner of the Land Square. An arrow indicates the stimulus dot which is being presented at each of the moments illustrated. A circle shows the perceived dot, which moves smoothly beyond the end of the edge in b, before it dies in c. The crosshatch area is the zone of inhibition following the perceived dot. In b inhibition suppresses initial stimulus points on the next edge. In c inhibition collapses and there is "retroactive readout" of stimulus points in the corner.

the stereopsis mechanism in effect assumes that there is a disparity and the motion mechanism assumes an opposite disparity so that the two cancel and the motion-depth perception is consistent with a display without disparity. A difficulty with this type of explanation for phenomenon 6 is that the delays which were found to result in depth and motion were 70 to 200 msec. or more. Ross and Hogben (1973) have found that short term memory for stereopsis only lasts about 50 msec. Furthermore evidence has been cited here which indicates that stereopsis and motion perception are processed at different levels of the visual system. In particular it was concluded in Chapter 3 that stereopsis depends upon interactions between separate populations of cells which code the images from the two eyes independently. On the other hand, evidence given in this chapter (see discussion of Figure 4:12) indicates that apparent motion follows binocular combination. An alternative explanation of the time delay dependent phenomena will be given here in which stereopsis is computed prior to motion and in a separate neural structure. These structures are the same as the models I have proposed for stereopsis and motion perception and are arranged as shown in Figure 4:23.

In Ross's random process stereogram with time delays points presented in the target area are presented to both eyes simultaneously and it is reasonable to assume these points are binocularly fused. However each point presented in the surrounding area is plotted to one eye 70 msec. or

more before its partner is plotted to the other eye. Even though there is no (relative) disparity in this case, it is reasonable to suspect that surround points do not fuse due to excessive time delay so are processed as if they were monocular stimuli. Under these assumptions Ross's stereogram is essentially equivalent to the well known Panum's limiting case. This is a stereogram composed of two vertical lines presented to one eye and a single line presented to the other as in Figure 4:24a. When this stereogram is binocularly viewed so that the single line of the right half image is fused with either line of the left half image and two lines appear in the "cyclopean" view, then one of these lines will be perceived in depth relative to the other. The same depth effect, whatever its cause, should be anticipated with Ross's display.

Panum's limiting case can be explained with the stereopsis model proposed here. (For empirical evidence see Kaufman, 1976). Suppose as one views the stereogram of Fig. 4:24a he converges his eyes so that lines B and C fuse. Generally speaking, vergence will not be so precise that the lines are fused with zero disparity. (This is not an unreasonable assumption. Ogle (1950) has found that when subjects fixate an object in space the actual point of fixation is usually slightly in front of or behind the object due to imprecise vergence). This means that the match cells which are activated by lines B and C will not be on the zero disparity

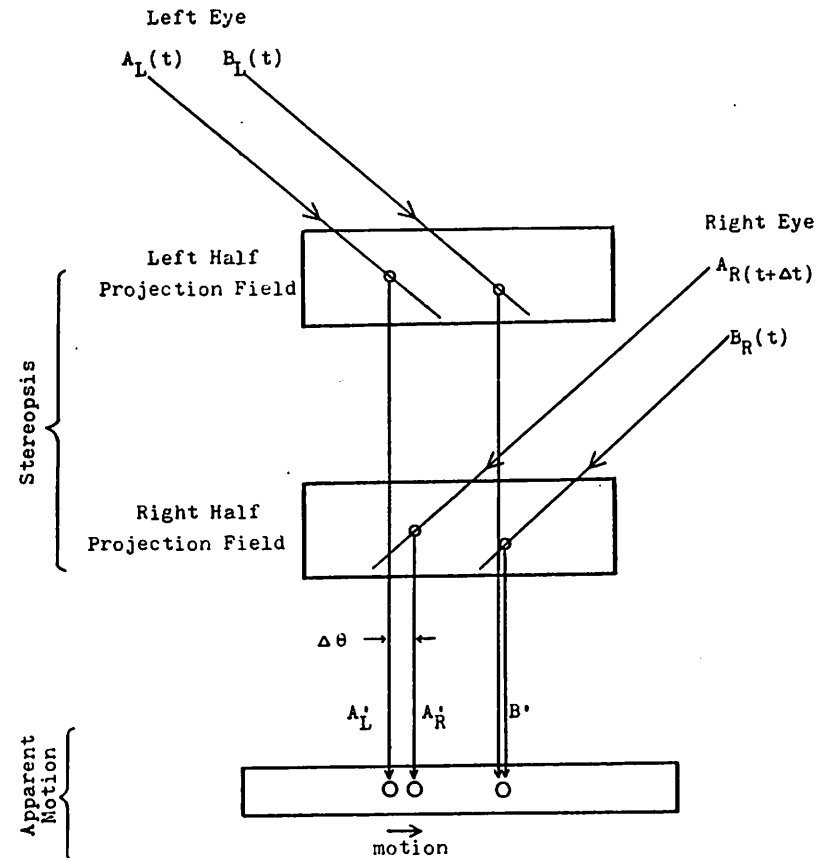


Figure 4:23

Ross's depth due to time delay may be due to a uniform disparity  $\Delta\theta$  and fusion of target area dots (B) but failure of fusion of surround area dots (A).

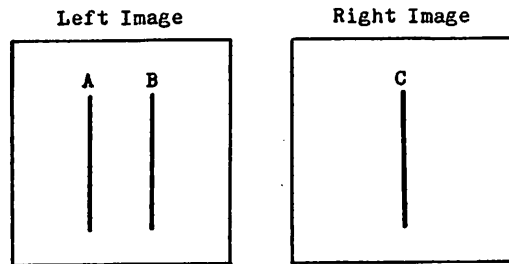


Figure 4:24a Panum's limiting case.

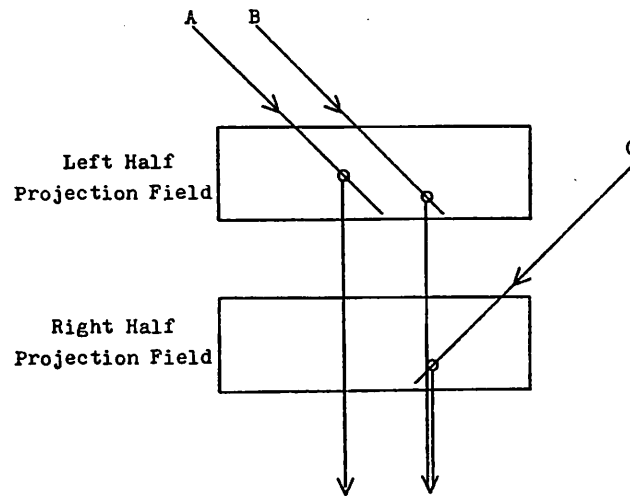


Figure 4:24b Pattern of match cells in the stereopsis structure which are activated by Panum's stereogram. If eye vergence does not allow fusion of B and C at zero disparity then depth will result.

plane of the half projection fields, as shown in Figure 4:24b. On the other hand, the cell activated by line A will be on the median line, since it is monocular. Therefore line A will appear at a different depth than the B-C line.

Now consider two dots presented binocularly in Ross's display, as shown in Figure 4:23. Dot A corresponds to a point in the surround region so that it is presented to the left eye at time  $t$  and to the right eye at  $t + \Delta t$ , where  $\Delta t$  is the Ross's time delay. Dot B is in the target area, so is presented to both eyes at the same time,  $t$ . Again  $B_L$  and  $B_R$  fuse with a nonzero disparity  $\Delta\theta$  due to a small error in eye vergence. Dots  $A_L$  and  $A_R$  do not fuse, so activate zero disparity match cells. This results in depth. Furthermore the disparity  $\Delta\theta$  becomes a spatial displacement between  $A'_L$  and  $A'_R$  which are the outputs of the stereopsis stage of processing. This  $\Delta\theta$  combined with the temporal disparity  $\Delta t$  becomes a stimulus for apparent motion in the next stage of processing.

#### Summary and Comments

In developing this model for apparent motion, I have made a number of assumptions, some of which are fundamental to the theory while others are simply useful for making the model concise. I will now restate what seem to me to be the four most fundamental assumptions.



The first assumption relates to the perceived continuity of an object's identity as its image moves on the retina. In the model, perceptual continuity is identified with the continued existence of a pattern of neural activity which represents the object, in an "integrating" structure of the visual system. When the object is perceived to move, this activity pattern moves within the structure. Beta motion occurs when object-related neural activity moves but object images do not.

The second assumption is that the "sensation" of motion occurs when certain motion sensitive elements are stimulated. Some of these elements may occur more peripherally in the visual system, but many must occur within the integrating structure mentioned above. Phi motion corresponds to the stimulations of motion sensitive elements without motion of the neural activity which represents objects.

The third underlying assumption of the model is that patterns of neural activity in the integration structure, which are evoked by brief stimulation, develop in a characteristic way over time. It is this development over time which places constraints on stimulus integration in apparent motion. The model behaves according to Korte's Laws when it is assumed that the rate at which activity patterns develop decreases with increased velocity, and increases with increased stimulus energy. The curious result that beta motion can occur between two but not a (sparse) sequence of patterns also

follows from these assumptions.

The final assumption is that there is a region of inhibition following any moving activity pattern in the integration structure, and another region of facilitation preceding it. This inhibition and facilitation originates in the processes which cause motion of the activity pattern. A number of results were shown to follow from this assumption, including "retrospective memory readout," phenomenon #3.

Further work on this theory might proceed in several directions. Further psychophysical experiments could be designed specifically to test the above assumption. The model should be examined also in the light of neurological data to determine where in the visual system motion detecting elements may occur, and where the lowest level of integration processes may occur.

## S U M M A R Y

It is often supposed that the first stage of visual information processing may be characterized as local feature extraction. This operation can be performed by relatively simple neurons, or groups of neurons, which function as perceptron-like detectors. However, it seems clear that such detectors, or even a hierarchy of detectors, cannot account for natural perception, and that more dynamic mechanisms must play a role.

Stimulus organization is proposed here as an early dynamic stage of image processing. Organizing processes perform two necessary functions when the visual field contains images of more than one object: first these processes segment the visual field into regions containing the features of individual objects (or parts of objects), and second, they control feature allocation in subsequent image processing, so that no single image feature is perceptually treated as part of more than one object. These segmentation and allocation operations must be performed at an early stage of visual processing because a high degree of spatial and temporal resolution is required in the coded visual information, and so that the computed image segments may direct higher level processes which are responsible for object recognition.

To a large extent stimulus organizing processes may be

stimulus driven. This is because segmentation can usually be based upon a few characteristics which nearly all object images have in common, such as the fact that features of a given object are localized in space and move together over time. (It is these characteristics which make studies of stereopsis and motion perception particularly relevant to a study of organizing processes.) While segmentation involves global information it may be still be computed on the basis of local features and local interactions in a retinotopically organized neural network. This possibility is demonstrated by the model proposed here in which the neural network was represented by an array of locally coupled processors. Each processor receives feature information from a corresponding region of the visual field. The processor then enters a state which is consistent with the input, and this state represents a perceptual quantity, such as a distance or velocity, which is associated with its input stimuli. In addition there are local consistency constraints which indicate when neighboring processors are in mutually consistent states. When these constraints are satisfied over the entire array, the array is said to be globally organized. It is these constraints which cause segments to form which reflect physical characteristics of objects. State change strategies are built into individual processors (or neural network properties) which cause the array to be self organizing, so that it will change from an initial

unorganized state to a globally organized state.

This array model for stimulus organization was illustrated with a simple computer simulation at the end of chapter 1. Arrays of formal neurons were defined which perform figure-ground separation of a line drawn image. The state change strategy of the model was based upon neural fatigue, while local consistency constraints took the form of lateral inhibition. It is interesting that these network properties not only resulted in net self organization in ways appropriate for representing figure-ground separation, but resulted also in spontaneous figure-ground reversals.

Stimulus organization based on stereopsis and motion cues depends upon the resolution of a local stimulus matching ambiguity. In particular, organizing processes are needed to determine how individual features of one image should be matched with features of another, when these images are the views of the two eyes at a given moment in time, or of one eye at slightly different moments in time. These problems are discussed, in general terms, in Chapter 1, and then reconsidered in the remaining chapters, where emphasis is given to the formulation of models which are psychologically and physiologically reasonable.

Principle conclusions of the study of stereopsis and the related phenomena of binocular rivalry and fusion (Chapters 2 and 3) are as follows:

- 1) Fusion and suppression are distinct perceptual states.

This conclusion is important because it means that binocular single vision cannot always be explained in terms of suppression of information from one eye. Processes must exist which determine when pairs of monocular features should fuse or be suppressed, as well as determine how features should be paired.

- 2) Neural binocular interactions responsible for suppression and stereopsis must take place at a stage of processing where visual information is monocularly coded. This conclusion is important because it is contrary to current projection field models of stereopsis and interpretations of related physiological data. In both cases interactions are assumed to take place between cells which have disparity-tuned binocular receptive fields.

- 3) At the level of the visual system at which binocular interactions take place, the image may be coded by activity in neurons with center-surround receptive fields. It is not necessary to assume that the code is feature specific to account for rivalry between differently oriented contours, and fusion of similarly oriented contours. These effects may result from spread of suppression due to lateral, recurrent, inhibition.

- 4) The code is "multi channelled", with different channels characterized by elements which differ in the size, but not the shape, of features coded. Stereopsis may result from fusion in one channel, while rivalry occurs in other channels. In this way fusion and rivalry may coexist in the same area of

of the visual field.

5) Binocular combination is stimulus driven. This conclusion has important implications for computer vision because it means that relatively unsophisticated processes, operating at an early stage of image processing, can compute depth from stereo images. It also supports the suggestion that stimulus organizing processes, in general, may be stimulus driven.

A model was proposed for binocular combination which was based upon these conclusions. Computer simulations demonstrated substantial agreement between the rivalry model and a range of empirical results. The principle innovation of the stereopsis model was that the single projection field of previous models was replaced by two coupled monocular projection fields. This structure resolves a number of difficulties with previous models and suggests the possible involvement of the lateral geniculate body in stereopsis.

Motion perception was further considered in Chapter 4. Natural vision typically takes place in an environment which is continually changing, as objects move relative to one another, and the observer moves through space. This possibility of image motion places very important constraints on how visual information is processed and coded. It was proposed that perception of an object in space corresponds to the development of a pattern of activity in a retinotopically organized neural structure. This patterned activity represents the image segments and constitutes a spatial short-term memory. As an

object image moves on the retina, the corresponding patterns of activity must move within the supporting neural structure. In this way image features may be integrated within an appropriate space-time window.

Apparent motion is modeled with a network in which a single brief stimulus may evoke an active transient response (as in the model of Wilson and Cowan, 1973). A sequence of stimuli may be integrated, and cause the perception of a single moving stimulus, if they contribute to the maintenance of a single network response. This will happen if the response activity moves within the neural structure at the appropriate velocity. This model has many properties which are consistent with a range of apparent motion phenomena.

A P P E N D I X    A  
 BINOCULAR INTERACTIONS IN THE  
 LATERAL GENICULATE NUCLEUS

Since the pioneering work of Hubel and Wiesel (1962), it has been known that most cells in cat striate cortex may be excited by an appropriate stimulus presented to either eye, and the receptive fields in each eye are "simple" and similar in orientation and location. More recently it has been discovered that the receptive fields of such binocular cells do not always occupy exactly corresponding positions in the two eyes and one may be displaced by a small angle relative to the other (Barlow, Blakemore and Pettigrew, 1967; Pettigrew, Nikara and Bishop, 1968). For a single object to maximally excite a given cortical neuron, its image must fall on the receptive fields of both eyes, which means it must be located not only in the right direction, but also at the right distance from the cat. "Complex" cells in Brodmann's area 18 of the monkey have also been found to be disparity specific in a similar way<sup>1</sup> (Hubel and Wiesel, 1970).

These discoveries have led to considerable speculation

---

<sup>1</sup>These observations of disparity tuned neurons are apparently not "robust" results, since when Hubel and Wiesel (1973) repeated the experiments of Barlow and Pettigrew, they concluded that binocular neurons in cat visual area 17 do not have disparate receptive fields!

that stereoscopic depth perception is mediated by populations of "disparity tuned" cortical neurons (Bishop, 1969; Blakemore, 1970). It is worth noting that the range of disparities discovered for receptive field disparities is comparable to the range of binocular disparities which yield stereoscopic depth in humans. Also, it is possible to account for stereopsis with complex, locally ambiguous stimuli, such as Julesz random dot stereograms (Julesz, 1960), if a system of inhibition is postulated between cortical cells tuned to different disparities but with receptive fields nearby in space (Dev, 1975; Nelson, 1975).

However, the assumption that disparity tuned cortical cells mediate stereopsis has not led to satisfactory resolution of all questions relating to binocular vision. For example, when different images are presented to the two eyes, the observer (now human) experiences binocular rivalry, and individual stimulus features of one or both images may be suppressed. As was noted in Chapter 2, there will typically be extended regions of the visual field in which all features from one image are suppressed while all features in the other image remain visible. This fact seems to indicate that suppression occurs at a level of the visual system where there is binocular interaction, but where afferent information from the two eyes is separately coded. This type of coding occurs in the lateral geniculate nucleus and layer 4 of striate cortex (Hubel, Wiesel, 1968), but not at the level

of binocular cells in striate cortex.

Another difficulty with the idea that stereopsis and binocular single vision are associated with disparity tuned cortical cells has to do with the phenomenon of allelotropia. When one views a stereogram in which a particular image feature is displaced in the image presented to one eye relative to its position in the other, the apparent position of the "fused" image in the binocularly combined view tends to be intermediate between its positions in the two half images. Allelotropia refers to this apparent displacement in image position. To explain this phenomenon, we might suppose that there is a direction, or "local sign," associated with each binocular cortical neuron which is midway between the visual directions of its two receptive fields and which is perceptually associated with any stimulus which activates that neuron. Since, presumably, the receptive fields are "wired in" and cannot change positions from moment to moment, the visual direction associated with the binocular neuron should also have a fixed, unvarying value. The difficulty arises from the fact that displacement due to allelotropia is not fixed for a given binocularly disparate stimulus, but may seem to change from moment to moment between extremes which are equal to the positions of the stimulus in either monocular image (Charnwood, 1951). This is evidence supporting the idea proposed by Julesz (1971) that disparate monocular images are "shifted" within the brain towards each other

prior to fusion. This internal shifting is dynamic, and monocular images may be shifted by different amounts. It follows from this view that the initial processing responsible for binocular combination must take place on a monocular level, prior to binocular convergence in visual cortex.

The stereopsis model proposed in Chapter 3 results from an attempt to account for these and other types of psychophysical evidence which are not readily explained if we assume stereopsis is mediated by disparity tuned cortical cells. While model structure was motivated principally by this attempt to account for psychophysical data, the models were formulated in terms of networks of idealized neurons, and it appeared that these networks might have natural analogues in the lateral geniculate nucleus (LGN). The objective of this appendix is to further examine this possibility by reviewing what is now known about the anatomy and physiology of this structure. However, it must be acknowledged at the onset that very little of a concrete nature is known about binocular interactions in the LGN, so my attempts to establish parallels with my models are very speculative.

I shall begin by briefly describing the anatomical structure of the LGN. This structure seems quite simple, and an understanding of it will be useful in the subsequent discussion of physiological data.

### Basic LGN Anatomy

The LGN is a thalamic relay nucleus between the retina and visual cortex, as shown in Figure 1a. In animals in which visual fields of the two eyes overlap, there is partial decussation in the optic chiasm so that ganglion cells of one retinal hemisphere of both eyes project to the LGN, which is located on the same side of the head. The LGN itself is divided into several lamina such that the principal relay neurons within a given lamina are enervated by ganglion cells of one eye only (this statement will be qualified slightly, below). The number of lamina differs for different animal species. Figure 1 is based on the cat which has three lamina: A, A<sub>1</sub> and B. Of these, A and A<sub>1</sub> are similar in structure, so presumably perform parallel functions for the two eyes. Principal cells of the LGN project by way of the optic radiation to visual areas of cortex. In the cat, major projections go to Brodmann areas 17 and 18, while in the monkey there is no projection to 18 (Hubel, Wiesel, 1969). There is also a cortex-geniculate efferent projection which seems to originate mainly in area 18 (Hollander, 1970). Individual efferent axons synapse in all layers of the LGN. The projections to LGN lamina are retinotopically organized, and the various layers are "in register," so that nearby neurons in different lamina have receptive fields in roughly the same visual direction, but in different eyes. This

organization is shown in Figure 1b, which is based on a drawing by Bishop et al. (1962). Neurons with similarly located receptive fields fall along "projection lines," as indicated here by the intersection of isoelevation and isoazimuth planes. Neurons located along a given projection line project to the same small area of cortex, where the retinotopic organization is maintained.

There are four classes of neurons in lamina A and A<sub>1</sub> of the cat LGN. The first three of these (classes 1, 2 and 3) were first described by Guillery (1966) and the fourth (class 5) was first described by Famiglietti (1970). (Neurons of class 4 occur only in lamina B, which we will not be concerned with here). The following descriptions of these cell classes follow summaries by Sanderson et al. (1971) and Szentagothai (1973).

Class 1. These are large (perikaryal diameter 25-40  $\mu$ m) multipolar neurons which occur throughout lamina A and A<sub>1</sub>, but principally in or near the central interlaminar nucleus (CIN). They have 4 to 12 dendrites which follow fairly straight radial courses and subdivide once or twice (see Figure 2a). The dendrites cross laminar boundaries freely, so these cells probably correspond to the binocularly enervated cells which are occasionally encountered in physiological studies, and which also are found primarily near the CIN. Axons of class 1 cells project to visual cortex.

Class 2. These are medium sized (perikaryal diameter 15-

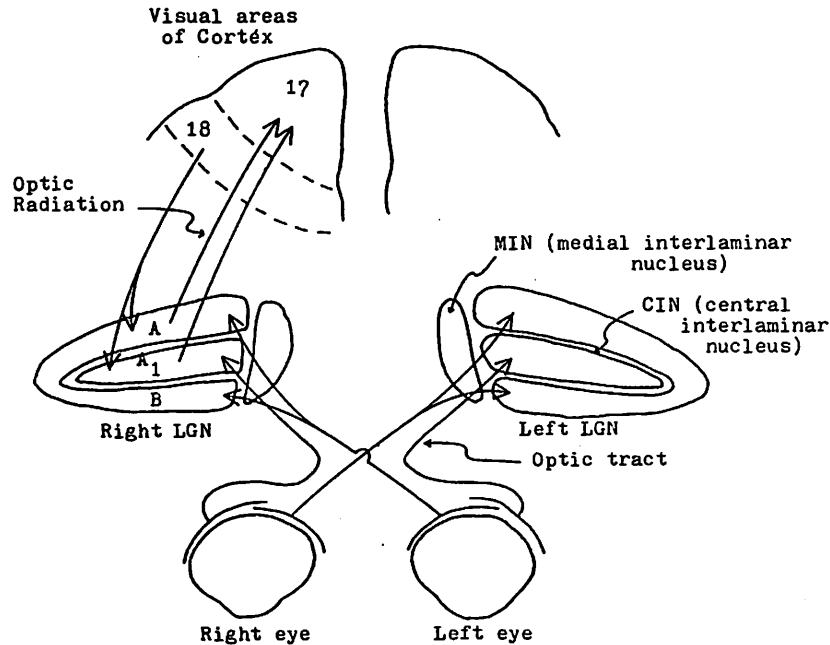


Figure A:1a Schematic drawing of LGN.

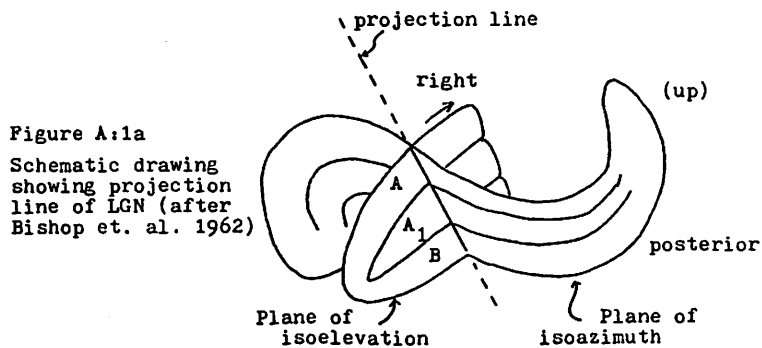


Figure A:1a  
Schematic drawing  
showing projection  
line of LGN (after  
Bishop et. al. 1962)

30  $\mu\text{m}$ ) cells which occur in lamina A and A<sub>1</sub> and which constitute the principal projection or relay cells of the LGN. Dendrites are sinuous (see Figure 2b) and smooth near the soma, but in the region of dendritic branching, there are a number of clusters of "grape-like" appendages where retinal afferents synapse with the relay cells. The dendritic region in which these specialized appendages occur is contained within one lamina, but peripheral portions of the dendrites sometimes cross to other lamina.

**Class 3.** Guillery's class three cells, (also called Golgi type 2 cells) are small interneurons (10 to 20  $\mu\text{m}$  perikaryal size) which occur in lamina A and A<sub>1</sub>. Tombol (1969) divides interneurons into two subclasses:

- a) These neurons have very short axons which do not extend appreciably beyond the dendritic tree. (The corresponding cell type in monkey may have no axon at all, LeVay, 1971). Class 3a neurons remain within a single lamina, where they synapse onto 5 to 10 other neurons, 1 to 3 of which are other interneurons.
- b) These are similar to subclass 3a interneurons, except that axons are longer and cross laminar boundaries, so may serve a binocular function.

**Class 5.** Class 5 neurons were identified by Famiglietti (1970) and occur in all lamina of cat LGN. Dendrites of these neurons appear to taper near their ends into processes which resemble axons.



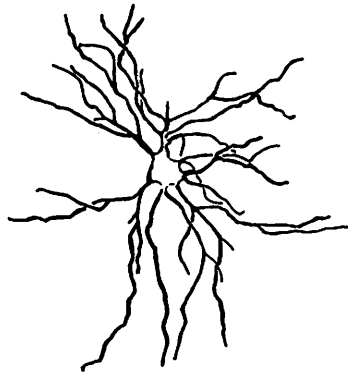


Figure A:2a Class 1 relay cell of rat LGN. (Drawn after Guillery, 1966)



Figure A:2b Class 2 relay cell of cat LGN showing "grape like" appendages near branching points. (Drawn after Guillery, 1966)

We now briefly consider the interconnections of neurons of the various classes. There are several features which deserve emphasis. First, the retinal ganglion cell axons enter the LGN and branch profusely in conical brush-like patterns, as shown in Figure 3a. These ramifications remain within single lamina where they overlap with the ramifications of other ganglion axons and encompass many cell bodies. However, physiological evidence suggests that actual synaptic contact is made with relatively few of these cells (see below). Golgi stains show that these axons end in "flower-like" arrangements, which, it is presumed, make contact with individual "grape-like" appendages on projection neurons (class 2) in synaptic glomeruli structures. These specialized synaptic structures include many individual synapses between several neuron appendages, as in Figure 3b. Usually the glomeruli are encapsulated by glial processes. Within the glomeruli the ganglion terminal occupies the central position and makes excitatory contact with the dendrite of the projection cell as well as with several other dendritic and axonal structures. These structures include two types of interneuron appendages which may be axons and dendrites of the same interneuron, or class of interneurons, and which are identified by flattened synaptic vesicles. In addition to the retinal afferent, there is one other type of excitatory axon ending in the glomeruli (identified by round vesicles), and these are probably the endings of cortex-geniculate efferent fibers.

As is indicated in Figure 3b, axo-axonal synapses occur between the optic fibers and interneurons and dendrite-dendrite synapses occur between interneurons and projection cells.

Other synapses of cortical efferent fibers and interneurons onto projection cells occur outside of the synaptic glomeruli. Figure 3c summarized the various types of interconnections.

#### LGN Physiology

It is thought that the LGN functions principally as a relay nucleus and performs little information processing (Shepard, 1974). A number of physiological facts may be cited in support of this proposition. First, visual images are coded by principal cell activity in the LGN in essentially the same way they are coded by retinal ganglion cell activity: the receptive fields of both types of cells are concentrically organized with on-or off-center and antagonistic surrounds. The sizes of these receptive fields are roughly the same (Maffei, Fiorentini, 1972). The number of optic tract fibers entering the LGN is roughly equal to the number of projection fibers leaving (1,000,000 in humans, according to Shepard (1974)). Stimulus integration is slight, as individual principal cells of the LGN appear to be enervated by only one or a very few ganglion cells (Kato et al., 1971). Also, interneurons are not numerous; in monkeys they are outnumbered

Figure A:3a Terminal "brush" of a ganglion cell in a lamina of the LGE. Holes in the branch pattern show the location of cell bodies. (After Cajal drawing reproduced in Szentagothai, 1973).

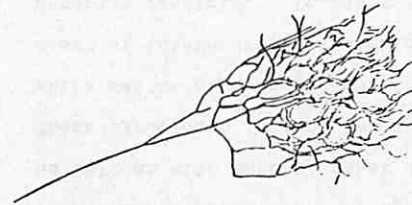
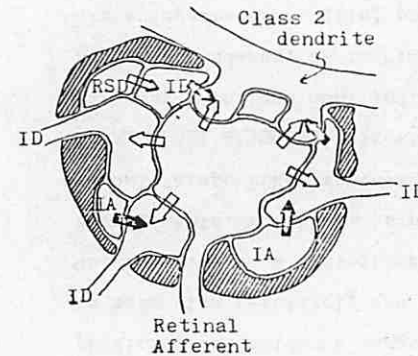


Figure a:3b Synaptic glomerulus.

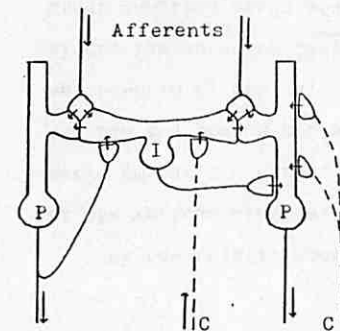


IA: interneuron axon  
ID: interneuron dendrite  
RSD: probably a corticofugal fiber terminal.

↑: excitatory synapse  
↑: inhibitory synapse  
■: glial cell

(based on drawing by Szentagothai, 1973)

Figure A:3c Shepard's (1974) circuit diagram for LGN.



P: principal cell  
I: interneuron  
C: corticofugal fiber

by principal cells by 10 to 1 (LeVay, 1971), so presumably they cannot mediate any complex information processing. Finally, individual characteristics of retinal afferent fibers are retained by the principal cells on which they synapse. For example, fast axon retinal ganglions synapse on fast axon principal cells, and slow synapse on slow (Stone, Hoffman, 1971).

On the other hand, there are also reasons to suspect that more takes place in the LGN than the simple relaying of action potentials. For example, while principal cells outnumber interneurons by 10 to 1, synapses of retinal afferents on principal cells and interneurons together constitute only 20% of the synapses in lamina A and A<sub>1</sub>, while 35% of the synapses are of interneurons onto principal cells and the remaining 45% of synapses are identified, "tentatively," as terminals of corticogeniculate fibers (Guillery, 1969b). In addition, the observation that the ends of some interneurons taper into axon-like processes, and the presence of many dendro-dendritic synapses in glomeruli suggests that portions of interneurons may constitute functional units. By this interpretation, the number of functional inhibitory units may greatly exceed the number of interneurons.

It should be noted also that while the receptive fields of geniculate neurons have the same center-surround organization as ganglion cell receptive fields, they differ in the sense that the center and surround contributions to cell

activity are more balanced in the LGN than in ganglion cells. Maffei and Fiorentini (1972) present evidence that the surround fields of geniculate cells do not correspond to the surround fields of their retinal afferents. Rather, both center and surround fields of LGN cells are enervated by the centers of different afferent cells.

The heavy efferent projection from cortex to LGN must modulate processing within the LGN, which again implies that the LGN has a more elaborate function than simple relay of afferent information. Harth (1976) has speculated on the function of the corticofugal projection from a theoretical point of view, while Kalil and Chase (1970) present physiological evidence of a complex influence on LGN activity. However, this efferent input seems not to be involved in LGN binocular interactions, so will not be considered further here.

There are also projections to the LGN from non-visual brain structures. An input from the reticular formation seems to modulate synaptic excitability in a way which is correlated with the animal's state of wakefulness or drowsiness (Coenen and Vendrick, 1972). A more interesting effect is related to oculomotor activities: whenever the animal executes a saccadic eye movement, there is a "corollary discharge" in the LGN (see review in Freund, 1973). The existence of this effect suggests that the LGN is somehow involved in maintaining visual field stability during eye movements.

The corollary discharge may serve only to suppress visual input during rapid eye movements, or it may cause appropriate displacement of pattern of neural activity in anticipation of afferent stimulus displacement (Burt, 1975). The latter function would be anticipated if the LGN is involved in dynamic "stimulus organizing" processes such as binocular fusion and stereopsis.

Binocular interactions. If the function of the LGN were simply to relay afferent action potentials from optic track fibers to optic radiation fibers, then we might wonder why the LGN is arranged so neatly into layers. The layers segregate afferent activity according to eye of origin, but segregation by layering does not seem necessary to maintain ocular dominance of an optic fiber and the relay cell on which it synapses. We have already remarked that other cell response characteristics, such as speed of impulse propagation, are shared by the optic fiber and relay cell, and this apparently does not require segregation of all cells with common properties into layers. It may be argued, therefore, that segregation of cells according to eye of stimulus origin must serve some purpose involving extensive lateral, and perhaps cooperative, interaction between cells coding information from one or the other eye.

The fact that the retinotopic projections to neighboring layers of the LGN are in register must also be functionally significant. We have already seen that several of the cell

classes of the LGN include cells with dendrites or axons which cross laminar boundaries, and we will review physiological evidence shortly which shows there are inhibitory interlaminar interactions. These facts certainly suggest binocular function in the LGN: the purpose of interlaminar communication may be to coordinate the monocular lateral interactions postulated above. The model proposed for stereopsis in Chapter 3 involves such a system of coupled monocular processes. But before we consider details of this possible stereopsis function, it is appropriate to examine the much simpler interactions required for binocular rivalry.

Rivalry. The principal hypothesis of the model for binocular rivalry are that 1) rivalry occurs at a level of the visual system where image information from the two eyes is separately coded, 2) coding is in terms of neural activity in cells which have antagonistic center-surround receptive fields (such cells respond well when there is a boundary between regions of different stimulus intensities within their receptive fields, and this response is not specific to boundary orientation), 3) there is reciprocal inhibition between cells which have receptive fields centered in the same visual direction but in opposite eyes, and 4) this interocular inhibition is spatially spread, so that an element coding information for one eye inhibits not only the corresponding element of the other eye, but elements which are neighbors of that element as well.

The first two of these hypotheses are certainly satisfied in the LGN. Physiological evidence, which will now be reviewed, shows that the inhibition described by the remaining two hypotheses is also present in the LGN. The question we will be left with will not be whether the inhibitory circuit exists, but whether inhibition is sufficiently strong to account for suppression due to binocular rivalry.

Early evidence of interocular inhibition at the level of the LGN was obtained by Suzuki and Kato (1966). The response of a principal relay cell was recorded intercellularly following stimulation of one or the other optic tract. It was observed that while stimulation of one optic tract would result in a single relay cell spike followed by a prolonged period of hyperpolarization, stimulation of the other optic tract would not produce a spike, but would produce the same hyperpolarization. This interocular, post-synaptic inhibition (IPSP) was observed in 75% of the principal relay cells tested.

Physiological evidence for interocular presynaptic inhibition has also been obtained (Marchiafava, 1966), although no anatomical evidence for axo-axonal synapses onto optic tract fiber terminals has been found. In these experiments, antidromic responses to LGN stimulation were recorded at points in the retina of one eye. It was found that this response could be increased by a conditioning stimulus presented to the corresponding area of the contralateral eye

just prior to LGN stimulation. No increase occurred if the conditioning stimulus was presented to non-corresponding areas of the contralateral retina.

Sanderson, Darian-Smith, Bishop (1969), Sanderson, Bishop, Darian-Smith (1971) and Singer (1970) have been able to map the inhibitory receptive field of relay cells in their non-dominant eye. Sanderson's group used two stimulus conditions. In the first, the activity of a cell was recorded while a bar was moved within the receptive field of the non-dominant eye. No stimulus was presented to the dominant eye. Here it is important to note that LGN cells typically maintain a spontaneous discharge rate of perhaps 5 or 10 spikes per second in the absence of any visual stimulus. Thus, in this first stimulus condition, stimulation of the non-dominant receptive field would cause a change, usually a marked decrease, in the rate of spontaneous discharge. The receptive field in the non-dominant eye could thus be mapped by determining the extent of the area in which a moving bar stimulus altered spontaneous activity. In the 1971 study 113 cells were tested for binocularity in this way. Receptive fields were found in the non-dominant eye of 82% of the cells, and of these, 88% were purely inhibitory. The sizes of non-dominant eye receptive fields varied between 1.5 and 6 degrees, and were considerably larger than the center region of the receptive field of the dominant eye. Furthermore, these two receptive fields occupied corresponding regions of the dominant

and non-dominant visual fields. Inhibition was not stimulus specific; it could be activated by any change in contrast, irrespective of orientation or direction of motion.

All of these results are consistent with the hypothesized interocular inhibition of the rivalry model. In addition, it is of interest to note that inhibition has a latency of about 50 msec. Psychophysical experiments show that suppression due to binocular rivalry may also have a latency of about 50 msec. (Kaufman, 1963).

The second stimulus condition used by Sanderson's group gave results which may be less consistent with the rivalry model hypothesis. In these experiments, the receptive field of the dominant eye was stimulated with a bar which moved continually back and forth, resulting in a considerably increased level of cell activity. When the receptive field of the non-dominant eye was stimulated with another moving bar, it proved nearly ineffective in reducing cell activity. This result may indicate that interocular inhibition is present but not sufficiently strong to mediate suppression due to binocular rivalry.

Singer (1970) obtained results in agreement with those of Sanderson's group: inhibitory receptive fields were found in the non-dominant eye for 32 of 41 neurons investigated, the receptive fields were 2 to 3 degrees in size, and inhibition had a latency of about 60 msec. "Quasi-intracellular" recording revealed long lasting IPSPs, which suggests that

inhibition is postsynaptic. However, Singer found that this interocular inhibition could be very effective, even when the dominant receptive field was stimulated simultaneously with the non-dominant receptive field. In these experiments, the dominant field was stimulated by a stationary light spot, 2 degrees in diameter, while the field of the non-dominant eye was explored with a moving white bar.

Sanderson's failure to find a strong inhibitory effect when both eyes are stimulated, may not be inconsistent with the proposition that suppression takes place in the LGN. We should bear in mind that in psychophysical experiments the occurrence of suppression depends strongly on the nature of the rivalrous binocular stimuli, and suppression is usually studied with stationary images. The effect requires some time to develop (50 msec. as indicated above) and a moving image presented to one eye is very likely to suppress a stationary image presented to the other (consistent with Singer's results). It is not clear from psychophysics that differently moving images presented to the two eyes (as in Sanderson's second stimulus condition) will be particularly effective in producing suppression. In any event, the rivalry model postulates recurrent inhibition, which means that if activity in one neuron is suppressed, then that neuron no longer exerts inhibition on the corresponding element of the other eye. Thus, if we record from a neuron which is dominant in a rivalry situation, we should not expect to find appreciable

inhibition due to a stimulus presented to the other eye, since we presume that stimulus has been suppressed.

We should note that both Sanderson and Singer showed that corticogeniculate efferent fibers could not be responsible for interocular inhibition in the LGN. Sanderson's group lesioned visual areas 17, 18 and 19 and the middle supra-sylvian gyrus while Singer cooled these areas of cortex. Interocular inhibition remained in each case. (Theoretical arguments against corticofugal mediation of suppression in the rivalry model are given in Chapter 2).

There is another demonstration of binocular interactions in the LGN which may be related to the above results. When kittens are raised with one eyelid sutured closed to prevent visual stimulation, the adult cats (with sutures removed) show several visual deficiencies. These cats do not respond behaviorally, and cortical cells do not respond physiologically, to stimuli presented to the deprived eye, and geniculate cells which normally would have been driven by the deprived eye are smaller than those driven by the open eye. A curious feature of these deficiencies is that they are confined to regions of the deprived system which correspond to parts of the visual field which are stimulated by the open eye. Thus deficiencies do not occur, or are much less pronounced, in the monocular peripheral field in the deprived eye, and in regions of normal binocular overlap, where the retina of the open eye has been

lesioned (Sherman et al., 1974, 1975). These results indicate that there is some sort of binocular interaction and competition at the level of the LGN which is important in the development of binocular vision.

Stereopsis. The model I have proposed for stereopsis involved two monocular "projection fields" such as the one shown in Figure 4. A projection field operates essentially as a switching network; each afferent fiber synapses with several relay cells, but only one of these contacts is functional at a given time. This synapse is enabled, and others are disabled by a system of interneurons. Suppose, for example, that the afferent fiber  $a_2$  carries a spike, and that interneuron  $i_2$  is active, then the spike will be transferred to relay cell  $r_2$ . On the other hand, if interneuron  $i_3$  is active, rather than  $i_2$ , then the afferent spike will be transferred to relay cell  $r_1$ . We assume that afferent and relay fibers are retinotopically organized and that the relay cells synapse onto retinotopically organized binocular cells (for example, cells of area 17 in cortex). These binocular cells are also enervated by relay fibers from the other projection field. Details of this system are given in Chapter 3; for the present purpose, the important point to note is that slightly disparate stimuli on the two retina may map to a single binocular cell if the "switches" are appropriately set in the two monocular projection fields. This allows a dynamic control of func-

tionally corresponding retinal points, which may account for the phenomenon of allelotropia mentioned earlier.

If the LGN is the brain structure which mediates stereopsis, then the two monocular projection fields correspond to different lamina; A, and A<sub>1</sub> in the case of the cat. The patterns of inhibition in the two projection fields are coupled and coordinated by interneurons which cross from one projection field to the other, and which also mediate suppression when the binocular stimulus patterns are rivalrous. We have already speculated on the existence of these interconnections between LGN lamina, so now we direct our attention to connections within single lamina which might correspond to those within a projection field.

The interneurons of the LGN are presumed to be inhibitory, so they may "disable" afferent-relay cell synapses directly by either pre- or postsynaptic inhibition. An interneuron might also "enable" a specific synapse by means of disinhibition, if it inhibits the action of other inhibitory neurons on that synapse. These types of connections may, in fact, occur in the LGN. The synapses of retinal afferents onto relay cells take place in the curious and highly specialized synaptic glomeruli mentioned earlier. Within a glomerulus there are many inhibitory interneuron synapses, and some of these inhibit relay cells postsynaptically, while others inhibit the inhibitors.

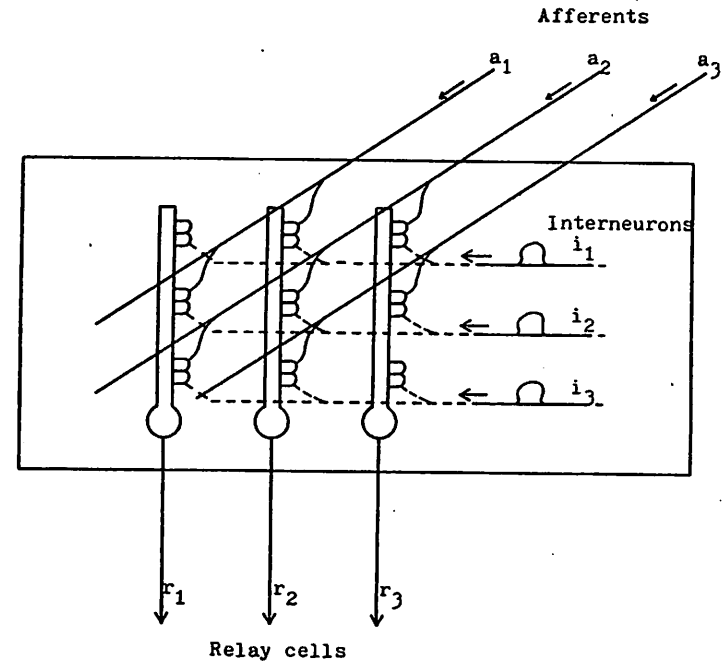


Figure A.4

Diagram of a three unit monocular projection field switching network.



It has already been mentioned that there is physiological evidence that individual afferent fibers synapse onto a few relay cells. This is consistent with the model. On the other hand, it is not clear from empirical data that synapses upon specific relay cells are ever "disabled." Perhaps each afferent fiber actually synapses onto many relay cells, but only a few of these connections are detected physiologically, since synapses with the others have been disabled.

For interneurons to perform the appropriate switching function, each interneuron should be associated with a particular displacement of several afferent fibers relative to relay cells, as is indicated by the vertical position of the interneuron in Figure 4. (We should emphasize here that the idealized neurons in this figure are neatly arranged to clarify their function, but it is not assumed that interneurons in the LGN would have to be similarly arranged in order to perform the same function). An interneuron might inhibit afferent relay synapses which are associated with other displacements, while disinhibiting those which cause the appropriate displacement. Interneurons should also inhibit one another so that within a small area of one lamina, only those interneurons which correspond to a specific displacement will remain active.

There seems to be little empirical data available to corroborate the above predictions. One other prediction is

that patterns of inhibition should change relatively slowly, so that once established, the pattern will appropriately direct subsequent afferent activity. This prediction is consistent with the observation that a stimulated relay cell will produce just one spike (others may follow after several milliseconds delay) while a stimulated interneuron will produce a burst of 7 to 13 spikes (Burke, Sefton, 1966). However, this observation was made for rat and may not hold for cat. Still, in cat, inhibitory effects are long-lived; both pre- and postsynaptic inhibition have effects which last 100 to 300 msec. (Marchiafava, 1966; Suzuki, Kato, 1966).

If we suppose that stereopsis is mediated by inhibitory processes in the LGN, why should binocular cells in cortex (area 17) have disparate receptive fields? (Barlow et al., 1967; Pettigrew et al., 1968). One answer to this question is provided by Hubel and Wiesel (1973), who failed to find such disparities. On the other hand, Sanderson (1971) found that these disparities may already exist in the lateral geniculate nucleus. Sanderson found a characteristic scatter of the dominant eye receptive fields of the principal relay cells which were encountered along a projection line. It was then possible to define a projection "column" in the LGN which contained 90% of LGN neurons with receptive fields centered in a particular direction of the visual field. The distribution of these receptive fields was similar to the distribution of one receptive field relative to the other in

binocular cortical cells. It was suggested, therefore, that single LGN projection columns project to single cortical columns, where synapses are random. Thus the disparities in the cortex result from receptive field scatter in the LGN.

None of these results substantiate a critical prediction of the stereopsis model: we expect that cortical cell receptive fields should vary in position with changes in the state of functional connectivity in the LGN. No shifting of receptive fields has been reported. However, that need not be interpreted as strong evidence against the model, since shift in position might not occur with monocular stimulation, if there is a preferred connectivity. With binocular stimulation, we might expect effects such as those reported for complex cells: the cell may be binocularly excited when the disparity between the monocular stimuli falls within a narrow range, wherever the stimuli are presented within a larger receptive field.

Summary. Many aspects of LGN anatomy and physiology seem to indicate that this structure plays a role in binocular vision. Interlaminar inhibition may mediate suppression due to binocular rivalry, as is suggested by similarities between the LGN and a model I have proposed for this function. The LGN may also play a role in stereopsis. A pattern of neural interconnections has been outlined which, if it exists within lamina of the LGN, allows dynamic shifting of one monocular

image relative to another prior to binocular combination. This interpretation of LGN function seems to be reasonably consistent with much of the empirical data, though no experimental results directly support the interpretation.

A P P E N D I X    B  
ANALYSIS OF DEPTH DOMAIN INHIBITION

A mechanism of inhibition between N cells will be described here which will suppress activity in all but one cell, and the output of that cell will be proportional to its input. The cell which initially becomes dominant will be the cell with the largest excitatory input, but dominance may change abruptly at a later time due to changes in the relative magnitudes of the excitatory inputs. These dominance changes are subject to hysteresis effects.

Suppose there are N match cells in the projection field which are excited by a single input line. For reasons given in Chapter 3, Section 3, these cells should reciprocally inhibit one another, so that when the system reaches a steady state condition, only one cell will be active. The inhibition between any two cells should not depend on the separation of these cells in the projection field, as in Nelson's stereopsis model, (Nelson, 1975), nor are any cells inhibited by other cells which are not members of this group of N cells. We assume that each cell inhibits all members of the group, including itself, by an amount which is proportional to its output, y, so that the total inhibition to each cell is the same for all cells, and may be represented as the output of a single inhibitory cell, as in Figure B:1. The inputs,  $x_i$ , to each of the match cells need not be equal, since, in

addition to the retinal stimulus, there will be excitatory inputs from other cells in the projection field due to space domain facilitation. In addition, each cell may recurrently excite itself.

We assume that the rate of change of a match cell's activity is proportional to the sum of its inputs, and that in the absence of an input, the cell activity decays exponentially to zero. Thus:

$$(Eq. 1) \quad \frac{dy_i(t)}{dt} = \alpha(x_i(t) + y_i(t) - I(t)) - y_i(t)\beta, \quad \text{for } i=1, N.$$

Similarly for the inhibitory cell:

$$(Eq. 2) \quad \frac{dI(t)}{dt} = \gamma \left( \sum_{i=1}^N y_i(t) \right) - \delta I(t).$$

The first equation is subject to the constraint that  $y_i$  can never be negative. Thus if  $y_i(t)=0$  and  $I(t) > x_i(t)$ , then  $dy_i(t)/dt=0$ .

We wish to determine values for the parameters  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  which will cause suppression of activity in all match cells but one. Suppose first that the excitatory inputs,  $x_i$ , are constant in time and  $x_j(t)=\hat{x}$  while  $x_i(t)=0$  for all  $i \neq j$ . When activity in the system reaches equilibrium so that the derivatives in equations 1 and 2 equal zero, we have  $y_j(t) = \hat{y}$  and  $I(t) = \hat{I}$ , where:

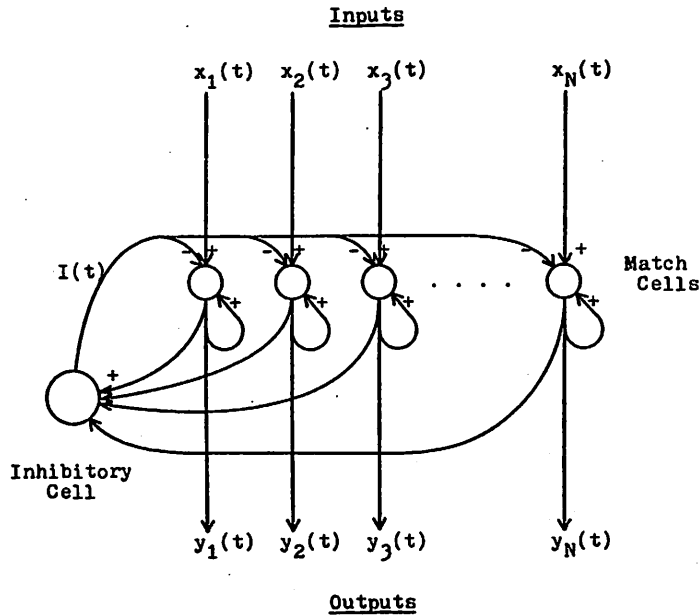


Figure B.1

Connectivity is shown here for a system in which one inhibitory cell suppresses activity in all except one match cell.

$$(Eq. 3) \quad \hat{y} = \frac{\delta \alpha \hat{x}}{\delta(\beta - \alpha) + \gamma \alpha}$$

and

$$(Eq. 4) \quad \hat{I} = \frac{\gamma \alpha \hat{x}}{\delta(\beta - \alpha) + \gamma \alpha}$$

This solution for  $y$  should be positive, so

$$(Eq. 5) \quad \alpha < \frac{\delta \beta}{\delta - \gamma}, \quad \text{or } \gamma > \delta$$

If  $\delta$  and  $\gamma$  are large compared to  $\alpha$  and  $\beta$ , then  $I(t)$  will "follow" changes in  $y_j(t)$  very rapidly, so that to good approximation

$$\begin{aligned} \frac{dy_j(t)}{dt} &= \alpha(\hat{x} + y_j(t) - \frac{\gamma y_j(t)}{\delta}) - \beta y_j(t) \\ &= \alpha \hat{x} + (\alpha(1 - \frac{\gamma}{\delta}) - \beta)y_j(t). \end{aligned}$$

It follows that when equation 5 is satisfied, this equilibrium state will be stable.

Now suppose that after equilibrium has been reached,  $x_j$  is held at  $\hat{x}$  while other inputs,  $x_i$ , are increased from zero. It is clear from equation 1 that none of these inputs will be large enough to activate its cell until it exceeds the inhibition  $\hat{I}$ . If  $\hat{I} > \hat{x}$ , then no  $x_i$  will activate a cell until it also exceeds  $\hat{x}$ , and this is a sufficient condition for hysteresis. From equation 4 we see that  $\hat{I} > \hat{x}$  when  $\alpha > \beta$ .

If at an initial point in time all cell outputs are equal, and we set cell inputs equal to  $\hat{x}$ , then no cell will become dominant and equilibrium will be reached when  $y_i = \hat{y}$  for all  $i$ , where

$$\hat{y} = \frac{\delta \alpha \hat{x}}{\delta(\beta - \alpha) + N\gamma}$$

This equilibrium is stable against equal changes in all  $x_i$  and positive when either condition of equation 5 is satisfied. Otherwise  $\hat{y} = 0$ . However this equilibrium is not stable against an arbitrarily small change in any one of the  $x_i$ . In particular suppose  $x_j$  is increased from  $\hat{x}$  to  $\hat{x} + \Delta$ . This will cause an increase in  $I(t)$ , which in turn will decrease all  $y_i$  except  $y_j$ . The difference between the stimulus to the  $j^{\text{th}}$  cell and the other cells will continue to increase until all  $y_i$  are driven to zero except  $y_j$  which goes to the equilibrium value in equation 3.

To summarize, if parameter values are chosen so that equation 5 is satisfied,  $\delta$  and  $\gamma$  are large compared to  $\alpha$  and  $\beta$ ,  $\alpha > \beta$ , and  $\delta > N\gamma$ , then the stable equilibrium state for this network is one in which only one match cell is active, and the output of this cell is proportional to its input. Furthermore the system exhibits hysteresis behavior because no change in dominance can occur until the input to some suppressed cell exceeds  $\hat{I}$ , which in turn exceeds the input to the dominant cell.

If this type of inhibitory mechanism exists amongst all the match cells stimulated by each input line in the projection field, then it is clear that a stable equilibrium of the net as a whole is one in which there is just one match cell excited for each input line. It also follows that if the retinal stimuli to a number of match cells are equal, the smallest additional excitation of any cell by spatial domain facilitation will cause that cell to become dominant. Thus lateral facilitation in the projection field can be very weak and yet be completely effective in organizing net activity. It also can be shown that the network as a whole will converge on a stable equilibrium and will not enter a limit cycle. For oscillation to occur, there must be at least two cells stimulated by the same input line which are vying for dominance. If these cells are not in an unstable equilibrium then the one with the largest total input will be increasing in activity while other cells' activity is decreasing. The only way in which oscillation could occur is if changes in activity of one of these cells could cause negative feedback upon itself through some sequence of net interactions. However the net is structured so that all paths lead to positive feedback except the cell's direct inhibition of itself, and this only limits the magnitude of the cell's equilibrium activity. Of course there may be changes in the magnitude of lateral facilitation to cells

of a particular mutually inhibiting group, and these input changes may change cell dominance within the group, but such changes in input stimulation cannot be initiated by changes of cell activity within the group, as would be required for oscillation.

## Bibliography

- Arbib, M. (1974) Informal notes on stereopsis.
- Arbib, M. A. (1975a) "Artificial intelligence and brain theory: unities and diversities," COINS Tech. Rep. 75C-8, University of Massachusetts.
- Arbib, M. A. (1975b) "Two papers on schemas and frames," COINS Tech. Rep. 75C-9, University of Massachusetts.
- Asher, H. (1953a) "Suppression theory of binocular vision," Brit. J. Ophthalmol., 37, 37-49.
- Asher, H. (1953b) "The suppression theory," Brit. Orthoptic J., 10, 1-9.
- Barlow, H. B., C. Blakemore and J. D. Pettigrew (1967) "The neural mechanism of binocular depth discrimination," J. Physiol. (London), 193, 327-342.
- Bishop, P. O. (1973) "Neurophysiology of binocular single vision and stereopsis," in R. Jung, ed., Handbook of Sensory Physiology, Vol. 7/3A. New York:Springer-Verlag.
- Bishop, P. O., W. Kozak, W. R. Levick, G. Vakkur (1962) "The determination of the projection of the visual field on to the lateral geniculate nucleus of the cat," J. Physiol., 163, 503-539.
- Blakemore, C. (1970) "The representation of three-dimensional space in the cat's striate cortex," J. Physiol., 209, 155-178.

- Boring, E. G. (1933) The physical dimensions of consciousness. New York:Century.
- Boynton, R. "Discrimination of homogeneous double pulses of light," Chap. 9 in Handbook of Sensory Physiology, VII/4, Visual Psychophysics.
- Burke, W., A. J. Sefton (1966) "Discharge patterns of principal cells and interneurons in LGN in the rat," J. Physiol., 187, 201-212.
- Burt, P. J. (1974) "An examination of possible low level spatial processing in the visual system," COINS Dept. Master's Project.
- Burt, P. J. (1975) "Computer simulations of a dynamic visual perception model," Int. J. Man-Machine Studies, 7, 529-546.
- Burt, P. J. (1976) "A computer program for simulating figure-ground separation in a two-dimensional image," in preparation.
- Charnwood, J. R. B. (1951) Essay on binocular vision. London: Halton Press, (Summarized in Dodwell, 1970).
- Coenen, A. M. L., A. J. H. Vendrik (1972) "Determination of the transfer ratio of cat's geniculate neurons through quasi-intracellular recordings and the relation with the level of consciousness," Exp. Brain Res., 14, 227-242.
- Crovitz, H. F., G. R. Lockhead (1967) "Possible monocular predictors of binocular rivalry of contours," Perception and Psychophysics, 2, 83-85.

- De Silva, H. R., S. H. Bartley (1930) "Summation and subtraction of brightness in binocular perception," Brit. J. Psych., 20, 241-250.
- Dev, P. (1975) "Perception of depth surfaces in random-dot stereograms: a neural model," Int. J. Man-Machine Studies, 7, 511-528.
- Dev, P. (1975) "Segmentation processes in visual perception: a cooperative neural model," Int. J. Man-Machine Studies.
- Dodwell, P. C. (1970) Visual pattern recognition. New York: Holt, Rinehart and Winston.
- Dodwell, P. C. and G. R. Engel (1963) "A theory of binocular fusion," Nature, 198, 39-40, 73-74.
- Enroth-Cugell, C., J. G. Robson (1966) "The contrast sensitivity of retinal ganglion cells of the cat," J. Physiol. (London) 187, 517-552.
- Famiglietti, E. V. (1970) "Dendro-dendritic synapses in the lateral geniculate nucleus of the cat," Brain Res., 20, 181-191.
- Fender, D. H., B. Julesz (1967) "Extension of Panum's fusional area in binocular stabilized vision," J. Opt. Soc. Am., 57, 919-930.
- Freund, H. J. (1973) "Neuronal mechanisms of the lateral geniculate body, in K. Jung, ed., Handbook of Sensory Physiology, Vol. 7/3B. New York:Springer-Verlag.

- Frisby, J. P., B. Julesz (1975) "Depth reduction effect in random line stereograms," Perception.
- Frisby, J. P., B. Julesz (1975) "Some new subjective contours in random-line stereograms," Perception.
- Graham, N. "Spatial-frequency channels in human vision: detecting edges without edge detectors," to appear in C. S. Harris, ed. Visual coding and adaptability. Hillsdale, N.J.: Lawrence Erlbaum Assocs. (distributor: Halsted Press, div'n of John Wiley and Sons, New York).
- Guillery, R. W. (1966) "A study of Golgi preparations from the dorsal lateral geniculate nucleus of the adult cat," J. Comp. Neurol., 128, 21-50.
- Guillery, R. W. (1969a) "The organization of synaptic interconnections in the laminae of the dorsal lateral geniculate nucleus of the cat," Z. Zellforsch., 96, 1-38.
- Guillery, R. W. (1969b) "A quantitative study of synaptic interconnections in the dorsal lateral geniculate nucleus of the cat," Z. Zellforsch., 96, 39-48.
- Hanson, A. R., E. M. Riseman (1975) "The design of a semantically directed vision processor," (revised and updated) COINS Tech. Rep. 75C-1.
- Harth, E. (1976) "Visual perception: a dynamic theory," Biol. Cybernetics, 22, 169-180.
- Harth, E., E. Tzawakon (1974) "ALOPEX: a stochastic method for determining visual receptive fields," Vis. Res., 14, 1475-1482.

- Helmholtz, H. von (1962) Treatise on physiological optics, Vol. 3, trans. from 3rd German ed., J. P. C. Southall, ed., Opt. Soc. Amer., 1925. Republished New York: Dover.
- Hochberg, J. E. (1964a) "Contralateral suppressive fields of binocular combination," Psychon. Soc., 1, 157-158.
- Hochberg, J. (1964b) "Depth perception loss with local monocular suppression: a problem in the explanation of stereopsis," Science, 145, 1334-1336.
- Hollander, H. (1970) "The projection from the visual cortex to the lateral geniculate body (LGB). An experimental study with silver impregnation methods in the cat," Exp. Brain Res., 10, 219-235.
- Hubel, D. H., T. N. Wiesel (1962) "Receptive fields, binocular interactions and functional architecture in the cat's visual cortex," J. Physiol., 160, 106-154.
- Hubel, D. H., T. N. Wiesel (1968) "Receptive fields and functional architecture of monkey striate cortex," J. Physiol., 195, 215-243.
- Hubel, D. H., T. N. Wiesel (1969) "Anatomical demonstration of columns in the monkey striate cortex," Nature, 221, 747.
- Hubel, D. H., T. N. Wiesel (1970) "Cells sensitive to binocular depth in Area 18 of the macaque monkey cortex," Nature, 225, 41-42.
- Hubel, D. H., T. N. Wiesel (1973) "A reexamination of stereoscopic mechanisms in area 17 of cat," J. Physiol., 232, 29-30.



- Huffman, D. A. (1971) "Impossible objects as nonsense sentences," Machine Intelligence, 6, Edinburgh University Press.
- Jastrow, J. (1899) "The mind's eye," Pop. Sci. Monthly, 54, 299-312.
- Johansson, G. (1975) "Visual motion perception," Sci. Am., 232, Num. 6 (June) 76-88.
- Julesz, B. (1960) "Binocular depth perception of computer generated patterns," Bell Syst. Tech. J., 39, 1125-1162.
- Julesz, B. (1964) "Binocular depth perception without familiarity clues," Science, 145, 356-362.
- Julesz, B. (1971) Foundations of cyclopean perception. Chicago:University of Chicago Press.
- Julesz, B. (1976) "Global stereopsis," to be published in Handbook of sensory physiology, Vol. VIII, "Perception," ed. Teuber and Held, Springer.
- Julesz, B., J.-J. Chang (1976) "Interaction between pools of binocular disparity detectors tuned to different disparities," Biol. Cybernetics, 22, 107-119.
- Julesz, B., J. E. Miller (1975) "Independent spatial-frequency-tuned channels in binocular fusion and rivalry," Perception, 4.

- Kalil, R. E., R. Chase (1970) "Corticofugal influence on activity of lateral geniculate neurons in the cat," J. Neurophys., 32, 459-474.
- Kato, H., M. Yamamoto, H. Nakahama (1971) "Intercellular recordings from the lateral geniculate neurons of cats," Jap. J. Physiol., 21, 307-323, summarized in Freund, 1973.
- Kaufman, L. (1963) "On the spread of suppression and binocular rivalry," Vis. Res., 3, 410-415.
- Kaufman, L. (1974) Sight and mind. New York:Oxford University Press.
- Kolers, P. A. (1972) Aspects of motion perception. New York: Pergamon Press.
- LeVay, S. (1971) "On the neurons and synapses of the lateral geniculate nucleus of the monkey, and the effects of eye enucleation," Z. Zellforsch., 113, 396-419.
- Levelt, W. J. M. (1965) On binocular rivalry. Soesterberg, The Netherlands:Institute of Perception.
- Levelt, W. J. M. (1965) Psychological studies on binocular rivalry. Mouton & Co.
- Maffei, L., A. Fiorentini (1972) "Retinogeniculate convergence and analysis of contrast," J. Neurophysiol., 35, 65-72.

Marchiafava, P. L. (1966) "Binocular reciprocal interaction upon optic fiber endings in the lateral geniculate nucleus of the cat," Brain Res., 2, 188-192.

Marr, D. (1974) "A note on the computation of binocular disparity in a symbolic, low-level visual processor," MIT AI Lab. Memo, #327.

Nelson, J. I. (1975) "Globality and stereoscopic fusion in binocular vision," J. Theor. Biol., 49, 1-88.

Ogle, K. N. (1950) Researches in binocular vision. Philadelphia: W. B. Sanders, Co.

Pettigrew, J. D., T. Nikara, and P. O. Bishop (1968) "Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence," Exp. Brain Res., 6, 391-410.

Pitblado, C. B. (1966) "Displacement of half-image during binocular viewing," PhD. diss., Boston University.

Pitts, W. H., W. S. McCulloch (1947) "How we know universals: the perception of auditory and visual forms," Bull. Math. Biophys., 9, 127-147.

Pritchard, R. M. (1961) "Stabilized images on the retina," Sci. Am., 204, 72-78.

Reiss, R. F. (1962) "A theory and simulation of rhythmic behavior due to reciprocal inhibition in small nerve nets," Am. Fed. Inf. Process Soc., Proc. Spr. Joint Computer Conf., 21, 171-194.

Richards, W. (1971) "Anomalous stereoscopic depth perception," J. Opt. Soc. Am., 61, 410-414.

Riggs, L. A., F. Ratliff, J. C. Cornsweet, T. N. Cornsweet (1953) "The disappearance of steadily fixated visual text objects," J. Opt. Soc. Amer., 43, 495-501.

Rosenfeld, A., R. A. Hummel, S. W. Zucker (1976) "Scene labeling by relaxation operations," IEEE Trans. on Systems, Man and Cybernetics, 6, 420-433.

Ross, J. (1972) "Analysis before perception," Research Rept. #4, Dept. of Psychology, U. of Western Australia.

Ross, J. (1974) "Stereopsis by binocular delay," Nature, 248, 363.

Ross, J. (1976) "The resources of binocular perception," Sci. Am., 234, March 80-86.

Ross, J., J. Hogben (1973) "Short-term memory in stereopsis," Vis. Res.

Sakitt, B. (1975) "Locus of short-term visual storage," Science, 190, 1318-1319.

Sanderson, K. J. (1971) "Visual field projection columns and magnification factors in the lateral geniculate nucleus of the cat," Exp. Brain Res., 13, 159-177.

Sanderson, K. J., P. O. Bishop, I. Darian-Smith (1971) "The properties of the binocular receptive fields of the lateral geniculate neurons," Exp. Brain Res., 13, 178-208.

- Sanderson, K. J., I. Darian-Smith, P. O. Bishop (1969)  
"Binocular corresponding receptive fields of single  
units in the cat lateral geniculate nucleus," Vis. Res.,  
9, 1297-1303.
- Shepard, G. M. (1974) The synaptic organization of the brain.  
New York:Oxford University Press.
- Shepard, R. N., S. A. Judd (1976) "Perceptual illusion of  
rotation of three-dimensional objects.
- Sherman, S. M., R. W. Guillery, J. H. Kass, K. J. Sanderson  
(1974) "Behavioral, electrophysiological and morphological  
studies of binocular competition in the development of  
the geniculo-cortical pathways of cats," J. Comp. Neur.,  
158, 1-18.
- Sherman, S. M., J. R. Wilson and R. W. Guillery (1975)  
"Evidence that binocular competition affects the post-  
natal development of Y-cells with cat's lateral genicu-  
late nucleus," Brain Res., 100, 441-444.
- Singer, W. (1970) "Inhibitory binocular interaction in the  
lateral geniculate body of the cat," Brain Res., 18,  
165-170.
- Sperling, G. (1970) "Binocular vision: a physical and a  
neural theory," Am. J. Psych., 83, 463-534.
- Sperling, G. (1975) "Movement perception in computer driven  
visual displays," paper presented at the National Con-  
ference on the Use of On-Line Computers in Psychology,  
Boulder, Colorado, Nov. 5, 1975.

- Stone, J., K. P. Hoffmann (1971) "Conduction velocity as a  
parameter in the organization of afferent relay in the  
cat's lateral geniculate nucleus," Brain Res., 22, 454-  
459.
- Suzuki, H., E. Kato (1966) "Binocular interaction at cat's  
lateral geniculate body," J. Neurophysiol., 29, 909-920.
- Szentagothai, J. (1973) "Neuronal and synaptic architecture  
of the lateral geniculate nucleus," in R. Jung, ed.,  
Handbook of sensory physiology, Vol. 7/3B. New York:  
Springer-Verlag.
- Tombol, T. (1969) "Two types of short axon (Golgi 2nd)  
interneurons in the specific thalamic nuclei," Acta morph.  
Acad. Sci. hung., 17, 285-297, summarized in Szentagothai,  
1973.
- Treisman, A. (1962) "Binocular rivalry and stereoscopic  
depth perception," Quart. J. Exp. Psychol., 14, 23-37.
- Tyler, C. W. (1975) Observations on binocular spatial fre-  
quency reduction in random noise," Perception, 4, 305-  
309.
- Waltz, D. (1975) "Understanding line drawings of scenes with  
shadows," in P. H. Winston, ed., The psychology of com-  
puter vision. New York:McGraw-Hill, 19-91.
- Wertheimer (1912), see Kaufman (1974).
- Wilde, K. (1938) "Figur und Fläche im Wettstreit," Psychol.  
Forsch., 22, 26-58.

Wilson, D. M. (1966) "Central nervous mechanisms for the generation of rhythmic behavior in arthropods," Symp. Soc. Exp. Biol., 20, 199-228.

Wilson, H. R., J. D. Cowan (1973) "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue," Kybernetik, 13.

Wittgenstein, L. (1953) Philosophical investigations.  
trans. G. E. M. Anscombe, Oxford.